

Malaysia Journal of Invention and Innovation

<https://journal.academicapress.org/aps/index.php/mjii>

Research Article

Analisis Ramalan Indeks Kualiti Udara berasaskan Pembelajaran Mesin melalui Aplikasi SEGAR

Azizuddin Hanif¹, Zuraini Zainol^{2,*}, Ummul Fahri Abdul Rauf³, Siti Hasliza Ahmad Rusmili⁴,
Yuhanim Hani Yahya⁵ dan Puteri NE Nohuddin⁶

¹ Jabatan Sains Komputer, Universiti Pertahanan Nasional Malaysia; azizuddinhanif15i@gmail.com

² Jabatan Sains Komputer, Universiti Pertahanan Nasional Malaysia; zuraini@upnm.edu.my

³ Pusat Asasi Pertahanan, Universiti Pertahanan Nasional Malaysia; ummul@upnm.edu.my

⁴ Fakulti Sains dan Teknologi Pertahanan, Universiti Pertahanan Nasional Malaysia; haslizarusmili@gmail.com

⁵ Jabatan Sains Komputer, Universiti Pertahanan Nasional Malaysia; yuhanim@upnm.edu.my

⁶ Faculty of Business, Higher Colleges of Technology, Sharjah, United Arab Emirates; pnohuddin@hct.ac.ae

* Correspondence: zuraini@upnm.edu.my.

Kata Kunci:

Indeks Kualiti Udara
Ramalan Kualiti Udara
Pembelajaran Mesin
Sistem Amaran Awal



Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstrak: Kajian ini membentangkan pembangunan aplikasi SEGAR, sebuah platform berteraskan pembelajaran mesin yang direka khusus untuk menganalisis dan meramal Indeks Kualiti Udara (AQI) di Shah Alam dan Cheras. Menggunakan set data 2018-2022, pembangunan model dimulakan dengan prosedur pembersihan data dan pengisian nilai hilang menggunakan teknik KNN-Imputer. Analisis perbandingan antara model MLR, SVR, dan RF membuktikan keupayaan prediktif algoritma RF yang paling jitu dalam mengenal pasti dinamik corak pencemaran yang kompleks dan tidak linear. Model RF tersebut kemudiannya diintegrasikan ke dalam aplikasi web SEGAR yang dilengkapi fungsi muat naik fail CSV, visualisasi data interaktif, dan mekanisme amaran awal. Rumusannya, SEGAR berpotensi bertindak sebagai sistem sokongan keputusan yang berwibawa bagi pihak berkuasa dalam memitigasi risiko kesihatan awam secara proaktif berasaskan data.

1. PENGENALAN

Udara bersih merupakan elemen asas yang penting dalam menjamin kesihatan manusia, kesejahteraan masyarakat, dan kelestarian alam sekitar. Namun demikian, peningkatan aktiviti perindustrian, pertambahan kenderaan, serta pembangunan pesat di kawasan bandar telah menyebabkan tahap pencemaran udara semakin membimbangkan (Wang et al., 2022; Zhang et al., 2022). Pencemaran udara merupakan antara faktor utama yang menyumbang kepada masalah kesihatan global dengan peningkatan kadar morbiditi serta kelahiran dan kematian pramatang di seluruh dunia (Karimi & Samadi, 2024; Zaini, 2025). Kajian juga menunjukkan bahawa pendedahan jangka panjang kepada bahan pencemar seperti Sulphur Dioxide (SO₂), Particulate Matter (PM_{2.5}/PM₁₀),

Ozone (O_3), Carbon Monoxide (CO) dan Nitrogen Dioxide (NO_2) mempunyai hubungan signifikan dengan penyakit kardiovaskular, pernafasan serta kanser paru-paru (Li et al., 2024; Neo et al., 2022). Selain itu, pencemaran udara juga telah menyumbang kepada jutaan kematian setiap tahun di peringkat global, sekali gus menekankan kepentingan pemantauan dan pengurusan kualiti udara secara berkesan (Orellano et al., 2024).

Di Malaysia, kawasan bandar dan perindustrian seperti Shah Alam dan Lembah Klang sering terdedah kepada risiko pencemaran akibat pelepasan industri, kesesakan lalu lintas dan aktiviti pembangunan pesat (Neo et al., 2022). Keadaan ini menyebabkan tahap kualiti udara boleh berubah secara dinamik dan sukar diramalkan, sekali gus meningkatkan kebimbangan terhadap kesihatan awam dan kualiti hidup. Walau bagaimanapun, aplikasi pemantauan sedia ada seperti MyIPU, APIMS dan ASMC, kebanyakannya hanya memberi tumpuan kepada pelaporan keadaan semasa tanpa keupayaan meramal keadaan masa hadapan. Kekangan ini menyukarkan pihak berkuasa dan masyarakat untuk merancang serta melaksanakan langkah pencegahan awal secara lebih berkesan. Selain itu, kaedah ramalan tradisional yang berasaskan model statistik didapati kurang berkesan dalam menangani corak pencemaran yang kompleks serta tidak linear, sekali gus mengehadkan tahap ketepatan ramalan yang dihasilkan.

Bagi menangani jurang tersebut, sebuah aplikasi ramalan kualiti udara yang dinamakan SEGAR (*System of Environmental Guardian for Air Quality & Response*) telah dibangunkan. Dengan mengintegrasikan teknik pembelajaran mesin, sistem ini berupaya menganalisis data siri masa, mengenal pasti pola pencemaran, serta meramal Indeks Kualiti Udara (AQI) berdasarkan piawaian USEPA (*United States Environmental Protection Agency*) bagi tempoh jangka pendek. Kajian ini memanfaatkan data dari Jabatan Alam Sekitar (JAS) yang memfokuskan kepada kawasan berkepadatan tinggi di Lembah Klang, khususnya stesen Shah Alam dan Cheras, merangkumi tempoh pemerhatian dari 1 Januari 2018 sehingga 31 Disember 2022. Set data tersebut mengandungi atribut pencemar kritikal seperti $PM_{2.5}$, PM_{10} , NO_2 , SO_2 , CO, dan O_3 . Bagi fasa pemodelan, algoritma *Multiple Linear Regression* (MLR), *Random Forest* (RF), dan *Support Vector Regression* (SVR) telah diaplikasikan serta dinilai secara perbandingan bagi menjamin penghasilan model ramalan yang paling rintang (*robust*) dan berprestasi tinggi.

Pembelajaran mesin (*machine learning*) merupakan salah satu cabang kecerdasan buatan yang membolehkan komputer mempelajari corak dan hubungan dalam data secara automatik untuk tujuan analisis dan ramalan. Dalam bidang peramalan kualiti udara, pendekatan ini semakin mendapat perhatian kerana keupayaannya untuk memodelkan hubungan yang kompleks dan tidak linear antara pelbagai parameter pencemar udara. Keupayaan tersebut menjadikan pembelajaran mesin sebagai alternatif yang lebih fleksibel berbanding kaedah statistik tradisional dalam menghasilkan ramalan yang tepat.

Kajian ini mengaplikasikan pendekatan pembelajaran mesin bagi membangunkan model ramalan yang berupaya menangani dinamik hubungan tidak linear antara pemboleh ubah pencemar udara secara fleksibel. Tiga algoritma utama, iaitu MLR, RF, dan SVR, telah dibangunkan serta dinilai secara perbandingan bagi mengenal pasti model dengan prestasi paling optimum. Pemilihan ketiga-tiga algoritma ini didasari oleh beberapa pertimbangan teknikal yang kritikal. Pertama, penggunaan set data primer bagi tempoh lima tahun (2018–2022) adalah sangat ideal untuk latihan model pembelajaran mesin tradisional, sekali gus meminimumkan risiko ralat *overfitting* yang sering kali terjadi pada model yang lebih kompleks. Kedua, pemilihan model-model ini menjamin kecekapan komputasi yang tinggi bagi integrasi aplikasi web, sekali gus membolehkan sistem SEGAR menjana ramalan Indeks Kualiti Udara (AQI) yang tepat dan responsif sebagai mekanisme amaran awal yang efektif kepada pengguna. Sehubungan dengan itu, objektif utama kajian ini adalah (i) mengumpul dan menyediakan set data kualiti udara dengan atribut yang bersesuaian bagi tujuan analisis dan pemodelan; (ii) membangunkan model ramalan AQI menggunakan algoritma pembelajaran mesin

seperti MLR, SVR dan RF; dan (iii) membangunkan satu aplikasi berasaskan web yang dapat memvisualisasikan keputusan ramalan kualiti udara serta menyediakan fungsi amaran awal kepada pengguna.

Secara keseluruhannya, walaupun terdapat pelbagai kajian global mengenai peramalan kualiti udara, keperluan terhadap sebuah sistem aplikasi yang berpusat, mesra pengguna, dan disesuaikan khusus dengan senario tempatan di Malaysia masih kritikal. Trend penyelidikan tempatan kini mula beralih daripada sekadar penilaian model teoretikal kepada pembangunan sistem sokongan keputusan interaktif berasaskan web. Sebagai contoh, dalam domain hidrologi, pemodelan statistik linear maju melalui pendekatan *Seasonal Autoregressive Integrated Moving Average* (SARIMA) telah berjaya diintegrasikan ke dalam aplikasi ramalHujan bagi meramal unjuran kerpasan jangka pendek secara sistematik (Mohd Puad et al., 2025). Selaras dengan perkembangan teknologi pengurusan risiko tersebut, kajian ini memperluaskan implikasi kecerdasan buatan dalam domain kualiti udara pula menerusi pembangunan aplikasi SEGAR.

2. KAJIAN LITERATUR

Penggunaan Indeks Kualiti Udara atau *Air Quality Index* (AQI) adalah penting kerana ia membolehkan kepekatan bahan pencemar ditransformasikan kepada nilai kuantitatif dan kategori risiko kesihatan yang lebih mudah difahami oleh masyarakat umum. Menurut U.S. Environmental Protection Agency (EPA, 2018), AQI berfungsi sebagai suatu mekanisme penyampaian maklumat kepada masyarakat umum berkaitan status kualiti udara di sesuatu kawasan, di samping mengenal pasti kumpulan populasi yang berpotensi terjejas. Jadual 1 menunjukkan klasifikasi AQI yang terbahagi kepada enam kategori utama seperti yang disarankan oleh U.S EPA.

Jadual 1. Klasifikasi Indeks Kualiti Udara (AQI) oleh U.S EPA

Warna AQI	Nilai Indeks	Tahap Pencemaran	Penerangan AQI
Hijau	0-50	Baik	Pencemaran udara berada pada tahap rendah dan kualiti udara adalah memuaskan serta tidak mendatangkan atau hanya memberi kesan minimum terhadap kesihatan.
Kuning	51-100	Sederhana	Kualiti udara adalah memadai. Walau bagaimanapun, sesetengah individu mungkin berisiko terjejas, terutamanya golongan yang sangat sensitif terhadap pencemaran udara.
Jingga (Orange)	101-150	Tidak sihat bagi kumpulan sensitif	Ahli dalam kumpulan sensitif berkemungkinan mengalami kesan kesihatan, manakala impak terhadap populasi umum adalah lebih rendah.
Merah	151-200	Tidak sihat	Sebahagian daripada populasi umum mungkin mengalami kesan kesihatan, manakala kumpulan sensitif berisiko menghadapi masalah kesihatan yang lebih serius.
Ungu	201-300	Sangat Tidak sihat	Amaran kesihatan dikeluarkan: seluruh populasi berada pada risiko tinggi untuk mengalami kesan kesihatan yang buruk.
Maroon	≥ 301	Berbahaya	Semua individu berkemungkinan besar terjejas dan situasi kecemasan kesihatan boleh berlaku, selaras dengan amaran kesihatan yang dikeluarkan.

Beberapa kajian terdahulu telah dijalankan untuk menangani isu pencemaran udara menggunakan teknik pembelajaran mesin di Malaysia. Sebagai contoh, Yohanna and San (2026) telah

membangunkan model ramalan kepekatan PM_{10} dan $PM_{2.5}$ di Pulau Pinang menggunakan algoritma SVR dan MLR. Kajian ini menggunakan data purata bulanan selama lima tahun (2018–2022) yang diperolehi daripada empat stesen pemantauan kawal selia, dengan pelbagai atribut seperti CO, NO_2 , O_3 , SO_2 , suhu, dan Aerosol Optical Depth (AOD). Dapatan analisis menunjukkan bahawa algoritma SVR secara konsisten mengatasi prestasi MLR untuk meramalkan zarah terampai di kawasan bandar yang dipengaruhi oleh monsun.

Seterusnya, Latif et al. (2023) telah menjalankan kajian perbandingan menggunakan algoritma Decision Tree (DT), RF, SVR, dan Automated Neural Network Multi-Layer Perceptron (ANN-MLP) untuk meramal kepekatan karbon monoksida (CO) di Malaysia bagi horizon masa 12 jam dan 24 jam. Kajian ini melibatkan tiga stesen pemantauan strategik di Lembah Klang merangkumi stesen SMK Seri Permaisuri, Cheras (S1), SK Bandar Utama, Petaling Jaya (S2) dan SMK(P) Raja Zarina, Klang (S3). Hasil kajian menunjukkan bahawa model ANN-MLP secara konsisten mempamerkan prestasi paling unggul dengan nilai pekali penentu (R^2) tertinggi mencecah 0.8942, terutamanya di stesen S2. Keberkesanan ANN-MLP turut diperkukuhkan melalui analisis ketidaktentuan yang mencatatkan tahap keyakinan sehingga 99.99% dengan nilai *d-factor* minimum 0.000250455.

Selain itu, Neo et al. (2022) telah mengusulkan model ramalan kualiti udara *end-to-end* menggunakan algoritma AdaBoost, SVR, RF, dan KNN, serta teknik deep learning iaitu MLP Regressor dan *Long Short-Term Memory* (LSTM) di empat bandar utama di Selangor (Petaling Jaya, Banting, Klang, dan Shah Alam). Kajian ini menekankan kepentingan proses pengoptimuman ciri (*feature optimization*) untuk mengurangkan dimensi data bagi meningkatkan kecekapan ramalan. Hasil kajian menunjukkan bahawa model LSTM memberikan prestasi paling cemerlang dengan nilai R^2 antara 0.918 hingga 0.998. Dapatan kajian turut mengenal pasti bahawa $PM_{2.5}$, PM_{10} , NO_2 , kelajuan angin, dan kelembapan merupakan elemen paling kritikal untuk dipantau.

Kajian oleh Abdul Rahman et al. (2023) pula menilai keberkesanan algoritma *Generalized Linear Model* (GLM), DT, dan *Support Vector Machine* (SVM) merentasi 125 lokasi rawak di seluruh Malaysia, termasuk bandar-bandar seperti Ayer Keroh, Taiping, Rawang, dan Kudat. Dengan memanfaatkan data daripada platform *eLichens*, kajian tersebut merumuskan bahawa model SVM mempamerkan ketepatan tertinggi dengan nilai R^2 melebihi 90%. Kajian ini juga mengenal pasti $PM_{2.5}$ sebagai parameter paling signifikan yang mempunyai pengaruh linear kuat dalam penentuan nilai Indeks Kualiti Udara (AQI) di Malaysia.

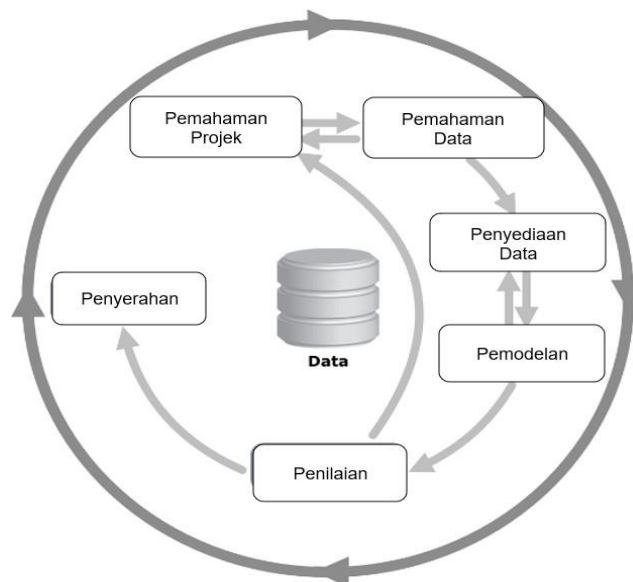
Berdasarkan pemerhatian terhadap kajian-kajian lepas, didapati bahawa kebanyakan penyelidikan seperti Latif et al. (2023), Neo et al. (2022), dan Abdul Rahman et al. (2023) lebih memberi fokus kepada pembangunan model ramalan dan penilaian prestasi model secara teoritikal. Walau bagaimanapun, terdapat jurang di mana kurang penekanan diberikan kepada pengintegrasian model terbaik ke dalam aplikasi praktikal yang boleh digunakan secara langsung oleh pengguna akhir. Oleh itu, kajian ini membangunkan aplikasi SEGAR dengan mengintegrasikan data sejarah Jabatan Alam Sekitar (JAS) dan pengiraan AQI berasaskan piawaian USEPA. Melalui pembangunan model MLR, SVR, dan RF, kajian ini memberi nilai tambah praktikal menerusi visualisasi data, fungsi amaran awal, dan sokongan sistem pemantauan yang lebih proaktif.

3. METODOLOGI

Kajian ini menggunakan metodologi *Cross Industry Standard Process for Data Mining* (CRISP-DM) sebagai kerangka utama dalam pembangunan Aplikasi Ramalan Kualiti Udara menggunakan Pembelajaran Mesin (SEGAR) rujuk Rajah 1. Pemilihan metodologi ini adalah bersesuaian dengan sifat kajian yang melibatkan analisis data, pembangunan model ramalan, penilaian prestasi model, dan pelaksanaan model ke dalam satu aplikasi. Secara umumnya, CRISP-DM mengandungi enam fasa

utama, iaitu pemahaman kajian, pengumpulan data, penyediaan data, pemodelan, penilaian, dan pelaksanaan. Pendekatan ini membolehkan setiap proses kajian dijalankan secara sistematik dan teratur, bermula daripada pemahaman masalah kajian sehingga kepada pembangunan aplikasi ramalan yang boleh digunakan dalam situasi sebenar.

Pada fasa pertama, kajian memberi penekanan kepada aspek pemahaman kajian yang melibatkan penentuan objektif, skop, dan keperluan aplikasi yang ingin dibangunkan. Dalam konteks ini, kajian memfokuskan kepada pembangunan satu aplikasi ramalan kualiti udara berasaskan pembelajaran mesin yang mampu meramal Indeks Kualiti Udara (AQI) berdasarkan data sejarah pencemar udara. Keperluan terhadap aplikasi seperti ini timbul kerana kaedah pemantauan sedia ada lazimnya lebih tertumpu kepada pelaporan bacaan semasa, manakala keupayaan untuk meramal keadaan kualiti udara pada masa hadapan masih terhad. Oleh itu, pembangunan SEGAR dilihat sebagai satu pendekatan yang dapat menyokong pemantauan kualiti udara secara lebih proaktif melalui fungsi ramalan awal, paparan keputusan, dan amaran kepada pengguna apabila tahap pencemaran dijangka meningkat.



Rajah 1. CRISP-DM diadaptasi daripada (Wirth & Hipp, 2000)

Bagi merealisasikan objektif tersebut, kajian ini menggunakan data kualiti udara yang diperoleh dari JAS, melibatkan dua lokasi utama di kawasan Lembah Klang, iaitu Shah Alam dan Cheras, bagi tempoh 1 Januari 2018 hingga 31 Disember 2022. Pemilihan kedua-dua lokasi ini dibuat kerana kawasan tersebut mewakili persekitaran bandar dan perindustrian yang sering terdedah kepada perubahan tahap pencemaran akibat pelepasan industri, peningkatan jumlah kenderaan, dan aktiviti pembangunan setempat. Set data yang digunakan mengandungi enam parameter pencemar utama, iaitu $PM_{2.5}$, PM_{10} , NO_2 , SO_2 , CO , dan O_3 , yang menjadi asas dalam pengiraan AQI serta pembangunan model ramalan dalam aplikasi SEGAR. Pemilihan atribut ini adalah berdasarkan kajian terdahulu yang menunjukkan kepentingan parameter tersebut dalam penilaian dan ramalan kualiti udara (Méndez et al., 2023; Zaini et al., 2025).

STATION ID	LOCATION	DATE TIME	PM10 (µg/m3)	PM2.5 (µg/m3)	SO2 (ppm)	NO2 (ppm)	NO (ppm)	NOX (ppm)	O3 (ppm)	CO (ppm)	WIND DIRECTION (°)	WIND SPEED (m/s)	RELATIVE HUMIDITY (%)	Ambient Temperature (°C)
CA20B	Shah Alam, SELANGOR	1/1/2018	17.806	12.785	0.0007	0.0085	0.0071	0.015	0.0092	0.736	277.6405	0.8411	88.1564	25.0239
CA20B	Shah Alam, SELANGOR	2/1/2018	28.261	20.15	0.0005	0.0232	0.0228	0.0455	0.0123	1.128	247.9378	0.6954	84.2035	26.5468
CA20B	Shah Alam, SELANGOR	3/1/2018	22.74	15.946	0.0005	0.0184	0.0112	0.03	0.01	0.908	213.8232	0.6746	87.1907	25.3622
CA20B	Shah Alam, SELANGOR	4/1/2018	31.314	22.461	0.0006	0.0235	0.0172	0.0393	0.0142	0.989	258.1867	1.1461	86.1193	25.7314
CA20B	Shah Alam, SELANGOR	5/1/2018	22.306	15.944	0.0006	0.0277	0.0129	0.0395	0.009	0.979	160.4057	0.7853	87.3457	25.3183
CA20B	Shah Alam, SELANGOR	6/1/2018	28.534	20.462	0.0007	0.022	0.0214	0.0423	0.0167	1.068	171.86	0.7791	74.8258	27.5127
CA20B	Shah Alam, SELANGOR	7/1/2018	39.421	29.821	0.0007	0.0162	0.0233	0.0376	0.0196	1.152	162.9502	0.8139	70.3385	28.6206
CA20B	Shah Alam, SELANGOR	8/1/2018	32.858	24.189	0.0007	0.0199	0.0029	0.0197	0.027	0.812	201.5386	1.18	73.5456	27.642
CA20B	Shah Alam, SELANGOR	9/1/2018	23.743	17.776	0.0005	0.0151	0.0003	0.011	0.0208	0.616	282.9692	1.2842	82.8628	26.0563
CA20B	Shah Alam, SELANGOR	10/1/2018	16.133	11.492	0.0004	0.0123	0.0007	0.0226	0.0203	0.73	213.3376	0.894	83.8276	25.6304
CA20B	Shah Alam, SELANGOR	11/1/2018	16.24	10.941	0.0004	0.0124	0.023	0.0354	0.0035	0.819	171.5684	0.7256	91.7304	22.9883
CA20B	Shah Alam, SELANGOR	12/1/2018	16.707	11.275	0.0004	0.0173	0.0192	0.0365	0.0034	0.943	136.6352	0.6313	91.0507	22.31
CA20B	Shah Alam, SELANGOR	13/1/2018	26.902	17.243	0.0006	0.0156	0.0156	0.0312	0.01	1.012	231.1531	0.6772	85.5136	23.73
CA20B	Shah Alam, SELANGOR	14/1/2018	36.517	24.458	0.0005	0.0172	0.0265	0.0411	0.0193	1.296	211.4855	0.449	77.7058	25.15
CA20B	Shah Alam, SELANGOR	15/1/2018	38.817	26.895	0.0009	0.0157	0.0151	0.0296	0.0247	1.107	188.299	0.8412	75.7451	25.786
CA20B	Shah Alam, SELANGOR	16/1/2018	45.453	33.162	0.0014	0.0308	0.015	0.0467	0.022	1.375	165.3485	0.5513	74.8827	26.7543
CA20B	Shah Alam, SELANGOR	17/1/2018	69.33	45.717	0.0012	0.0251	0.0186	0.0423	0.027	1.273	180.0055	0.7718	76.8423	26.5474
CA20B	Shah Alam, SELANGOR	18/1/2018	35.023	28.648	0.0015	0.027	0.0058	0.0329	0.0243	0.985	190.211	0.8061	82.2749	26.4845
CA20B	Shah Alam, SELANGOR	19/1/2018	38.603	29.641	0.0017	0.0227	0.0116	0.0337	0.0205	0.968	248.9901	0.8041	84.9282	25.9122
CA20B	Shah Alam, SELANGOR	20/1/2018	27.433	21.368	0.0019	0.0196	0.0079	0.0271	0.0184	0.794	208.2918	0.7107	82.0717	26.4815
CA20B	Shah Alam, SELANGOR	21/1/2018	33.189	25.948	0.0021	0.0181	0.008	0.0248	0.0212	0.909	177.5322	0.6842	82.2552	26.6739
CA20B	Shah Alam, SELANGOR	22/1/2018	25.094	18.597	0.0012	0.0106	0.0032	0.0137	0.0222	0.791	251.762	0.9625	79.9034	27.1898
CA20B	Shah Alam, SELANGOR	23/1/2018	29.604	20.02	0.0018	0.0155	0.0042	0.0196	0.0143	0.771	270.4216	1.2032	82.1144	26.7356
CA20B	Shah Alam, SELANGOR	24/1/2018	21.339	14.086	0.0011	0.0139	0.0092	0.0229	0.011	0.868	233.6617	0.9728	85.6262	25.4942
CA20B	Shah Alam, SELANGOR	25/1/2018	23.28	18.686	0.0019	0.0162	0.0165	0.0366	0.0173	1.096	168.2263	0.8794	81.797	26.1927
CA20B	Shah Alam, SELANGOR	26/1/2018	44.921	33.233	0.0019	0.0276	0.0285	0.0555	0.0133	1.553	137.4445	0.6564	86.962	25.3047
CA20B	Shah Alam, SELANGOR	27/1/2018	26.449	18.455	0.0013	0.0153	0.0136	0.0272	0.0236	0.987	211.4006	0.7983	76.575	26.9901
CA20B	Shah Alam, SELANGOR	28/1/2018	29.375	22.637	0.0013	0.0149	0.0081	0.0187	0.0211	0.931	222.6907	0.9755	79.2495	27.3446
CA20B	Shah Alam, SELANGOR	29/1/2018	21.936	15.994	0.0012	0.0113	0.0066	0.0176	0.0155	0.91	173.1822	1.1799	80.6568	26.7914
CA20B	Shah Alam, SELANGOR	30/1/2018	28.449	18.979	0.0021	0.0133	0.0138	0.0268	0.0193	1.039	240.4534	1.0395	76.9676	27.9613
CA20B	Shah Alam, SELANGOR	31/1/2018	29.63	21.944	0.0018	0.0148	0.0052	0.0204	0.0191	1.035	195.434	0.823	76.2864	27.9882
CA20B	Shah Alam, SELANGOR	1/2/2018	28.881	21.586	0.0013	0.0208	0.0172	0.0424	0.019	1.37	180.9832	0.6783	79.1937	27.3352

Rajah 2. Contoh Set Data Mentah Udara Stesen Shah Alam, Selangor.

Rajah 2 menunjukkan set data mentah yang digunakan dalam kajian ini dikesan mempunyai beberapa isu integriti data seperti nilai hilang (missing values), rekod pendua, ketidakkonsistenan format, serta kehadiran titik data luar julat (outliers). Bagi menjamin kualiti data yang optimum untuk fasa analisis, satu prosedur penyediaan data yang sistematik telah dilaksanakan. Prosedur ini merangkumi penyelarasan format bacaan, penyingkiran rekod pendua, serta pengkelasan bacaan yang tidak logik sebagai nilai hilang. Seterusnya, nilai-nilai hilang tersebut diisi semula menggunakan teknik KNN Imputer bagi mengekalkan korelasi antara pemboleh ubah (Rajah 3). Akhir sekali, proses penskalaan data (*feature scaling*) dijalankan bagi menyeragamkan julat atribut, sekali gus memastikan setiap parameter dapat memberikan nilai yang seimbang semasa proses pemodelan pembelajaran mesin.

```

if len(calon_num) > 0:
    imputer = KNNImputer(n_neighbors=5, weights="distance")
    df[calon_num] = imputer.fit_transform(df[calon_num])

print("\n>>> Ringkasan nilai hilang selepas interpolasi + KNNImputer:")
print(df[calon_num].isna().sum() if len(calon_num) > 0 else "(Tiada lajur numerik)")

```

Rajah 3. Paparan kod Python untuk baris yang kehilangan data menggunakan teknik *KNN-Imputer*.

Dalam fasa pemodelan, tiga algoritma pembelajaran mesin iaitu MLR, SVR dan RF telah digunakan untuk membangunkan model ramalan AQI. Bagi menjamin kebolehpercayaan model dan mengelakkan isu ketirisan data (*data leakage*), set data telah dibahagikan kepada set latihan dan set ujian berdasarkan urutan masa. Prestasi setiap model dinilai secara kuantitatif menggunakan metrik piawai industri iaitu Min Ralat Mutlak (MAE), Punca Min Kuasa Dua Ralat (RMSE), dan R^2 . Selain penilaian berangka, perbandingan visual melalui plot graf antara nilai ramalan dan nilai sebenar turut dilaksanakan bagi menganalisis kecekapan model dalam menangkap trend pencemaran.

Berdasarkan analisis perbandingan dalam Jadual 1, model RF dikenal pasti sebagai model paling unggul dengan mencatatkan nilai ralat yang paling minimum, iaitu MAE = 0.1505 dan RMSE = 1.0388. Keupayaan prediktif RF turut dibuktikan melalui nilai R^2 yang sangat tinggi iaitu 0.9944, yang menunjukkan keupayaan model dalam menerangkan 99.44% variasi data AQI dengan ralat ramalan yang sangat marginal. Sebagai perbandingan, model SVR menunjukkan prestasi yang kompetitif dengan nilai R^2 = 0.9352, namun masih menunjukkan ralat yang lebih tinggi berbanding RF (MAE = 1.8190; RMSE = 3.5458). Manakala, model MLR mencatatkan prestasi paling rendah dengan nilai R^2 = 0.7413 dan ralat yang ketara (RMSE = 7.0851), sekali gus membuktikan bahawa pendekatan linear tradisional kurang efektif dalam menangani dinamik data kualiti udara yang kompleks. Secara keseluruhannya, keputusan ini mengukuhkan pemilihan RF sebagai model teras bagi pembangunan

Aplikasi SEGAR kerana keupayaannya dalam mengendalikan hubungan data bukan linear secara rintang (robust), stabil, dan jitu.

Jadual 1. Laporan Perbandingan Nilai MAE, RMSE dan R²

Model	MAE	RMSE	R ²
MLR	5.3268	7.0851	0.7413
SVR	1.8190	3.5458	0.9352
RF	0.1505	1.0388	0.9944

Seterusnya, fasa pembangunan aplikasi SEGAR dilaksanakan dengan mengintegrasikan model ramalan terbaik ke dalam persekitaran web yang interaktif. Proses ini dimulakan dengan fasa Penyediaan Persekitaran Pembangunan yang melibatkan konfigurasi platform serta *library* pengaturcaraan yang diperlukan. Selepas persekitaran sedia ada, fasa Pembangunan Antara Muka Pengguna (UI) dijalankan bagi memastikan aplikasi mempunyai reka bentuk yang mesra pengguna dan mudah dinavigasi. Antara muka ini direka khusus untuk membolehkan pengguna memuat naik data kualiti udara dalam format CSV secara sistematik.

Fasa kritikal seterusnya ialah Integrasi Model Pembelajaran Mesin, di mana model *Random Forest* (RF) yang telah dioptimumkan disepadukan ke dalam logik aplikasi bagi membolehkan proses ramalan berlaku secara automatik sebaik sahaja data dimuat naik. Bagi memberikan nilai tambah kepada pengguna, fasa Pembangunan Visualisasi Data diimplementasikan untuk memaparkan trend pencemaran dalam bentuk graf yang mudah difahami. Akhir sekali, fasa pembangunan sistem amaran awal dibangunkan sebagai mekanisme notifikasi automatik sekiranya bacaan AQI ramalan menunjukkan tahap yang tidak sihat, sekaligus memenuhi objektif utama aplikasi sebagai sistem sokongan keputusan yang proaktif.

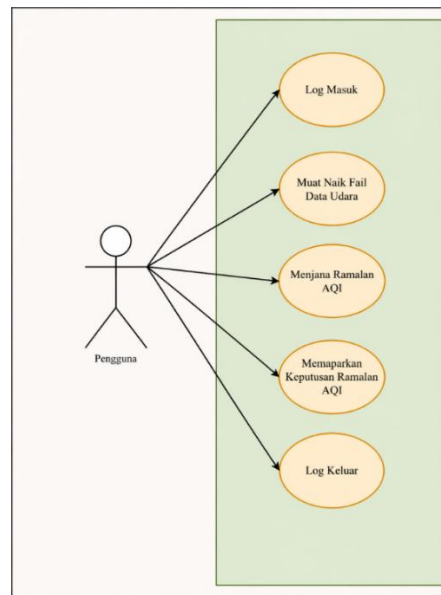
4. ANALISIS APLIKASI

Fasa analisis merupakan komponen kritikal dalam kitaran hayat pembangunan sistem bagi memastikan aplikasi yang dibangunkan memenuhi keperluan pengguna dan objektif kajian. Bagi aplikasi SEGAR, proses analisis dibahagikan kepada empat domain utama iaitu analisis keperluan, analisis data dan proses, analisis objek, serta transisi reka bentuk.

4.1 Fasa Pemodelan Keperluan (Analisis Use Case)

Pemodelan keperluan dilaksanakan untuk mengenal pasti interaksi antara pengguna dan sistem. Rajah 4 memaparkan Rajah Kes Guna (Use Case Diagram) yang menggariskan fungsi-fungsi teras aplikasi SEGAR. Pengguna mempunyai autoriti untuk melaksanakan beberapa operasi utama, antaranya:

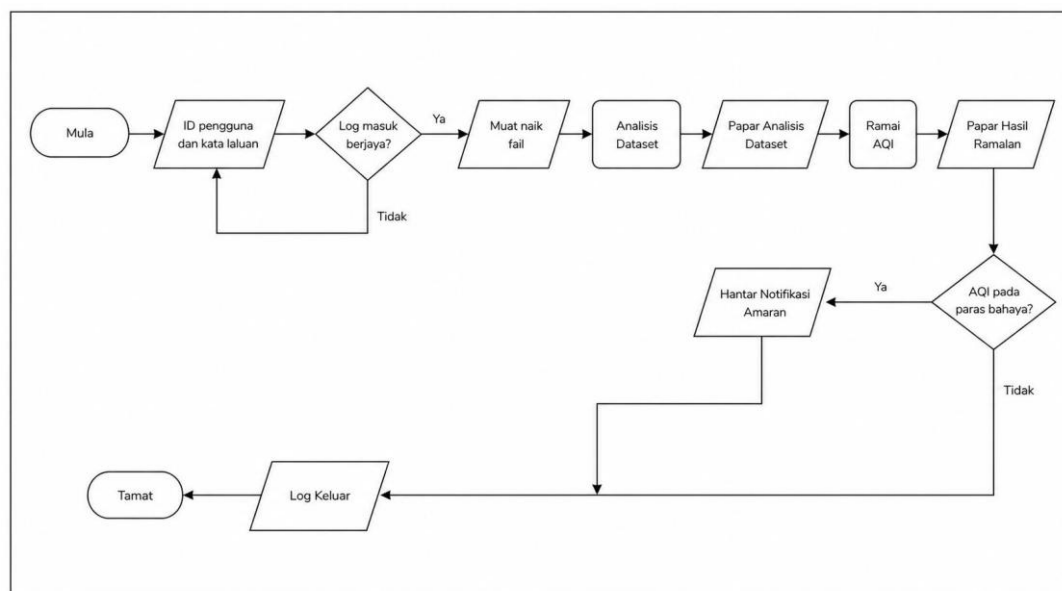
- Log Masuk: Mekanisme pengesahan identiti bagi menjamin keselamatan data.
- Muat Naik Fail CSV: Proses penginputan dataset kualiti udara mentah untuk diproses oleh enjin pembelajaran mesin.
- Jana Ramalan AQI: Fungsi teras yang mengaktifkan model Random Forest (RF) untuk meramal tahap pencemaran.
- Paparan Keputusan: Visualisasi hasil ramalan dalam bentuk grafik dan jadual interaktif



Rajah 4. Gambarajah *use case* bagi aplikasi SEGAR.

4.2 Pemodelan Data dan Proses (Carta Alir)

Bagi memperincikan logik operasi sistem, sebuah carta alir proses telah dirangka seperti yang ditunjukkan dalam Rajah 5. Aliran kerja dimulakan dengan fasa pra-pemprosesan data automatik sebaik sahaja fail dimuat naik. Sistem kemudiannya akan melaksanakan rutin ramalan menggunakan model yang telah dilatih. Komponen penting dalam proses ini adalah Logik Amaran Awal, di mana sistem akan melakukan penilaian ambang (*threshold assessment*) berdasarkan piawaian US EPA (rujuk Jadual 1). Sekiranya nilai ramalan melebihi tahap julat sihat, sistem akan menjana notifikasi amaran secara automatik bagi menyokong tindakan mitigasi awal oleh pengguna.



Rajah 5. Carta Alir bagi Aplikasi SEGAR.

5. MODEL RAMALAN & APLIKASI SEGAR

Reka Pembangunan aplikasi SEGAR merupakan fasa penterjemahan model ramalan terbaik kepada satu platform digital yang berfungsi secara interaktif. Aplikasi ini dibangunkan menggunakan kerangka kerja berasaskan Python untuk memastikan kelancaran proses manipulasi data, pelaksanaan model pembelajaran mesin, dan visualisasi hasil ramalan dalam satu ekosistem yang bersepadu.

PANDUAN PENGGUNAAN SISTEM SEGAR
Sistem Ramalan Kualiti Udara

1 LOG MASUK KE SISTEM

Masukkan ID Pengguna dan Kata Laluan untuk log masuk.

a. Masukkan ID Pengguna
b. Masukkan Kata Laluan
c. Klik "Log Masuk"

2 HALAMAN UTAMA (DASHBOARD)

Halaman utama dipaparkan selepas log masuk berjaya.

a. Nota penting penggunaan sistem.
b. Butang untuk memuat naik dataset (fail CSV).
c. Butang untuk menjalankan analisis dan pembersihan data.

3 ANALISIS DATASET

Lihat ringkasan data sebelum membuat ramalan.

a. Maklumat dataset dipaparkan:
• Bilangan baris
• Bilangan kolom
• Nilai hilang
• Contoh data
• Ringkasan kolom

b. Terus ke bahagian tetapan ramalan.

4 LIHAT HASIL RAMALAN

a. Nota sistem dan ringkasan AQI dipaparkan
b. Graf Trend AQI
c. Infografik Jadual AQI (Rujukan Kategori)
d. Tiga tab rekod:
• Rekod Amaran
• Hasil Ramalan
• Ramalan Masa Hadapan

Hasil Ramalan AQI

Nota Sistem:
• Notifikasi e-mel: DIHIDUPKAN (a*****@gmail.com) - Status: DHANTAR.
• Notifikasi emel berjaya dihantar ke aishahadlan178@gmail.com.

Bil. Amaran: 22
1.2% daripada keseluruhan

Min AQI: 23.3
Purata AQI: 67.0
Maksimum AQI: 189.7

Ringkasan: Terdapat 22 rekod melebihi/bersamaan ambang 150.0. Sila ambil langkah berjaga-jaga.

Trend AQI (Graf)
Tetapan Trend: Bulanan
Muat Turun Cleaned CSV

Rekod Amaran

Tarikh/Masa	AQI	Kategori AQI
2018-08-17 00:00:00	157.3	Unhealthy
2019-03-20 00:00:00	157.1	Unhealthy
2019-09-05 00:00:00	151.6	Unhealthy
2019-09-06 00:00:00	157.5	Unhealthy
2019-09-07 00:00:00	152.2	Unhealthy

Ramalan Masa Hadapan

Navigasi Rekod (Ramalan Masa Hadapan)
Baris / Halaman: 50

Jumlah: 5 Halaman: 1 / 1 Julat: 1 - 5

Tarikh/Masa	Amaran	AQI	Kategori AQI
2023-01-01 00:00:00	TIADA	70.0	Moderate
2023-01-02 00:00:00	TIADA	72.4	Moderate
2023-01-03 00:00:00	TIADA	66.9	Moderate
2023-01-04 00:00:00	TIADA	70.7	Moderate
2023-01-05 00:00:00	TIADA	68.6	Moderate

Ramalan masa hadapan bermula selepas rekod terakhir dataset (skut unit yang dipili: jam/hari).

Infografik Jadual AQI
Rujukan kategori AQI, julat bacaan dan peringatan kesihatan.

Air Quality	Air Quality Index	PM2.5 (µg/m³)	Health Advisory
Good	0-50	0-15	None
Moderate	51-100	16-40	Unusually sensitive people should consider reducing prolonged or heavy exertion.
Unhealthy for Sensitive Groups	101-150	41-65	People with heart or lung disease, older adults, and children should reduce prolonged or heavy exertion.
Unhealthy	151-200	66-150	People with heart or lung disease, older adults, and children should avoid all physical activity outdoors. Everyone else should avoid prolonged or heavy exertion.
Very Unhealthy	201-300	151-250	People with heart or lung disease, older adults, and children should avoid all physical activity outdoors. Everyone else should avoid prolonged or heavy exertion.

Rajah 6. Paparan Langkah atau Tatacara Menggunakan Aplikasi SEGAR

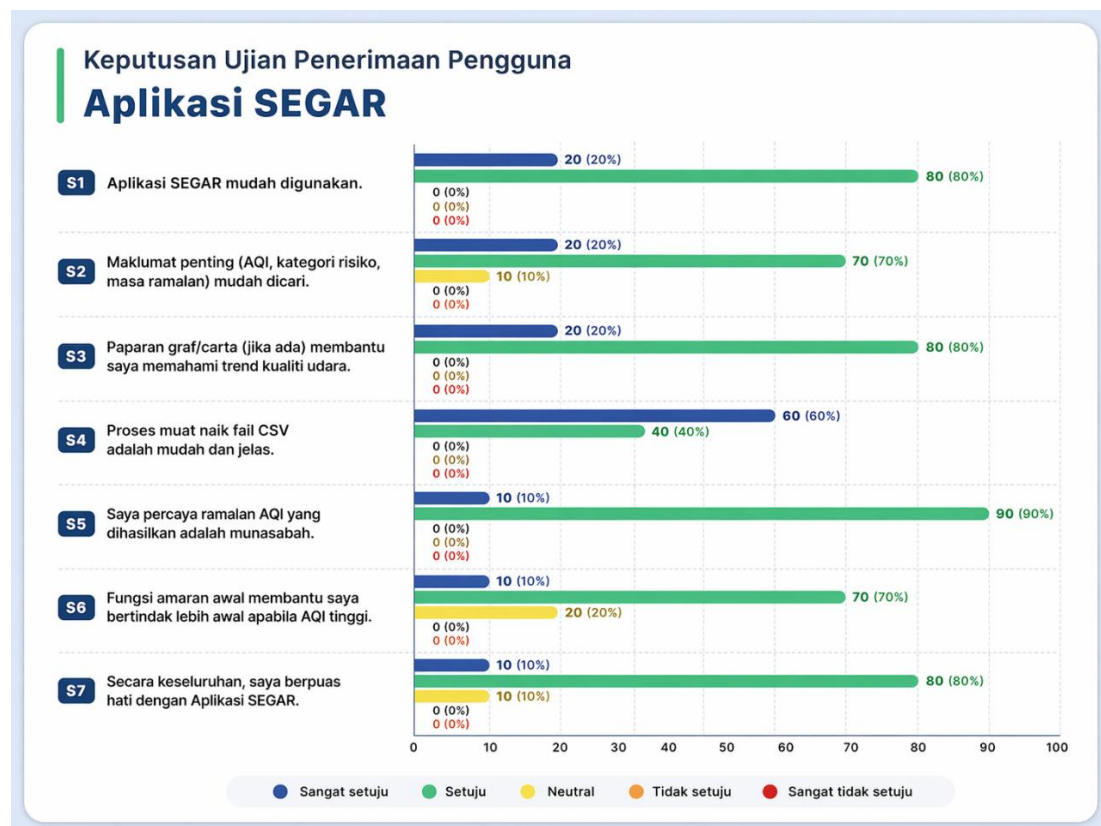
Antaramuka aplikasi SEGAR direka dengan memberikan tumpuan kepada kemudahan pengguna (user-centric design). Reka bentuk ini dipecahkan kepada beberapa komponen utama yang mengikut aliran kerja sistematik seperti yang ditunjukkan dalam Rajah 6:

- Langkah 1: Papan Pemuka (*Dashboard*) Utama: Berfungsi sebagai pusat navigasi utama yang memberikan gambaran keseluruhan fungsi aplikasi kepada pengguna.
- Langkah 2: Modul Input Data: Membolehkan pengguna memuat naik dataset kualiti udara dalam format CSV.
- Langkah 3: Analisis dan Ringkasan Dataset: Aplikasi secara automatik memaparkan analisis deskriptif bagi mengesahkan integriti data sebelum proses ramalan dijalankan.
- Langkah 4: Penjana Ramalan AQI dan Visualisasi Trend: Pengguna mengaktifkan enjin ramalan yang dikuasakan oleh algoritma RF untuk menghasilkan unjuran nilai AQI yang jitu. Selain itu, hasil akhir dipaparkan dalam bentuk graf interaktif bagi memudahkan interpretasi data merentasi masa.

6. IMPLEMENTASI DAN PENGUJIAN

Keberfungsian aplikasi SEGAR dinilai melalui satu siri ujian lapangan dan maklum balas daripada pengguna sasaran bagi memastikan sistem ini mencapai standard kebolegunaan yang ditetapkan. Keputusan ujian penerimaan pengguna (UAT) berdasarkan borang soal selidik yang melibatkan 10 orang responden diringkaskan dalam Rajah 7.

Dapatan kajian menunjukkan tahap penerimaan yang sangat positif terhadap sistem ini. Dari aspek kebolegunaan, 100% responden bersetuju dan sangat setuju bahawa reka bentuk aplikasi SEGAR adalah mesra pengguna dan mudah difahami. Selain itu, integrasi elemen visualisasi data terbukti berkesan apabila keseluruhan responden mengakui bahawa paparan graf sangat membantu dalam memahami trend kualiti udara secara intuitif.



Rajah 7. Maklumbalas Ujian Penerimaan Pengguna bagi Aplikasi SEGAR.

Dari sudut teknikal ramalan, majoriti responden (90%) menyatakan keyakinan terhadap ketepatan hasil ramalan AQI yang dijana oleh sistem, yang mengukuhkan kredibiliti model Random Forest (RF) yang digunakan. Walau bagaimanapun, fungsi Amaran Awal mencatatkan maklum balas neutral sebanyak 20%, yang memberikan isyarat penting untuk penambahbaikan mekanisme notifikasi pada masa hadapan agar lebih responsif terhadap keperluan pengguna. Secara keseluruhan, tahap kepuasan pengguna yang tinggi (90% setuju dan sangat setuju) membuktikan SEGAR merupakan instrumen yang praktikal bagi menyokong tindakan awal terhadap risiko pencemaran udara.

7. KESIMPULAN

Kajian ini telah berjaya merintis pembangunan Aplikasi SEGAR, sebuah platform berasaskan web yang mengintegrasikan kecerdasan buatan untuk pemantauan dan ramalan kualiti udara di Malaysia. Melalui perbandingan tiga algoritma pembelajaran mesin, model Random Forest dikenal pasti sebagai model paling unggul dengan nilai R^2 mencapai 0.9944, sekali gus membuktikan keupayaannya dalam menangani dinamik data kualiti udara yang kompleks dan bukan linear secara jitu. Pelaksanaan sistem ini bukan sahaja berjaya mendigitalkan proses pemantauan secara sistematik, malah berupaya mengurangkan kebergantungan kepada analisis manual yang memakan masa. Hasil ujian UAT mengesahkan bahawa SEGAR memenuhi kriteria utama pengguna dari segi kefungsiian dan kebolehgunaan. Walaupun kajian ini mencapai objektifnya, beberapa kekangan telah dikenal pasti, terutamanya dari aspek kebolehwakilan spatial yang terhad kepada stesen darat sahaja. Justeru, penyelidikan masa hadapan disarankan untuk mengintegrasikan data satelit seperti Aerosol Optical Depth (AOD) dan faktor meteorologi tambahan bagi memperkukuhkan liputan ramalan. Secara tuntasnya, SEGAR berpotensi besar untuk diadaptasi sebagai sistem sokongan keputusan yang berkesan oleh pihak berkuasa dan masyarakat umum ke arah pengurusan alam sekitar yang lebih mampan dan proaktif.

Penghargaan: Penulis ingin merakamkan setinggi-tinggi penghargaan dan ribuan terima kasih kepada semua pihak yang telah memberikan sokongan padu dalam menjayakan kajian ini, terutamanya kepada Fakulti Sains dan Teknologi Pertahanan (FSTP) dan Universiti Pertahanan Nasional Malaysia (UPNM).

Rujukan

- Abdul Rahman, A. S., Idris, A. Y., & Abdul Rahman, S. (2023). Evaluation of machine learning in predicting air quality index. *Mathematical Sciences and Informatics Journal (MIJ)*, 4(1), 1-10.
- Epa, U. (2018). Technical assistance document for the reporting of daily air quality–The Air Quality Index (AQI). *Environmental Protection Agency*.
- Karimi, B., & Samadi, S. (2024). Long-term exposure to air pollution on cardio-respiratory, and lung cancer mortality: a systematic review and meta-analysis. *Journal of Environmental Health Science and Engineering*, 22(1), 75-95.
- Latif, S. D., Almalayih, M., Yafouz, A., Ahmed, A. N., Zaini, N. a., Irwan, D., AlDahoul, N., Sherif, M., & El-Shafie, A. (2023). Prediction of atmospheric carbon monoxide concentration utilizing different machine learning algorithms: A case study in Kuala Lumpur, Malaysia. *Environmental Technology & Innovation*, 32, 103387.

- Li, Z.-H., Wang, X.-M., Liao, D.-Q., Zhang, Q., Chen, Z.-T., Qiu, C.-S., Tang, X.-L., Li, H.-M., Du, L.-Y., & Zhang, P.-D. (2024). Long-term air pollutants exposure and respiratory mortality: A large prospective cohort study. *Ecotoxicology and Environmental Safety*, 274, 116176.
- Méndez, M., Merayo, M. G., & Núñez, M. (2023). Machine learning algorithms to forecast air quality: a survey. *Artificial intelligence review*, 1.
- Mohd Puad, A. I. H., Zainol, Z., & Abdul Rauf, U. F. (2025). Analisis ramalan siri masa hujan menggunakan aplikasi ramalHujan. *Malaysia Journal of Invention and Innovation*, 4(2), 58-68.
- Neo, E. X., Hasikin, K., Mokhtar, M. I., Lai, K. W., Azizan, M. M., Razak, S. A., & Hizaddin, H. F. (2022). Towards integrated air pollution monitoring and health impact assessment using federated learning: a systematic review. *Frontiers in public health*, 10, 851553.
- Orellano, P., Kasdagli, M.-I., Velasco, R. P., & Samoli, E. (2024). Long-term exposure to particulate matter and mortality: an update of the WHO global air quality guidelines systematic review and meta-analysis. *International journal of public health*, 69, 1607683.
- U.S. EPA. *Communicating Air Quality Conditions: The Air Quality Index*. Available online at: <https://www.epa.gov/wildfire-smoke-course/communicating-air-quality-conditions-air-quality-index> (accessed Disember 13, 2025).
- Wang, B., Wang, B., Lv, B., & Wang, R. (2022). Impact of motor vehicle exhaust on the air quality of an urban city. *Aerosol and Air Quality Research*, 22(8), 220213.
- Yohanna, E., & San, L. H. (2026). Comparative analysis of machine learning models for predicting PM₁₀ and PM_{2.5} Concentrations in Pulau Pinang, Malaysia. *Air Quality, Atmosphere & Health*, 19(2), 21.
- Zaini, N. a. (2025). *Air quality forecasting and mapping in Malaysian urban areas: A hybrid deep learning approach*
- Zaini, N. a., Ean, L. W., Malek, M. A., Ahmed, A. N., & Chow, M. F. (2025). Forecasting of CO and O₃ concentrations using optimized deep learning: A case study for Kuala Lumpur, Malaysia. AIP Conference Proceedings,
- Zhang, X., Han, L., Wei, H., Tan, X., Zhou, W., Li, W., & Qian, Y. (2022). Linking urbanization and air quality together: A review and a perspective on the future sustainable urban development. *Journal of Cleaner Production*, 346, 130988.