

# PEMANFAATAN TEKNOLOGI MACHINE LEARNING DENGAN PENDEKATAN BAYESIAN UNTUK MENINGKATKAN POTABILITAS AIR DI INDONESIA

Jennifer Ardelia Limicia<sup>1)</sup>, Crysanthia Monica Lim<sup>1)</sup>, Cherylene Callista  
Reksohartono<sup>2)</sup>, Gladys Lionardi<sup>2)</sup>

<sup>1</sup>Data Science, School of Computer Science, Universitas Bina Nusantara,  
Indonesia

<sup>2</sup>Data Science, School of Computer Science, Universitas Bina Nusantara,  
Indonesia

Penulis korespondensi: [jennifer.limicia@binus.ac.id](mailto:jennifer.limicia@binus.ac.id)

## ABSTRAK

Air memegang peranan yang sangat penting dalam kehidupan manusia sebagai sumber daya alam yang sangat krusial yang dimana jika tidak dapat memenuhi standar keamanan dapat menimbulkan konsekuensi serius terhadap kesehatan masyarakat. Air yang tercemar memiliki potensi sebagai penyebab penyakit berbahaya. Tingginya jumlah kasus penyakit yang terkait dengan kualitas air yang rendah semakin menjadi perhatian serius. Situasi ini berkaitan erat dengan *Sustainable Development Goals* (SDG) nomor enam, yang menargetkan akses universal terhadap air bersih dan sanitasi. Penelitian ini mengadopsi metode *Bayesian Data Analysis* untuk menganalisis faktor-faktor yang mempengaruhi potabilitas air di Indonesia. Tujuan penelitian adalah untuk menyumbangkan pemahaman mendalam tentang hubungan antar variabel yang mempengaruhi potabilitas air (kualitas air). Dengan meningkatkan potabilitas air, diharapkan dapat berkontribusi pada pencapaian SDG nomor tiga, yaitu “*Good Health and Well-Being*”. Data yang digunakan merupakan dataset “*Water\_Potability*” dari *Kaggle* yang bersifat kuantitatif. Dataset dianalisis dengan model *logistic regression*, dengan menggunakan analisis *Bayesian*. Metode *Bayesian* adalah pendekatan di mana kombinasi antara keyakinan sebelumnya dan data yang diamati dapat memberi suatu keyakinan baru mengenai parameter model. Hasil dari penelitian ini menunjukkan bahwa model tidak dapat mendeskripsikan faktor apa saja yang mempengaruhi kualitas air.

Kata-kata kunci: pencemaran air, potabilitas air, *Sustainable Development Goals*, *logistic regression*, analisis *Bayesian*

## ABSTRACT

*Water plays a very important role in daily human life as a crucial natural resource which safety standards must be met lest there be serious consequences towards citizens' health. Contaminated water holds the potential to be the cause of dangerous disease. High numbers of diseases related to low water quality are increasingly becoming a serious concern. This situation is highly correlated to Sustainable Development Goals (SDG) number six, which targets universal access towards clean water and sanitation. This research adopts Bayesian Data Analysis*

*Method to analyze the factors that affect water potability in Indonesia. The purpose of this research is to contribute deep understanding about the relationship between variables which affects water potability (water quality). By increasing water potability, it's hoped that there's contribution to the achievement of SDG number three, that is "Good Health and Well-Being". The data used is the "Water\_Potability" dataset obtained from Kaggle which is quantitative. The dataset is analyzed using a logistic regression model with Bayesian analysis. The Bayesian Method is an approach where the combination from previous beliefs and current observed data can give a new belief regarding model parameters. The outcome of this research shows that the model couldn't describe the factors that affect water quality.*

*Keyword: water contamination, water potability, Sustainable Development Goals, logistic regression, Bayesian analysis*

## Pendahuluan

Air merupakan sumber daya alam yang paling krusial bagi kehidupan manusia. Penggunaan air tidak hanya terbatas pada kebutuhan minum, namun juga melibatkan aspek-aspek penting lainnya, seperti persiapan makanan, kebersihan diri, dan berbagai aktivitas sehari-hari (KSDA Jateng, 2024). Meskipun air memiliki peran krusial dalam kehidupan, isu potabilitas air di Indonesia menjadi perhatian serius. Potabilitas air merujuk pada kualitas air yang memenuhi standar kebersihan dan keamanan sehingga aman untuk dikonsumsi oleh manusia.

Sayangnya, situasi potabilitas air di Indonesia cenderung mengkhawatirkan. Dalam kehidupan sehari-hari, manusia secara terus-menerus bergantung pada air untuk berbagai keperluan, seperti minum, memasak, mandi, dan mencuci. Pentingnya air yang bersih dan aman tidak bisa dipandang sebelah mata, namun fakta menunjukkan bahwa kondisi potabilitas air di Indonesia belum mencapai standar yang diinginkan.

Kondisi potabilitas air yang belum mencapai standar di Indonesia secara langsung memiliki relevansi yang kuat dengan *Sustainable Development Goals* (SDG), terutama dengan SDG nomor 6, yaitu "*Clean Water and Sanitation*" atau "Air Bersih dan Sanitasi" (Gemeda dkk., 2021). SDG 6 bertujuan untuk memastikan akses universal terhadap air bersih dan sanitasi yang aman, serta meningkatkan keberlanjutan pengelolaan sumber daya air.

Ketidakmampuan memenuhi standar potabilitas air dapat berdampak serius pada kesehatan masyarakat. Air yang terkontaminasi dapat menjadi sumber penyakit berbahaya. Dampak negatif ini semakin memprihatinkan dengan meningkatnya jumlah kasus penyakit yang dapat dihubungkan dengan kualitas air yang rendah.

Dalam penelitian ini metode *Bayesian Data Analysis* digunakan untuk menganalisis faktor-faktor yang mempengaruhi standar keamanan air atau potabilitas air di Indonesia. Pendekatan ini akan membantu dalam memahami kompleksitas masalah dan memperoleh pemahaman yang lebih mendalam mengenai hubungan antara berbagai variabel yang mempengaruhi kualitas air. Karena itu, diharapkan penelitian ini dapat memberikan kontribusi positif dalam menyelesaikan permasalahan potabilitas air di Indonesia. Meningkatnya potabilitas air akan sekaligus meningkatkan kualitas kesehatan di Indonesia, yang akan berkontribusi pada pencapaian SDG 3 yaitu "*Good Health and Well-Being*" atau "Kesehatan dan Kesejahteraan yang Baik" (Binns dkk., 2022).

Dengan menggali informasi dari sumber-sumber pustaka terkini, penelitian ini akan menghadirkan pemahaman yang mutakhir mengenai potabilitas air dan faktor-

faktor yang mempengaruhinya di Indonesia. Melalui analisis *Bayesian*, diharapkan variabel-variabel kunci yang perlu diperhatikan dapat diidentifikasi dalam upaya meningkatkan kualitas dan keamanan air di negara ini.

### Metode

Penelitian ini penulis lakukan dengan menggunakan data sekunder yang berjudul “Water\_Potability” dari *Kaggle* mengenai potabilitas air dengan tujuan untuk membuat model *machine learning* yang dapat mengklasifikasikan air yang layak maupun tidak layak untuk diminum berdasarkan variabel-variabel yang ada. Berikut ini adalah variabel input yang digunakan dalam menentukan potabilitas air:

Tabel 1. Fitur-fitur dalam dataset beserta penjelasannya

Fitur	Penjelasan
ph	Indikator asam atau basa air (dalam rentang 0-14)
Hardness (mg/L)	Kapasitas air untuk mengendapkan sabun
Solids (ppm)	Jumlah zat padat yang terlarut
Chloramines (ppm)	Jumlah Chloramines (desinfektan untuk mengolah air minum, terbentuk ketika amonia dicampur dengan klorin)
Sulfate (mg/L)	Sulfat yang terlarut
Conductivity ( $\mu\text{S}/\text{cm}$ )	Konduktivitas listrik air
Organic_carbon (ppm)	Jumlah karbon organik (terbentuk dari bahan organik alami yang membusuk dan sumber sintetis)
Trihalomethanes ( $\mu\text{g}/\text{L}$ )	Jumlah Trihalomethanes – bahan kimia yang ditemukan dalam air yang diolah dengan klorin
Turbidity (NTU)	Ukuran sifat pemancaran cahaya air
Potability	Indikator untuk menentukan apakah air aman untuk dikonsumsi (1 untuk iya dan 0 untuk tidak)

Dari data yang penulis punya, pertama-tama penulis melakukan proses *data cleaning*. Data cleaning merupakan metode untuk mengatasi ketidaklengkapan data (*missing value*) melalui berbagai pendekatan, seperti menghilangkan duplikasi data, memeriksa inkonsistensi data, mengisi atau menghapus nilai data yang kosong, dan lain sebagainya (Azmi dkk., 2022). Dalam penelitian ini, proses *data cleaning* penulis lakukan dengan cara meng-imputasi *missing value* menggunakan median. Hal ini dikarenakan imputasi menggunakan nilai median lebih konsisten dan tidak bias ketika data tidak berdistribusi normal (Purwanto, 2022). Reduksi dimensi tidak dilakukan karena dimensi dari data yang digunakan tidak terlalu besar. Oleh karena itu, dinilai bahwa tidak diperlukan adanya reduksi dimensi.

Langkah selanjutnya dalam penelitian penulis adalah membuat model klasifikasi. Dalam penelitian ini, model yang penulis buat menggunakan pendekatan *Bayesian* dengan 3 model *logistic regression*, dengan rumus :

$$\text{logit}(q_i) = \eta_i = \sum_{j=1}^J X_{ij}\beta_j$$

Rumus 1. Rumus Logistic Regression dengan Pendekatan Bayesian

*Logistic regression* sendiri adalah suatu model yang mencerminkan korelasi antara variabel respons dan variabel prediktor, di mana variabel respons bersifat biner. Variabel biner ini memiliki dua kemungkinan nilai yang dapat diwakili oleh 0 (gagal) dan 1 (sukses). Penggunaan regresi logistik dapat dilakukan untuk meramalkan atau mengklasifikasikan, sambil juga mengidentifikasi faktor-faktor yang memiliki dampak signifikan terhadap variabel respons. Sedangkan, *logistic regression* dengan menggunakan pendekatan *Bayesian* akan menghasilkan estimasi parameter dengan mengkombinasikan *likelihood* dari data sampel dengan distribusi *prior*, dan hasilnya disebut sebagai distribusi *posterior* (Aryani dkk., 2023).

Dalam penelitian ini, distribusi posterior untuk semua model diperoleh dengan menggunakan algoritma *Markov Chain Monte Carlo* (MCMC) berdasarkan 6.000 iterasi dengan 5.000 MCMC sampel setelah burn-in 1.000 sampel. Kekonvergenan dari setiap parameter model dicek dengan menggunakan metode Geweke dan Gelman-Rubin. Terakhir, ketiga model tersebut akan penulis bandingkan performanya untuk mencari model terbaik dengan menggunakan metode DIC (*Deviance Information Criteria*), WAIC (*Watanabe-Akaike Information Criteria*), serta *posterior-predictive p-value* (PPP).

Hasil dari 3 model tersebut memberikan prediksi antara 2 pilihan *output* yang bersifat *binary*, di mana 0 memiliki arti bahwa air berkualitas buruk dan tidak layak untuk dikonsumsi, sedangkan 1 berarti air berkualitas baik dan layak untuk dikonsumsi. Pada model pertama, penulis menggunakan distribusi *Normal* sebagai distribusi *prior* dengan rata - rata 0 dan varians 0.01. Untuk *likelihood* yang digunakan adalah distribusi *Bernoulli*. Setelah itu, dijalankan model kedua yang menggunakan seleksi fitur dari SSVS (*Stochastic Search Variable Selection*) dengan *prior* yang menggunakan distribusi *Bernoulli* dengan probabilitas 0.5 (untuk parameter  $\gamma$ ) dan *Normal* (untuk parameter  $\delta$ ) dengan rata - rata 0 dan varians 1 untuk memilih variabel mana saja yang mempengaruhi potabilitas dari air secara signifikan. Kemudian model ketiga menggunakan distribusi *Normal* dengan rata - rata 0 dan varians 1 sebagai *prior* dan distribusi *Bernoulli* sebagai *likelihood*.

### Hasil dan Pembahasan

Ketiga model dijalankan menggunakan data yang sama. Karena model kedua hanya digunakan untuk melihat variabel yang memiliki pengaruh signifikan terhadap potabilitas air, tidak dilakukan estimasi kovariat beta pada model tersebut.

Tabel 1. Estimasi kovariat beta dari model 1 dan 3

Kovariat Beta	Model 1			Model 3		
	Rata-rata	Q 2.5%	Q 97.5%	Rata-rata	Q 2.5%	Q 97.5%
$\beta$ ph	0.00622	-0.06369	0.07697	0.00624	-0.06469	0.07795
$\beta$ Hardness	-0.02819	-0.09955	0.04191	-0.02729	-0.09693	0.04358
$\beta$ Solids	0.06869	-0.00258	0.14024	0.06759	-0.00394	0.13960
$\beta$ Chloramines	0.05259	-0.01673	0.12220	0.05221	-0.01965	0.12344
$\beta$ Sulfate	-0.03412	-0.10648	0.03871	-0.03442	-0.10646	0.03756
$\beta$ Organic_carbon	-0.01743	-0.08757	0.05454	-0.01587	-0.08596	0.05501
$\beta$ Trihalomethanes	-0.06016	-0.13268	0.01188	-0.05999	-0.13135	0.01265
$\beta$ Turbidity	0.0111	-0.06163	0.08184	0.01224	-0.05806	0.08042
$\beta$ Potability	0.00008	-0.06852	0.07169	-0.00003	-0.07038	0.07146

Nilai pada kolom rata-rata merupakan estimasi kovariat beta untuk masing-masing model. Besaran hasil estimasi beta berpengaruh pada *output*, dimana bila estimasi beta naik maka nilai *output* akan naik juga. Berbeda dengan itu, nilai negatif pada estimasi beta berarti bahwa semakin besar nilai estimasi, semakin kecil nilai *output*. Selain itu, menggunakan 95% *confidence interval*, kuantil beta yang mengandung nilai 0 dalam 95% jangkauannya termasuk tidak signifikan. Berdasarkan data yang diperoleh, semua variabel dapat dikatakan tidak signifikan.

Data juga dimasukkan ke dalam model kedua, dimana signifikansi setiap variabel akan dikalkulasi dengan metode SSVS.

Tabel 2. Signifikansi variabel berdasarkan metode SSVS

Kovariat Beta	Inc_prob
$\beta$ ph	0.0414
$\beta$ Hardness	0.0452
$\beta$ Solids	0.1834
$\beta$ Chloramines	0.0918
$\beta$ Sulfate	0.0606
$\beta$ Organic_carbon	0.0394
$\beta$ Trihalomethanes	0.1489
$\beta$ Turbidity	0.0414
$\beta$ Potability	0.0365

Pada tabel di atas terlihat bahwa semua variabel memiliki nilai “Inc\_prob” yang sangat kecil, sehingga dapat dikatakan bahwa semua variabel tidak signifikan.

Kemudian konvergensi diperiksa untuk ketiga model, menggunakan Geweke ( $|value| < 2$  menunjukkan konvergensi) dan Gelman-Rubin (1 menunjukkan konvergensi). Dengan Geweke, pada model pertama nilai mutlak semua variabel lebih kecil daripada 2, yang berarti model sudah bersifat konvergen. Pada model ketiga, sama seperti model pertama, nilai mutlak semua variabel di bawah 2. Jadi dalam sisi konvergensi berdasarkan Geweke, semua model konvergen. Hasil dari Gelman-Rubin untuk model pertama dan ketiga semua bernilai 1, sehingga dapat dikatakan konvergen.

*Autocorrelation* dihitung untuk memeriksa apakah per variabelnya bersifat independen dan tidak saling berkorelasi. Hasil autokorelasi pada kedua *chain* dalam masing-masing model sangatlah kecil, dengan nilai autokorelasi mutlak terbesar 0.05 di model pertama, 0.036 di model ketiga, sehingga dapat dibuktikan bahwa tidak ada ketergantungan antar variabel.

Performa dari model pertama dan ketiga dibandingkan dengan *Deviance Information Criteria* (DIC), *Watanabe-Akaike Information Criteria* (WAIC), serta *posterior-predictive p-value* (PPP). Tabel berikut merupakan tabel perbandingan ketiga model tersebut.

Tabel 3. Perbandingan performa model satu dan tiga berdasarkan DIC dan WAIC

	<b>Model 1</b>	<b>Model 3</b>
<b>DIC</b>	4391	4391
<b>WAIC</b>	4391.713	4391.783
<b>PPP</b>	0.487	0.4938

Dapat dilihat pada tabel di atas, kedua model memiliki nilai DIC (*penalized deviance*) dan WAIC yang hampir sama. Model 1 memiliki nilai WAIC yang lebih rendah. Semakin rendah nilai WAIC, semakin baik kualitas model tersebut. *Posterior predictive p-value* kedua model ini mendekati 0.5, berarti hasil prediksi model-model tersebut menyerupai data asli dan tidak bias (terlalu tinggi/rendah). Dari kedua model, model ke-3 memiliki *posterior predictive p-value* terbaik.

### Kesimpulan

Berdasarkan pemodelan data yang dibuat telah dibuktikan bahwa model sudah konvergen dan tidak ada autokorelasi atau sifat saling ketergantungan antar variabel. Meski demikian, didapat bahwa tidak ada satu variabel pun yang signifikan terhadap potabilitas air, baik melalui perhitungan SSVS, *confidence interval*, maupun hasil estimasi beta tiap variabel. Terdapat beberapa hal yang patut dipertimbangkan dalam hal ini, seperti pemilihan *prior* yang kurang tepat atau pemilihan dataset yang kurang tepat.

Saran untuk penelitian selanjutnya yang berkaitan dengan penelitian ini yaitu untuk mencari *prior* yang lebih tepat dengan tinjauan pustaka yang lebih luas. Dengan demikian diharapkan dapat ditemukan faktor yang mempengaruhi potabilitas air serta meningkatkan kualitas dan keamanan air di negara ini.

### **Ucapan Terima Kasih**

Terima kasih kepada Universitas Bina Nusantara yang telah memberi kesempatan atas penyusunan dan penyelesaian kegiatan ilmiah ini.

### **Kontribusi Penulis**

Penulis Satu yaitu Jennifer Ardelia Limicia berkontribusi dalam melakukan pengembangan model kedua dan ketiga, menjalankan percobaan serta dalam penulisan artikel; Penulis Dua yaitu Crysanthia Monica Lim berkontribusi dalam pengembangan model pertama dan penulisan artikel; Penulis Tiga yaitu Cherylene Callista Reksohartono berkontribusi dalam penulisan artikel; Penulis Terakhir yaitu Gladys Lionardi berkontribusi dalam pengembangan model pertama, menjalankan percobaan dan penulisan artikel; Dosen Pendamping yaitu Karli Eka Setiawan berkontribusi dalam memberi arahan dan masukan mengenai kegiatan ilmiah yang dilaksanakan.

### **DAFTAR PUSTAKA**

- Azmi, B., Hermawan, A. dan Avianto, D. 2022. Analisis Pengaruh PCA Pada Klasifikasi Kualitas Air Menggunakan Algoritma K-Nearest Neighbor dan Logistic Regression. *Jurnal Sistem dan Teknologi Informasi*. 7 (2):94-103.
- Bain, R., Johnston, R., Khan, S., Hancioglu, A. dan Slaymaker, T. 2021. Monitoring Drinking Water Quality in Nationally Representative Household Surveys in Low- and Middle-Income Countries: Cross-Sectional Analysis of 27 Multiple Indicator Cluster Surveys 2014–2020. *Environmental Health Perspectives*. 129 (9). doi:<https://doi.org/10.1289/ehp8459>
- Cronin, A.A., Odagiri, M., Arsyad, B., Nuryetty, M.T., Amannullah, G., Santoso, H., Darundiyah, K. dan Nasution, N. 'Aisyah. 2017. Piloting water quality testing coupled with a national socioeconomic survey in Yogyakarta province, Indonesia, towards tracking of Sustainable Development Goal 6. *International Journal of Hygiene and Environmental Health*. 220 (7):1141–1151. doi:<https://doi.org/10.1016/j.ijheh.2017.07.00>
- Djana, M. 2023. Analisis Kualitas Air Dalam Pemenuhan Kebutuhan Air Bersih di Kecamatan Natar Hajimena Lampung Selatan. 8(1).
- Envihsafkm. 2021. *Krisis Air Bersih – Envihsa FKM UI 2021*. URL: <https://envihsa.fkm.ui.ac.id/2021/09/30/krisis-air-bersih/>. Diakses tanggal 17 Februari 2024.



- Gemeda, S.T., Springer, E., Gari, S.R., Birhan, S.M. dan Bedane, H.T. (2021). The importance of water quality in classifying basic water services: The case of Ethiopia, SDG6.1, and safe drinking water. *PLOS ONE*, 16(8), p.e0248944. doi:<https://doi.org/10.1371/journal.pone.0248944.1>.
- Jateng, R.K. 2024. *Sumber Daya Alam Air*. [online] KSDA Jateng. URL: <https://ksdajateng.id/blog/sumber-daya-alam-air/>. Diakses tanggal 18 Februari 2024.
- Media, K.C. 2021. *Alasan Air Penting bagi Kehidupan Manusia dan Lingkungan*. [online] KOMPAS.com. URL: <https://www.kompas.com/skola/read/2021/10/04/153000469/alasan-air-penting-bagi-kehidupan-manusia-dan-lingkungan>. Diakses tanggal 18 Februari 2024.
- Novriadhy, D. 2021. Kualitas Air Bersih dan Potensi Dampaknya Terhadap Kesehatan Masyarakat di Kabupaten Muara Enim. *Ekologi Kesehatan*, 19(1), pp.35–44. doi:<https://doi.org/10.22435/jek.v19i1.2360>.
- Purwanto, A. 2019. Studi Simulasi : Perbandingan Metode Imputasi Mean Substitution dan Median Substitution pada Person Trait Estimation. *Skripsi*. Fakultas Ilmu Sosial dan Ilmu Politik Universitas Brawijaya, Malang.
- Riyantoko, P. A., Fahrudin, T. M., dan Hindrayani, K. M. 2021. Analisis Sederhana Pada Kualitas Air Minum Berdasarkan Akurasi Model Klasifikasi Dengan Menggunakan Lucifer Machine Learning. *Seminar Nasional Sains Data 2021 (SENADA 2021)*. 2021, Jawa Timur. pp.12-18.
- Sari, Y.V., Muallifah, Z., dan Fanani, A. 2023. Klasifikasi Kualitas Air Menggunakan Metode Extreme Learning Machine (ELM). *Jupiter*, 15(2):983–994.
- Setioningrum, R.N.K., Sulistyorini, L., dan Rahayu, W.I. 2019. Gambaran Kualitas Air Bersih Kawasan Domestik di Jawa Timur Pada Tahun 2019. *Ikesma*, 16(2).
- Savitri, L. dan Nursalim, R. 2023. Klasifikasi Kualitas Air Minum menggunakan Penerapan Algoritma Machine Learning dengan Pendekatan Supervised Learning. *Diophantine Journal of Mathematics and Its Applications*. 2(1):29-36.
- Transitioning to Good Health and Well-Being. (t.t.). URL: <https://library.oapen.org/bitstream/handle/20.500.12657/59127/1/%5BHBK5%20Final%20Binder%20196%20pgs.pdf#page=129>. Diakses tanggal 20 November 2023.
- Utomo, M dan Suliadi. 2021. Penerapan Metode Bayesian pada Analisis Regresi untuk Menguji Pengaruh Kepadatan Penduduk dan Aktivitas Ekonomi terhadap Kasus Terkonfirmasi Covid-19 di Indonesia. *Prosiding Statistika*. Bandung, Indonesia. pp.684-691.
- Wijaya, R. 2019. Pengukuran Kualitas Air Terhadap Status Kesehatan Masyarakat di Kelurahan Jembatan Lima Kecamatan Tambora Jakarta Barat. *Skripsi*.

Program Studi Ilmu Kesehatan Masyarakat Sekolah Tinggi Ilmu Kesehatan  
Indonesia Maju, Jakarta.

## Lampiran 1. Biodata Ketua dan Anggota, serta Dosen Pendamping

### Biodata Ketua

#### A. Identitas Diri

1	Nama Lengkap	Jennifer Ardelia Limicia
2	Jenis Kelamin	Perempuan
3	Program Studi	Data Science
4	NIM	2602105090
5	Tempat dan Tanggal Lahir	Pontianak, 30 Agustus 2004
6	Alamat Email	jennifer.limicia@binus.ac.id
7	Nomor Telepon/HP	0895613280818

#### B. Kegiatan Kemahasiswaan Yang Sedang/Pernah Diikuti

No	Jenis Kegiatan	Status dalam Kegiatan	Waktu dan Tempat
1	Data Science Club	Aktif	2023-sekarang, Universitas Bina Nusantara
2	Associate Member IT Division	Aktif	2023-2024, Universitas Bina Nusantara
3			

#### C. Penghargaan Yang Pernah Diterima

No.	Jenis Penghargaan	Pihak Pemberi Penghargaan	Tahun
1			

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan PKM-AI.

Jakarta, 18-02 -2024

Ketua Tim



(Jennifer Ardelia Limicia)

## Biodata Anggota 1

## A. Identitas Diri

1	Nama Lengkap	Crysantha Monica Lim
2	Jenis Kelamin	Perempuan
3	Program Studi	Data Science
4	NIM	2602090076
5	Tempat dan Tanggal Lahir	Batam, 23 Juli 2004
6	Alamat Email	crysantha.lim@binus.ac.id
7	Nomor Telepon/HP	089624742647

## B. Kegiatan Kemahasiswaan Yang Sedang/Pernah Diikuti

No	Jenis Kegiatan	Status dalam Kegiatan	Waktu dan Tempat
1	HIMTI	Aktif	2024-2025, Universitas Bina Nusantara
2	Associate Member IT Division	Aktif	2024-2025, Universitas Bina Nusantara
3	KMK	Aktif	2022 – sekarang, Universitas Bina Nusantara

## C. Penghargaan Yang Pernah Diterima

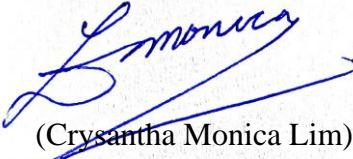
No.	Jenis Penghargaan	Pihak Pemberi Penghargaan	Tahun
1	Beasiswa Best School Scholarship	Universitas Bina Nusantara	2022 - sekarang

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan PKM-AI.

Jakarta, 18-02 -2024

Anggota Tim 1



(Crysantha Monica Lim)

## Biodata Anggota 2

## A. Identitas Diri

1	Nama Lengkap	Cherylene Callista Reksohartono
2	Jenis Kelamin	Perempuan
3	Program Studi	Data Science
4	NIM	2602087024
5	Tempat dan Tanggal Lahir	Surabaya, 23 Oktober 2004
6	Alamat Email	cherylene.reksohartono@ binus.ac.id
7	Nomor Telepon/HP	085334755789

## B. Kegiatan Kemahasiswaan Yang Sedang/Pernah Diikuti

No	Jenis Kegiatan	Status dalam Kegiatan	Waktu dan Tempat
1	Himpunan Teknik Informatika	Aktif	2022 - sekarang, Universitas Bina Nusantara

## C. Penghargaan Yang Pernah Diterima

No.	Jenis Penghargaan	Pihak Pemberi Penghargaan	Tahun
1			

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan PKM-AI.

Jakarta, 18-02 -2024

Anggota Tim 2

(Cherylene Callista Reksohartono)

### Biodata Anggota 3

#### A. Identitas Diri

1	Nama Lengkap	Gladys Lionardi
2	Jenis Kelamin	Perempuan
3	Program Studi	Data Science
4	NIM	2602073076
5	Tempat dan Tanggal Lahir	Jakarta, 20 Desember 2004
6	Alamat Email	gladys.lionardi@binus.ac.id
7	Nomor Telepon/HP	082112609473

#### B. Kegiatan Kemahasiswaan Yang Sedang/Pernah Diikuti

No	Jenis Kegiatan	Status dalam Kegiatan	Waktu dan Tempat
1	Binus Mandarin Club	Aktif	2022 – sekarang, Universitas Bina Nusantara

#### C. Penghargaan Yang Pernah Diterima

No.	Jenis Penghargaan	Pihak Pemberi Penghargaan	Tahun

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan PKM-AI.

Jakarta, 18-02 -2024  
Anggota Tim 3

(Gladys Lionardi)

### Biodata Dosen Pendamping

## A. Identitas Diri

1	Nama Lengkap (dengan gelar)	Karli Eka Setiawan., S.Si., M.Kom.
2	Jenis Kelamin	Laki-laki
3	Program Studi	Data Science
4	NIP/NIDN	0323119304
5	Tempat dan Tanggal Lahir	Tangerang, 23 November 1993
6	Alamat Email	karli.setiawan@binus.ac.id
7	Nomor Telepon/HP	082213685668

## B. Riwayat Pendidikan

No	Jenjang	Bidang Ilmu	Institusi	Tahun Lulus
1	Sarjana (S1)	Matematika	Universitas Padjadjaran	2015
2	Magister (S2)	Computer Science	Universitas Bina Nusantara	2023

## C. Rekam Jejak Tri Dharma PT

## Pendidikan/Pengajaran

No	Nama Mata Kuliah	Wajib/Pilihan	sks
1.	Machine Learning	Wajib	3
2.	Data Security	Wajib	2
3.	Research Methodology in Computer Science	Wajib	2
4.	Data Science and Applications	Wajib	4
5.	Computer Vision	Wajib	2
6.	Big Data Infrastructure and Technology	Wajib	2
7.	Scientific Computing	Wajib	2

## Penelitian

No	Judul Penelitian	Penyandang Dana	Tahun
1.			
2.			

## KPengabdian Kepada Masyarakat

No	Judul Pengabdian kepada Masyarakat	Penyandang Dana	Tahun
1.			
2.			

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi. Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan PKM-AI.

Jakarta, 18-02 -2024  
Dosen Pendamping



(Karli Eka Setiawan., S.Si., M.Kom.)



**Lampiran 2. Kontribusi Ketua, Anggota, dan Dosen Pendamping**

<b>No</b>	<b>Nama</b>	<b>Posisi penulis</b>	<b>Bidang Ilmu</b>	<b>Kontribusi</b>
1	Jennifer Ardelia Limicia	Ketua	Data Science	Melakukan pengembangan model kedua dan ketiga, menjalankan percobaan serta penyusunan metode.
2	Crysantha Monica Lim	Anggota	Data Science	Melakukan pengembangan model pertama dan penyusunan pendahuluan dan abstrak.
3	Cherylene Callista Reksohartono	Anggota	Data Science	Penyusunan pendahuluan, abstrak, dan daftar pustaka.
4	Gladys Lionardi	Anggota	Data Science	Melakukan pengembangan model pertama, percobaan, dan penyusunan hasil dan pembahasan.
5	Karli Eka Setiawan	Dosen Pendamping	Data Science	Memberi arahan dan masukan mengenai kegiatan ilmiah yang dilaksanakan.

### Lampiran 3. Surat Pernyataan Ketua Pengusul

#### SURAT PERNYATAAN KETUA TIM PENGUSUL

Yang bertanda tangan di bawah ini :

Nama Ketua Tim	:	Jennifer Ardelia Limicia
Nomor Induk Mahasiswa	:	2602105090
Program Studi	:	Data Science
Nama Dosen Pendamping	:	Karli Eka Setiawan., S.Si., M.Kom.
Perguruan Tinggi	:	Universitas Bina Nusantara

Dengan ini menyatakan bahwa proposal PKM-AI saya dengan judul:

"Pemanfaatan Teknologi Machine Learning dengan Pendekatan Bayesian untuk Meningkatkan Potabilitas Air di Indonesia" yang diusulkan untuk tahun anggaran 2024 adalah:

1. Asli karya kami, belum pernah dibiayai oleh lembaga atau sumber dana lain
2. Tidak dibuat dengan menggunakan kecerdasan buatan/artificial intelligence (AI).

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku dan mengembalikan seluruh biaya yang sudah diterima ke kas Negara.

Demikian pernyataan ini dibuat dengan sesungguhnya dan sebenar – benarnya.

Jakarta, 18-02 -2024

Yang menyatakan,



(Jennifer Ardelia Limicia)

2602105090.

**Lampiran 4. Surat Pernyataan Sumber Tulisan****SURAT PERNYATAAN SUMBER TULISAN PKM-AI**

---

Saya yang menandatangani Surat Pernyataan ini:

Nama Ketua Tim	:	Jennifer Ardelia Limicia
Nomor Induk Mahasiswa	:	2602105090
Program Studi	:	Data Science
Nama Dosen Pendamping	:	Karli Eka Setiawan., S.Si., M.Kom.
Perguruan Tinggi	:	Universitas Bina Nusantara

1. Menyatakan bahwa PKM-AI yang saya tuliskan bersama anggota tim lainnya benar bersumber dari kegiatan yang telah dilakukan:
  - a. Sumber tulisan dari hasil kegiatan yang telah dilakukan berkelompok oleh tim penulis, yaitu: Kegiatan Ilmiah Pemodelan Data
  - b. Topik Kegiatan: Prediksi Kualitas Air dengan Pendekatan Machine Learning
  - c. Tahun dan Tempat Pelaksanaan: 2024, Jakarta Barat.
2. Naskah ini belum pernah diterbitkan/dipublikasikan dalam bentuk prosiding maupun jurnal sebelumnya dan diikuti dalam kompetisi.
3. Kami menyatakan kesediaan artikel ilmiah ini ditampilkan pada laman simbelmawa.

Demikian Surat Pernyataan ini dibuat dengan penuh kesadaran tanpa paksaan pihak manapun juga untuk dapat digunakan sebagaimana mestinya.

Jakarta, 18-02 -2024

Yang menyatakan,

(Jennifer Ardelia Limicia)  
2602105090.

**Lampiran 5. Hasil Pengecekan Plagiasi dengan Indeks Similaritas  
Maksimum 25%**