



Implementasi Metode CRISP-DM dalam Analisis Dampak Produksi Sampah dan Limbah terhadap Kualitas Lingkungan Hidup serta Kaitannya dengan *Sustainability Development Goals* (SDGs)

Ratna Hapsari Wulandari^{1*}, Andy Dwi Bayu Bawono²

^{1,2} Fakultas Ekonomi dan Bisnis, Universitas Muhammadiyah Surakarta

* E-mail Korespondensi: ratnahpsr17@gmail.com

Information Article

History Article

Submission: 01-12-2024

Revision: 12-12-2024

Published: 14-12-2024

DOI Article:

10.24905/permana.v16i2.550

ABSTRAK

Penelitian ini mengkaji implementasi metode CRISP-DM dalam analisis dampak produksi sampah dan limbah terhadap kualitas lingkungan hidup di Indonesia, serta hubungannya dengan Tujuan Pembangunan Berkelanjutan atau SDGs. Metode CRISP-DM berfokus pada enam fase. Data yang digunakan berasal dari Badan Pusat Statistik (BPS), mencakup indeks kualitas udara (IKU), indeks kualitas air (IKA), dan produksi sampah. Metode yang digunakan dalam penelitian ini yaitu analisis deskriptif Model regresi logistik digunakan dalam analisis ini, dengan akurasi sebesar 86,36%. Hasil analisis menunjukkan bahwa terdapat hubungan negatif antara jumlah produksi sampah dengan kualitas udara dan air, yang berpotensi mengganggu pencapaian SDGs, terutama terkait air bersih dan sanitasi, serta ekosistem berkelanjutan.

Kata Kunci: CRISP-DM, Produksi Sampah, Indeks Kualitas Udara, Indeks Kualitas Air, SDGs

ABSTRACT

This study examines the implementation of the CRISP-DM method in analyzing the impact of waste and sewage production on environmental quality in Indonesia, as well as its relationship with the Sustainable Development Goals or SDGs. The CRISP-DM method focuses on six phases. The data used comes from the Central Statistics Agency (BPS), including the air quality index (IKU), water quality index (IKA), and waste production. The method used in this study is descriptive analysis. The logistic regression model is used in this analysis, with an accuracy of 86.36%. The results of the analysis show that there is a negative relationship between the amount of waste production and air and water quality, which has the potential to disrupt the achievement of SDGs, especially related to clean water and sanitation, as well as sustainable ecosystems.

Acknowledgment

Key word: CRISP-DM, Waste Production, Air Quality

PENDAHULUAN

Data analyst merupakan suatu profesi yang bertanggung jawab untuk mengolah data menggunakan sejumlah alat kemudian menerjemahkannya ke dalam bentuk informasi yang mudah dipahami oleh semua orang. Implementasi keahlian seorang *data analyst* tidak terbatas pada bisnis saja, tetapi dalam konteks isu lingkungan menjadi sangat krusial dan relevan pada era ini. Seorang *data analyst* bertugas untuk mengumpulkan, mengolah, dan menganalisis data guna menemukan tren, pola, dan wawasan yang relevan. Beberapa alat yang digunakan oleh seorang *data analyst* yaitu seperti *Excel*, *SQL*, *Python* dapat digunakan untuk mengelompokkan dan mengeksplorasi data, misalnya data mengenai kualitas air, udara atau perubahan iklim. Selain itu, perangkat lunak visualisasi data seperti *Tableau* atau *Power BI* dapat digunakan untuk menyajikan hasil analisis yang jelas dan mudah dipahami yang memungkinkan untuk pengambilan keputusan. Dengan demikian, peran seorang *data analyst* tidak hanya terbatas pada bisnis, tetapi juga memiliki dampak yang signifikan dalam mendukung tindakan berkelanjutan untuk perlindungan lingkungan.

Meyrena & Amelia (2020) menyebutkan bahwa minimnya peran serta masyarakat dalam penanganan serta pengelolaan sampah menimbulkan dampak yang cukup besar terhadap keseimbangan lingkungan. Masyarakat dan pemerintah memiliki peran penting sebagai kekuatan yang mampu mengatasi masalah sampah, salah satunya melalui inovasi teknologi tepat guna yang memanfaatkan limbah plastik untuk produksi *ecopaving* atau *paving block plastic*.

Pertumbuhan keanekaragaman industri yang berkelanjutan, menyebabkan limbah bahan berbahaya dan beracun (B3) turut meningkat di seluruh dunia. Pertumbuhan industri ini tentu membawa dampak negatif, salah satunya adalah meningkatnya jumlah limbah yang dihasilkan. Penerapan pengelolaan limbah B3 sesuai dengan standar lingkungan menjadi tantangan mendesak bagi para penghasil limbah, sekaligus penting untuk mendukung upaya perlindungan lingkungan dan kesehatan (Nursabrina et al., 2021).

Shahzad et al., (2022) menyimpulkan bahwa dibutuhkan adopsi inovasi ramah lingkungan dalam lingkungan bisnis yang penuh tantangan saat ini, khususnya dalam industry ma-

nufaktur. Selain itu, penelitian ini menggarisbawahi peran penting inovasi bisnis ramah lingkungan dan ukuran organisasi dalam memoderasi inisiatif ramah lingkungan di berbagai skala perusahaan. Penelitian ini juga menekankan pentingnya mengintegrasikan praktik ramah lingkungan dan teknologi inovatif untuk mencapai tujuan keberlanjutan secara efektif.

Yurnalisdell, (2023) menyimpulkan bahwa Pengelolaan limbah bahan berbahaya dan beracun (B3), mulai dari tahap pengumpulan, pemisahan, penyimpanan, pengangkutan, pengolahan, hingga penimbunan, harus dilakukan dengan benar dan aman sesuai dengan peraturan perundang-undangan yang berlaku. Langkah ini bertujuan untuk meminimalkan dampak negatif yang dapat timbul akibat pengelolaan limbah berbahaya dan beracun yang tidak sesuai prosedur.

Oleh sebab itu, saran untuk *Healthy, Safety, and Environment* (HSE) perlu memperhatikan dan meningkatkan system pengelolaan limbah B3 khususnya mengenai kemasan yang digunakan sebagai wadah limbah B3 dan penyimpanan sementara dari masing-masing divisi. Kemudian disarankan adanya pemeriksaan rutin pada penyimpanan sementara agar sesuai dengan peraturan yang berlaku (Nurlina, 2021) .

Berdasarkan penelitian terdahulu yang telah dilakukan, adanya krisis keadaan lingkungan mendorong adanya semangat teknologi dalam mengakselerasi upaya melestarikan lingkungan di Indonesia. Adanya pertumbuhan industri yang berkembang pesat menyebabkan peningkatan produksi sampah dan limbah secara signifikan di Indonesia. Produksi sampah dan limbah yang semakin besar memberikan dampak serius terhadap lingkungan hidup seperti penurunan kualitas udara, terganggunya kesehatan masyarakat, menurunnya nilai estetika, kerugian ekonomi, dan terganggunya ekosistem alami (Permadi & Murni, 2013). Pencemaran udara dan air merupakan salah satu dampak dari produksi sampah dan limbah yang tidak dikelola dengan baik. Sebagai contoh, pembakaran sampah dapat menghasilkan gas beracun yang mencemari udara, sementara pembuangan limbah kimia yang tidak diolah terlebih dahulu dapat mengkontaminasi sumber air yang berdampak buruk pada kualitas lingkungan hidup. Selain itu, paparan zat berbahaya dari limbah dapat menyebabkan berbagai penyakit seperti iritasi pada mata dan kulit, gangguan pernapasan, kanker, paru-paru, dan kematian (Hayuning & Sudarmaji, 2016). Oleh karena itu, diperlukan tindakan yang tepat dalam mengelola sampah dan limbah agar terjaganya kualitas hidup yang berkelanjutan di Indonesia.

Salah satu langkah penting dalam mencapai hal ini adalah melalui penerapan *sustainability reporting*. *Sustainability reporting* atau laporan keberlanjutan adalah proses di mana perusahaan atau organisasi melaporkan dampak sosial, lingkungan, dan ekonomi dari operasional mereka secara transparan dan terukur. Mengintegrasikan pengelolaan sampah dan limbah ke dalam *sustainability reporting*, perusahaan dapat memantau dan melaporkan upaya mereka dalam mengurangi dampak negatif terhadap lingkungan. Hal ini tidak hanya meningkatkan akuntabilitas dan transparansi perusahaan, tetapi juga mendorong praktik bisnis yang lebih bertanggung jawab dan berkelanjutan. Dalam laporan keberlanjutan, perusahaan dapat mencantumkan strategi dan tindakan yang telah diambil untuk mengurangi produksi sampah dan limbah, seperti daur ulang, pengurangan penggunaan bahan berbahaya, dan implementasi teknologi ramah lingkungan. Dengan demikian, *sustainability reporting* tidak hanya berfungsi sebagai alat pemantauan dan evaluasi, tetapi juga sebagai pendorong perubahan positif yang berkelanjutan bagi lingkungan dan masyarakat.

Dayan (2020) menyebutkan bahwa sebelum tahun 2017, tingkat penerapan *sustainability reporting* di Indonesia masih sangat rendah. Hal ini dikarenakan peraturan khusus terkait pelaporan SR di Indonesia baru ada sejak 2017 melalui POKL No 51 / POJK.03 / 2017. Dibuktikan dengan data dari Bursa Efek Indonesia (BEI) yang menyebutkan bahwa dalam konteks perusahaan publik, 49 perusahaan yang *listing* telah menerbitkan *sustainability report* sampai dengan akhir tahun 2016 (Otoritas Jasa Keuangan, 2017) dan sampai dengan tahun 2018, *sustainability report* masih bersifat sukarela (*voluntary*) di Indonesia, sebagian besar perusahaan mengacu pada GRI. Belum adanya standar *sustainability reporting* yang diberlakukan standar setter di Indonesia, sehingga perusahaan menggunakan standar yang berlaku secara global.

METODE PENELITIAN

Metode yang digunakan dalam penelitian ini yaitu analisis deskriptif. Tujuan analisis deskriptif adalah untuk memberikan gambaran yang jelas tentang karakteristik dasar dari suatu set data. Data yang digunakan dalam penelitian ini merupakan data yang diperoleh dari Badan Pusat Statistik (BPS) Indonesia, berupa Data Statistik Lingkungan Hidup Indonesia Tahun 2023. Adapun variable yang digunakan yaitu Indeks Kualitas Udara (IKU), Indeks Kualitas Air (IKA), Indeks Kualitas Lingkungan Hidup (ILKH), Limbah yang Dihasilkan (Manufaktur), Limbah yang Dikelola (Manufaktur) Dan Jumlah Produksi Sampah.

Tools yang digunakan dalam penelitian ini antara lain : 1) *Google Spreadsheet* digunakan untuk mengolah data mentah yang berasal dari Badan Pusat Statistik (BPS). Data yang terinput pada Dataset Fix yang Digunakan merupakan data dari tahun 2021 dan 2022., 2) *Google Colab* adalah sebuah IDE untuk pemrograman Python yang memanfaatkan server Google dengan perangkat keras berperforma tinggi untuk menjalankan proses komputasi (Gelar Guntara, 2023), 3) *Power BI* adalah *platform business intelligence* yang dikeluarkan oleh Microsoft, dirancang untuk membantu organisasi dan individu menggali wawasan dari sebuah data, dan 4) Canva untuk membuat *deck presentation*.

Tabel 1. Indikator Variabel

Variabel	Indikator
Provinsi (P)	Data jumlah provinsi yang ada di Indonesia.
Indeks Kualitas Udara (IKU)	Data pengukuran yang memberikan gambaran tentang seberapa baik atau buruk kualitas udara di berbagai lokasi provinsi yang ada di Indonesia.
Indeks Kualitas Air (IKA)	Data pengukuran yang menilai kualitas air dalam hal keamanan dan kesesuaian untuk berbagai kebutuhan manusia dan ekosistem.
Indeks Kualitas Lingkungan Hidup (IKLH)	Data pengukuran yang lebih luas yang mencakup sejumlah faktor yang mempengaruhi kualitas keseluruhan lingkungan, termasuk udara, air, tanah, keanekaragaman hayati, dan kesejahteraan manusia.
Limbah Yang Dhasilkan (Manufaktur)	Jumlah limbah padat, cair, atau gas yang dihasilkan oleh proses manufaktur suatu industri.
Limbah Yang Dikelola (Manufaktur)	Jumlah limbah dari proses manufaktur yang dikelola dengan benar dan secara bertanggungjawab, misalnya melalui daur ulang, pemrosesan khusus, atau disposisi aman.
Jumlah Produksi Sampah	Ukuran kuantitatif dari total sampah yang dihasilkan oleh suatu populasi, wilayah, atau entitas dalam jangka waktu tertentu. Ini bisa termasuk sampah rumah tangga, komersial, industri, atau konstruksi, dan seringkali menjadi fokus bagi upaya pengelolaan sampah dan daur ulang.

Sumber: data diolah (2024)

HASIL

Pilihan untuk mengadopsi metodologi *Cross-Industri Standard Process for Data mining (CRISP-DM)* yaitu karena metodologi tersebut menawarkan pendekatan yang terstruktur dan sistematis untuk menjalankan proyek *data mining*, mulai dari memahami kebutuhan bisnis hingga penerapan solusi di lingkungan operasional. Berikut ini merupakan enam proses CRISP-DM.

Penerapan *data mining* pada penelitian ini berhubungan dengan data produksi sampah dan limbah untuk menggali pengetahuan tentang suatu kuantitas terhadap kualitas lingkungan hidup yang berpotensi memberikan dampak buruk. *Business understanding* dalam penelitian ini dapat dirumuskan dalam empat poin utama, yaitu pertumbuhan industri yang pesat pada era globalisasi, meningkatnya produksi sampah dan limbah, banyaknya dampak negatif sampah dan limbah, adanya penurunan kualitas udara, air, dan lingkungan hidup.

Pada penelitian ini digunakan teknik *Exploratory Data Analysis (EDA)* dan Statistik Deskriptif. EDA dan Statistik Deskriptif ini membantu *Analyst* untuk merangkum dan mengevaluasi setiap variabel yang digunakan untuk mengidentifikasi parameter yang paling berpengaruh. Selanjutnya penelitian ini juga menggunakan *Correlations Analysis* atau analisis korelasi dilakukan untuk memahami hubungan antara parameter pencemar dan kualitas lingkungan hidup. Melalui analisis korelasi, dapat diketahui sejauh mana satu parameter mempengaruhi yang lain, yang sangat penting untuk perumusan strategi pengendalian pencemaran lingkungan yang efektif.

Sumber data yang digunakan yaitu Data Statistik Lingkungan Hidup Indonesia Tahun 2023 yang dipublikasikan oleh Badan Pusat Statistik Indonesia. Data tersebut merupakan ragam data yang luas mencakup statistik lingkungan hidup dan keterbatasan data yang tersedia menjadi tantangan yang terus diupayakan secara komprehensif.

Tabel 2. Dataset untuk Google Colaboratory

PROVINSI	INDEKS KUALITAS AIR	INDEKS KUALITAS UDARA	INDEKS KUALITAS LINGKUNGAN HIDUP	JUMLAH PRODUKSI SAMPAH	LIMBAH YANG DIHASILKAN (MANUFAKTUR)	LIMBAH YANG DIKELOLA (MANUFAKTUR)	TAHUN
Aceh	57,14	89,63	75,54	255,81	1075,4	1052,3	2021
Sumatera Utara	53,72	89,55	71,15	1746	190492,66	192716,2	2021
Sumatera Barat	52,55	90,22	73,68	643,76	182303,56	182303,1	2021
Riau	52,25	90,13	70,72	977	29165,72	20000,53	2021
Jambi	48,96	67,08	69,04	437,5	1672,66	1672,2	2021
Sumatera Selatan	58,25	86,28	69,7	1204,97	90383,92	90383,7	2021
Bengkulu	49,81	90,81	71,46	775	3,53	3,22	2021
Lampung	57,77	85,46	68,56	1100	349,79	322,16	2021
Kepulauan Bangka Belitung	58,37	90,39	72,05	140	27,05	25,87	2021
Kepulauan Riau	55,15	90,91	73,19	158,07	125499,45	110672,16	2021
DKI Jakarta	44,19	66,52	54,43	8,52	111765,07	108617,46	2021
Jawa Barat	43,09	79,34	62,68	1723,4	29300026,59	29855079,32	2021
Jawa Tengah	47,94	84,6	66,27	5600,88	554354,18	542746,64	2021



PROVINSI	INDEKS KUALITAS AIR	INDEKS KUALITAS UDARA	INDEKS KUALITAS LINGKUNGAN HIDUP	JUMLAH PRODUKSI SAMPAH	LIMBAH YANG DIHASILKAN (MANUFAKTUR)	LIMBAH YANG DIKELOLA (MANUFAKTUR)	TAHUN
DI Yogyakarta	45,73	88,59	65,66	303,13	1180,74	1057,6	2021
Jawa Timur	53,57	83,2	68,29	1783,68	2413488,54	2332528,31	2021
Banten	54,95	74,14	64,14	572,48	1836173,38	1703311,21	2021
Bali	54,29	89,28	70,7	866,61	172,22	172,1	2021
Nusa Tenggara Barat	45,1	88,52	69,89	255,3	8,96	8,96	2021
Nusa Tenggara Timur	58,28	90,51	74,97	227,93	0,35	0	2021
Kalimantan Barat	54,35	90,71	72,9	1851,85	2430905,07	2430907,65	2021
Kalimantan Tengah	55,34	90,39	75,11	1012,75	4131	4126	2021
Kalimantan Selatan	54,75	89,15	71,03	511,98	9674,03	96733,42	2021
Kalimantan Timur	51,92	88,84	75,06	587,25	14998,51	14998,41	2021
Kalimantan Utara	57,34	93,43	80,85	132,97	6,79	6,79	2021
Sulawesi Utara	49,69	91,27	72,38	291,1	147,23	146,24	2021
Sulawesi Tengah	55,84	91,33	77,78	197,54	35447,86	32462,89	2021
Sulawesi Selatan	56,82	89,13	73,23	1002,53	49631,61	26896,23	2021
Sulawesi Tenggara	53,26	90,89	74,97	246,15	18,15	18,15	2021
Gorontalo	53,46	93,96	77,32	141,49	0	0	2021
Sulawesi Barat	56,04	90,97	75,72	140,93	19,4	19,2	2021
Maluku	55,56	90,7	78,23	246,4	0	0	2021
Maluku Utara	53,08	91,64	77,45	46,31	282,85	282,85	2021
Papua Barat	54,44	95,6	81,8	165,63	82,85	81,41	2021
Papua	57,83	95,61	80,23	253,27	0	0	2021
Aceh	60,41	95,62	78,29	257,81	1082,53	1075,4	2022
Sumatera Utara	61,2	95,63	74,11	1748	192716,52	190492,66	2022
Sumatera Barat	55,64	95,64	74,98	645,76	182303,89	182303,56	2022
Riau	53,88	95,65	72,1	987	30095,19	29165,72	2022
Jambi	49,49	95,66	70,32	439,5	2181,46	1672,66	2022
Sumatera Selatan	59,85	95,67	72,12	1206,97	91370,04	90383,92	2022
Bengkulu	46,18	95,68	70,82	875	3,68	3,53	2022
Lampung	54,72	95,69	69,1	1140	353,16	349,79	2022
Kepulauan Bangka Belitung	57,68	95,7	71,67	160	28,87	27,05	2022
Kepulauan Riau	56,79	95,71	74,17	178,07	130672,16	125499,45	2022
DKI Jakarta	41,17	95,72	54,65	9,52	218617,46	111765,07	2022
Jawa Barat	47,13	95,73	64,66	1728,4	29855079,32	29300026,59	2022
Jawa Tengah	48,16	95,74	66,85	5680,88	592746,64	554354,18	2022
DI Yogyakarta	39,31	95,75	65,96	363,13	2257,6	1180,74	2022
Jawa Timur	56,13	95,76	69,92	1793,68	2532528,31	2413488,54	2022

PROVINSI	INDEKS KUALITAS AIR	INDEKS KUALITAS UDARA	INDEKS KUALITAS LINGKUNGAN HIDUP	JUMLAH PRODUKSI SAMPAH	LIMBAH YANG DIHASILKAN (MANUFAKTUR)	LIMBAH YANG DIKELOLA (MANUFAKTUR)	TAHUN
Banten	55,49	95,77	63,41	592,48	2203311,21	1836173,38	2022
Bali	53,5	95,78	70,89	886,61	172,6	172,22	2022
Nusa Tenggara Barat	43,39	95,79	70,56	265,3	8,96	8,96	2022
Nusa Tenggara Timur	52,62	95,8	73,49	237,93	0,35	0	2022
Kalimantan Barat	55,52	95,81	71,9	1871,85	2430907,65	2430905,07	2022
Kalimantan Tengah	56,69	95,82	76,62	1052,75	41,56	41,31	2022
Kalimantan Selatan	54,63	95,83	71,97	541,98	96753,42	96741,03	2022
Kalimantan Timur	53,02	95,84	74,46	589,25	14998,41	4280,11	2022
Kalimantan Utara	54,46	95,85	80,67	138,97	6,79	6,79	2022
Sulawesi Utara	48,24	95,86	72,37	321,1	147,23	146,24	2022
Sulawesi Tengah	57,71	95,87	78,66	207,54	35462,89	35447,86	2022
Sulawesi Selatan	57,79	95,88	74,17	1032,53	49631,61	26896,23	2022
Sulawesi Tenggara	56,21	95,89	76,88	276,15	18,15	16,89	2022
Gorontalo	58,07	95,9	79,79	161,49	0	0	2022
Sulawesi Barat	56,41	95,91	76,77	150,93	19,4	17,8	2022
Maluku	52,08	95,92	77,69	256,4	0	0	2022
Maluku Utara	55,69	95,93	78,84	49,31	282,85	282,85	2022
Papua Barat	62	95,94	84,22	169,63	82,85	81,41	2022
Papua	55,07	95,95	81,46	263,27	0	0	2022

Sumber: data diolah (2024)

Tahap awal dalam *data preparation* yaitu menentukan atribut. Atribut yang digunakan adalah semua kolom dan ada penambahan kolom berupa kolom kategori yaitu Kategori Kualitas Lingkungan Hidup, Kualitas Udara, dan Kualitas Air. Langkah selanjutnya adalah memisahkan data menjadi *data train* dan *data test*. Pemisahan ini dilakukan dengan tujuan untuk mengevaluasi kinerja model *machine learning* dengan lebih akurat dan menghindari *overfitting*. *Data train* digunakan untuk melatih model. Model belajar dari data ini untuk mengenali pola-pola yang ada dan menyesuaikan parameter-parameter model. *Data test* digunakan untuk menguji kinerja model yang sudah dilatih. Selain itu, *data test* juga membantu menentukan seberapa baik model mampu memprediksi data yang belum pernah dilihat sebelumnya, dan mengukur kemampuan generalisasi model, yaitu kemampuan untuk bekerja baik pada data baru. Berikut merupakan kode untuk pembagian data set.

```
[ ] # Memisahkan data menjadi set train dan test.  
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.4, random_state=42)
```

Gambar 1. Pembagian Data Set

Setelah memisahkan data, langkah selanjutnya yaitu memeriksa apakah ada data yang hilang dengan menggunakan `df.isna().sum()` fungsinya untuk menghitung nilai yang hilang (NaN atau null) dalam setiap kolom.

```
df.isna().sum() #menghitung nilai yg hilang
PROVINSI 0
INDEKS KUALITAS AIR 0
INDEKS KUALITAS UDARA 0
IKLH 0
JUMLAH PRODUKSI SAMPAH 0
LIMBAH YANG DIHASILKAN (MANUFAKTUR) 0
LIMBAH YANG DIKELOLA (MANUFAKTUR) 0
TAHUN 0
Kategori Kualitas Lingkungan 0
Kategori Kualitas Udara 0
Kategori Kualitas Air 0
Kategori 0
dtype: int64
```

Gambar 2. Perhitungan Nilai yang Hilang

Sumber: data diolah (2024)

Pada tahap *Analytical Approach and Modelling*, proses dimana model dengan teknik regresi yang dipilih menghasilkan suatu pola informasi yang dapat memudahkan pihak yang berkepentingan dalam pengambilan keputusan. Pola regresi yang dihasilkan oleh teknik *data mining* ini digunakan untuk memprediksi dampak dari produksi sampah dan limbah terhadap kualitas lingkungan hidup. Adapun *tools* atau alat bantu dalam pengolahan data mining yaitu *Google Collaboratory*. Tahap ini dapat dimulai dari :

Menentukan algoritma yang sesuai untuk analisis data dengan cara mencoba tiga model yaitu *Linear Regression*, *Random Forest*, dan *Logistic Regression* untuk mencari nilai akurasi yang paling baik. Kriteria akurasi dapat dikatakan baik biasanya saat hasil akurasi $>0,7$ dan kurang dari 1,0. Akurasi 1,0 dapat dikatakan sebagai *overfitting*, yaitu keadaan saat suatu model mempelajari *data train* terlalu baik sehingga akan membuat kinerjanya buruk pada data baru.

Menginterpretasikan hasil model dengan cara analisis *correlation* untuk melihat hubungan antara masing-masing variabel yang digunakan dengan bantuan visualisasi dari *Heatmap* Korelasi. *Heatmap* korelasi memberikan gambaran yang jelas mengenai kekuatan dan arah hubungan antar variabel, sehingga dapat mempercepat identifikasi korelasi antar variabel baik positif maupun negatif. Kriteria koefisien korelasi adalah 1) Semakin dekat nilai koefisien korelasi dengan 1, maka semakin tinggi hubungan positif antara dua variabel tersebut. 2) Semakin dekat nilai koefisien korelasi dengan -1, maka semakin tinggi hubungan

negatif antara dua variabel tersebut. 3) Nilai 0 menunjukkan tidak adanya korelasi linier antara dua *variable*.

Tabel 3. Heatmap korelasi

INDEKS KUALITAS AIR	1,00	0,26	0,64	-0,14	-0,29	-0,29	0,05	-0,54
INDEKS KUALITAS UDARA	0,26	1,00	0,48	-0,07	-0,13	-0,14	0,64	-0,50
IKLH	0,64	0,48	1,00	-0,31	-0,31	-0,31	0,07	-0,89
JUMLAH PPRODUKSI SAMPAH	-0,14	-0,07	-0,31	1,00	0,21	0,21	0,01	0,31
DIHASILKAN (MANUFAKTUR)	-0,29	-0,13	-0,31	0,21	1,00	1,00	0,00	0,26
LIMBAH YANG DIKELOLA (MANUFAKTUR)	-0,29	-0,14	-0,31	0,21	1,00	1,00	0,00	0,26
TAHUN	0,05	0,64	0,07	0,01	0,00	0,00	1,00	-0,07
KATEGORI	-0,54	-0,5	-0,89	0,31	0,26	0,26	-0,07	1,00

Sumber: data diolah (2024)

Gambar diatas dapat diinterpretasikan sebagai berikut : 1) Hubungan antara Indeks Kualitas Air dan Indeks Kualitas Lingkungan Hidup Sangat Erat (0.64), menunjukkan bahwa Kualitas Air dan Kualitas Lingkungan Hidup berkorelasi Positif Kuat. 2) Hubungan antara Jumlah Produksi Sampah dengan Indeks Kualitas Air (-0.14), Indeks Kualitas Udara (-0.07), dan Indeks Kualitas Lingkungan Hidup (-0.31) adalah Negatif, menunjukkan bahwa semakin tinggi Produksi Sampah, semakin rendah Kualitas Air Dan Udara, serta semakin buruk Kategori Lingkungan Hidup. 3) Hubungan antara Tahun dengan Indeks Kualitas Udara (0.64) menunjukkan korelasi positif yang cukup kuat, mungkin menunjukkan perubahan seiring waktu dalam Kualitas Udara. 4) Hubungan antara Indeks Kualitas Lingkungan Hidup dengan Kategori (-0.89) sangat dekat dengan -1, menunjukkan Korelasi Negatif yang sangat kuat, mungkin menunjukkan bahwa Indeks Kualitas Lingkungan Hidup adalah faktor penentu utama dalam menentukan Kategori Lingkungan Hidup.

Model yang digunakan untuk melihat tren dan parameter yang berpengaruh yaitu : *Linear regression* digunakan untuk melihat parameter mana yang paling berpengaruh terhadap kategori Kualitas Lingkungan (Sangat Baik, Baik, Tidak Sehat, Buruk).

```
Linear Regression

[ ] Import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score

# membaca data dari file csv
df = pd.read_csv('file_data_kategori.csv')

# pilih kolom-kolom yang akan digunakan sebagai fitur
features = ['IKU', 'IKLH', 'JPS']

# variabel prediktor (X) dan target (y)
X = df[features]
y = df['kategori']

# membagi data menjadi set train dan test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# melakukan standardisasi fitur
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

[ ] # Hitung korelasi Pearson antara setiap fitur dan variabel target
correlations = {}
for feature in features:
    correlation, _ = pearsonr(df[feature], df['kategori'])
    correlations[feature] = correlation

# Menampilkan korelasi Pearson untuk setiap fitur
for feature, correlation in correlations.items():
    print(f"Korelasi Pearson antara {feature} dan kategori: {correlation}")

Korelasi Pearson antara IKU dan kategori: -0.541057120918
Korelasi Pearson antara IKLH dan kategori: -0.90320192114906
Korelasi Pearson antara JPS dan kategori: 0.21421948117448
```

Gambar 3. Analisis Modeling

Gambar diatas dapat diinterpretasikan sebagai berikut : 1) Korelasi pearson antara IKU dan Kategori = -0,541, menunjukkan adanya korelasi negatif sedang antara IKU dan Kategori. 2) Korelasi pearson antara IKLH dan Kategori = -0,903, menunjukkan adanya korelasi negatif yang sangat kuat antara IKLH dan Kategori. 3) Korelasi pearson antara JPS dan Kategori = 0,214, menunjukkan adanya korelasi positif yang lemah antara JPS dan Kategori.

```
# Membuat model Random Forest Classifier
model = RandomForestClassifier()

# Melatih model menggunakan data pelatihan
model.fit(X_train, y_train)

# Melakukan prediksi pada data pengujian
y_pred = model.predict(X_test)

# Menghitung akurasi
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

Accuracy: 1.0
```

Gambar 4. Kode Random Forest

Dalam gambar diatas terdapat *Metrix evaluated: accuracy: 1,0* artinya model ini mungkin *overfitting* terhadap data, sehingga performa pada data baru dapat menurun.

```

* LogisticRegression
LogisticRegression()

* Melakukan prediksi pada data pengujian dan mengevaluasi performa model
y_pred = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))

Accuracy: 0.8636363636363636
    
```

Gambar 5. Kode Logistic Regression

Pada gambar diatas nilai *Metrics evaluated: accuracy* : 0,8636363636363636. *Area Under Curve* (AUC): 0,8 artinya akurasi sebesar 86,63% menunjukkan model ini cukup baik dalam memprediksi data. AUC sebesar 0,8 menunjukkan kemampuan model dalam membedakan antara kelas dengan baik.

Setelah melalui tahap modeling dengan tiga algoritma yang berbeda, yakni *Linear Regression*, *Random Forest*, dan *Logistic Regression*, langkah selanjutnya yaitu melakukan evaluasi menggunakan beberapa metrik evaluasi yang digunakan meliputi korelasi Pearson, akurasi, dan *Area Under Curve* (AUC) untuk menentukan model yang paling sesuai untuk tujuan penelitian ini. Berikut adalah evaluasi dari masing-masing model:

Tabel 4. Evaluasi 3 Model

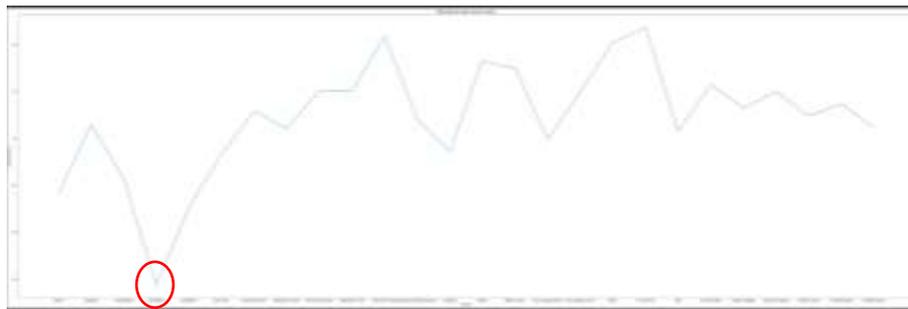
<i>Linear Regression</i>	<i>Random Forest</i>	<i>Logistic Regression</i>
<p><i>Metrics Evaluated:</i> Korelasi Pearson antara: 1) IKU dan Kategori: - 0.541 2) IKLH dan Kategori: - 0.903 3) JPS dan Kategori: 0.214</p>	<p><i>Metrics Evaluated:</i> Accuracy: 1.0</p>	<p><i>Metrics Evaluated:</i> 1) Accuracy: 0.86363636363636 2) Area Under Curve (AUC): 0.8</p>
<p><i>Linear Regression</i> menunjukkan korelasi yang cukup baik antara Indeks Kualitas Lingkungan Hidup (IKLH) dan kategori kualitas lingkungan. Namun, hasilnya tidak sebaik model lain dalam hal akurasi dan pemisahan kelas.</p>	<p>Model ini mungkin <i>overfitting</i> terhadap data, sehingga performa pada data baru dapat menurun.</p>	<p>Accuracy sebesar 86.36% menunjukkan model ini cukup baik dalam memprediksi data. AUC sebesar 0.8 menunjukkan kemampuan model dalam membedakan antara kelas dengan baik.</p>

Sumber: data diolah (2024)

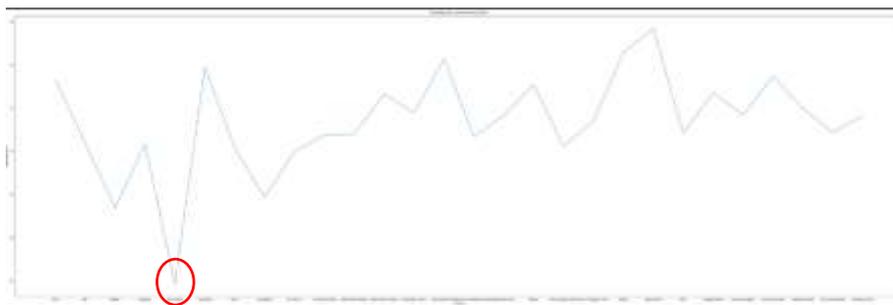
Perbandingan ketiga model diatas adalah 1) *Linear Regression* memiliki performa yang baik dalam hal prediksi variabel kontinu, tetapi kurang cocok untuk klasifikasi. 2) *Ran-*

dom Forest menunjukkan kinerja yang sangat baik dengan akurasi sempurna, tetapi ada kemungkinan *overfitting*. 3) *Logistic Regression* memberikan hasil yang baik dengan akurasi tinggi dan AUC yang cukup tinggi, menjadikannya model yang paling cocok untuk tujuan klasifikasi dalam penelitian ini. Model yang dipilih untuk digunakan dalam penelitian ini adalah *Logistic Regression*, karena menunjukkan kinerja terbaik dalam hal prediksi dan evaluasi metrik dengan Akurasi sebesar 86.36%.

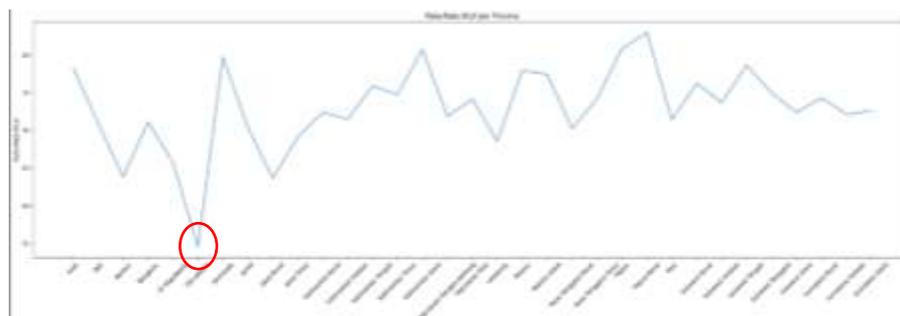
Setelah melalui tahap evaluasi, di mana menilai secara detail hasil dari sebuah model maka dilakukan pengimplementasian dari keseluruhan model yang telah dibangun. Selain itu juga dapat dilakukan penyesuaian terhadap model sehingga mendapatkan hasil yang sesuai dengan target awal tahap CRISP-DM ini. Langkah berikutnya adalah *Deployment*. Berikut ini merupakan hasil perbandingan rata-rata Indeks Kualitas Lingkungan Hidup Per Provinsi Tahun 2021 dan 2022:



Gambar 6. Rata-Rata IQLH per Provinsi Tahun 2021

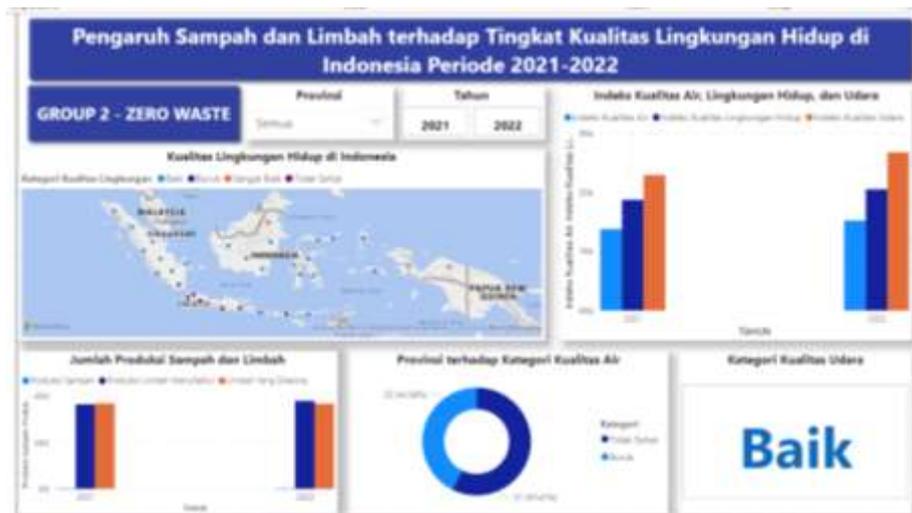


Gambar 7. Rata-Rata IQLH per Provinsi Tahun 2022



Gambar 8. Rata-Rata IQLH per Provinsi Tahun 2021-2022

Setelah melakukan analisis, dapat disusun sebuah visualisasi sebagai berikut:



Gambar 9. Visualisasi

Pengaruh produksi sampah dan limbah terhadap tingkat kualitas lingkungan hidup di Indonesia periode 2021-2022, adalah sebagai berikut :

Terdapat 57,41% provinsi di Indonesia memiliki kualitas air yang tidak sehat dan 42,59% provinsi di Indonesia memiliki kualitas air yang buruk. Hal ini menunjukkan bahwa kualitas air masih menjadi tantangan utama di banyak provinsi. Pada tahun 2021-2022 kualitas udara di Indonesia tergolong baik. Pada tahun 2022 dilaporkan skor IKU di Indonesia sebesar 88,06 poin. Skor tersebut telah memenuhi target pada tahun 2022 sebesar 84,3 poin. Skor ini mengalami kenaikan 0,8% dibandingkan tahun 2021. Pada tahun 2021, skor IKU Indonesia terpantau sebesar 87,36 poin. Dari 162 titik pemantauan, sebanyak 57 kabupaten/kota mengalami penurunan kualitas udara. Namun, IKU mengalami kenaikan 0,97 poin karena pada pemantauan di 341 titik lainnya mengalami perbaikan, atau 215 kabupaten/kota mengalami perbaikan. Jumlah produksi sampah pada tahun 2021 ke 2022 mengalami peningkatan, baik dalam hal produksi sampah rumah tangga maupun limbah industry. Namun, volume sampah yang dikelola juga meningkat, artinya mengindikasikan bahwa adanya upaya yang lebih baik dalam pengelolaan sampah pada tahun 2022.

Berdasarkan interpretasi dari visualisasi yang telah dibuat, kaitannya dengan 17 Tujuan Pembangunan Berkelanjutan (SDGs) diantaranya sebagai berikut :

Tujuan 6: Akses Air Bersih dan Sanitasi .Air bersih dan sanitasi yang layak merupakan kebutuhan dasar manusia. Namun, dari data yang telah didapatkan dan diolah, kualitas air di Indonesia sebagian besar tergolong tidak sehat dan buruk. Hal ini menunjukkan perlu

adanya upaya peningkatan kualitas air agar dapat mencapai tujuan SDGs ke 6. Akses air bersih dan sanitasi yang layak dapat dicapai dengan mengurangi polusi air, memperhatikan pembuangan limbah, melindungi dan memperbaiki ekosistem terkait air, hutan, dan resapan air.

Tujuan 11: Kota dan Pemukiman yang Berkelanjutan. Adanya peningkatan jumlah produksi sampah dan limbah terkait erat dengan poin ke 11 yaitu Kota dan Pemukiman yang Berkelanjutan. Menekankan pada pentingnya mengurangi dampak lingkungan negatif di kota-kota terutama yang disebabkan oleh sampah dan limbah. Mewujudkan kota yang aman dan berkelanjutan dapat dilakukan melalui pengelolaan sampah yang efektif. Langkah-langkah lain yang mendukung tujuan ini meliputi investasi dalam transportasi umum, penyediaan ruang hijau untuk masyarakat, serta pengembangan perencanaan dan pengelolaan kota yang bersifat inklusif.

Tujuan 13: Penanganan Perubahan Iklim. Kualitas udara yang “baik” pada tahun 2022 mencerminkan langkah-langkah yang dapat mendukung SDGs poin 13, yang berfokus pada mitigasi perubahan iklim. Pengelolaan limbah dan polusi udara yang lebih baik dapat membantu mengurangi emisi gas rumah kaca dan mendukung target untuk menangani perubahan iklim.

Tujuan 14: Ekosistem Lautan dan Kehidupan Bawah Air dan Tujuan 15: Ekosistem Daratan dan Kehidupan di Bumi. Kualitas air yang buruk di beberapa wilayah Indonesia berpotensi mempengaruhi ekosistem laut dan darat. Tujuan poin ke 14 dan 15 berfokus untuk melindungi ekosistem bawah air dan daratan dari kontaminasi polusi, yang sangat terkait dengan manajemen limbah dan kualitas air.

SIMPULAN

Penelitian yang dilakukan yaitu menganalisis dampak produksi sampah dan limbah terhadap kualitas lingkungan hidup. Menggunakan data dari Badan Pusat Statistik (BPS) dengan mengidentifikasi korelasi antara berbagai parameter lingkungan, seperti kualitas udara, air, dan kualitas lingkungan. Model regresi logistik menunjukkan hasil terbaik dengan akurasi 86,36% dan AUC 0,8, menegaskan bahwa model ini sebagai model yang paling sesuai untuk klasifikasi penelitian ini. Menggunakan metode *Cross-Industry Standard Process for Data Mining (CRISP-DM)* yang terdiri dari enam tahapan, yaitu: *Business Understanding, Data Understanding, Data Preparation, Analytical Approach and Modelling,*

Evaluation, dan Deployment.

Hasil visualisasi analisis data yang dilakukan, menunjukkan bahwa produksi sampah dan limbah berkontribusi negatif terhadap kualitas lingkungan, sejalan dengan tujuan keberlanjutan SDGs, khususnya terkait air bersih, kualitas lingkungan, iklim, serta ekosistem darat dan laut. Guna mendukung keberlanjutan SDGs dan meningkatkan kualitas lingkungan hidup, ada beberapa inovasi yang dapat dilakukan seperti : 1) Program pemanfaatan lubang biopori untuk mengolah sampah organik menjadi pupuk. 2) Teknologi pengolahan sampah terintegrasi untuk mendukung circular economy. 3) Implementasi kebijakan insentif digital bagi pengelolaan sampah yang baik. 4) Implementasi kebijakan pelaporan pengelolaan limbah industry bagi perusahaan

Berdasarkan kesimpulan dan keterbatasan penelitian, beberapa saran yang dapat penulis berikan untuk pengembangan penelitian di masa mendatang: 1) Memperpanjang durasi penelitian untuk mencakup analisis jangka panjang. 2) Melakukan pengumpulan data tambahan yang lebih luas dan terperinci. 3) Melibatkan pemangku kepentingan untuk rencana pengelolaan limbah yang lebih efektif. 4) Mengintegrasikan penelitian dengan program pemantauan kualitas lingkungan berkelanjutan

DAFTAR PUSTAKA

- Dayan, D. W. (2020). *Sustainability Reporting di Indonesia: Peran, Tantangan, dan Dilema dibalik Penerapannya*. <https://doi.org/10.6084/m9.figshare.15655692.v1>
- Gelar Guntara, R. (2023). Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(1), 55–60. <https://doi.org/10.47233/jteksis.v5i1.750>
- Hayuning, I., & Sudarmaji. (2016). Pengelolaan Limbah B3 Dan Keluhan Kesehatan Pekerja Di Pt. Inka (Persero) Kota Madiun B3 Waste Management and Health Workers Complaint In. Inka (Persero) Madiun City. *JURNAL Kesehatan Lingkungan*, 8(1), 118–127.
- Otoritas Jasa Keuangan. (2017). *POJK 51 Tahun 2017*.
- Meyrena, S. D., & Amelia, R. (2020). Analisis Pendayagunaan Limbah Plastik Menjadi Ecopaving Sebagai Upaya Pengurangan Sampah. *Indonesian Journal of Conservation*, 9(2), 96–100. <https://doi.org/10.15294/ijc.v9i2.27549>
- Nurlina. (2021). Toxic And Hazardous Waste (B3) Management At Pt. Pal Indonesia (Persero). *The Indonesian Journal of Public Health*, 16(3), 449–460. <https://doi.org/10.20473/ijph.v16i1.2021.449-460>
- Nursabrina, A., Joko, T., & Septiani, O. (2021). Kondisi Pengelolaan Limbah B3 Industri Di



Indonesia Dan Potensi Dampaknya: Studi Literatur. *Jurnal Riset Kesehatan Poltekkes Depkes Bandung*, 13(1), 80–90. <https://doi.org/10.34011/juriskesbdg.v13i1.1841>

Permadi, I. M. A., & Murni, R. . R. (2013). Dampak pencemaran lingkungan akibat limbah dan upaya penanggulangannya di kota denpasar. *Kertha Negara*, 1, 3–7.

Shahzad, M., Qu, Y., Rehman, S. U., & Zafar, A. U. (2022). Adoption of green innovation technology to accelerate sustainable development among manufacturing industry. *Journal of Innovation and Knowledge*, 7(4), 100231. <https://doi.org/10.1016/j.jik.2022.100231>

Yurnalisdell, Y. (2023). Analisis Pengelolaan Limbah Bahan Berbahaya dan Beracun (B3) di Indonesia. *Jurnal Syntax Admiration*, 4(2), 201–208. <https://doi.org/10.46799/jsa.v4i2.562>