Analisis Pengelolahan Limbah Bahan Kimia dari Fasilitas Industri di Amerika Serikat Menggunakan Double Optimization dan Topic Modelling

Kotak Riset SC

Kotak Riset SC

Rhafael Chandra

Krisna Bayu Dharma Putra

Vincent Yeozekiel

Daftar Isi

Pendahuluan

1 Latar Belakang

Industri adalah **penyumbang terbesar** limbah global. Pengelolaan limbah kimia yang buruk tidak hanya **mencemari lingkungan** tetapi juga mencerminkan inefisiensi yang dapat **merugikan perusahaan** secara finansial. [1]

Analisis dan pemodelan diperlukan untuk mendukung **pengelolaan limbah kimia** yang lebih efektif, terutama dalam menghadapi tantangan lingkungan yang semakin kompleks akibat aktivitas industri. [2] [3]

[1] A. T. Charette, M. B. Collins, dan J. E. Mirowsky, “Assessing residential socioeconomic factors associated with pollutant releases using EPA’s Toxic Release Inventory,” J Environ Stud Sci, vol. 11, no. 2, hlm. 247–257, Jun 2021, doi: 10.1007/s13412-021-00664-7.
[2] A. Marvuglia, M. Kanevski, dan E. Benetto, “Machine learning for toxicity characterization of organic chemical emissions using USEtox database: Learning the structure of the input space,” Environment International, vol. 83, hlm. 72–85, Okt 2015, doi: 10.1016/j.envint.2015.05.011.
[3] S.-R. Lim, C. W. Lam, dan J. M. Schoenung, “Quantity-based and toxicity-based evaluation of the U.S. Toxics Release Inventory,” Journal of Hazardous Materials, vol. 178, no. 1, hlm. 49–56, Jun 2010, doi: 10.1016/j.jhazmat.2010.01.041.

Tujuan dan Manfaat

Tujuan

Untuk menganalisis dan memodelkan limbah produksi yang dihasilkan pada setiap fasilitas.

Manfaat

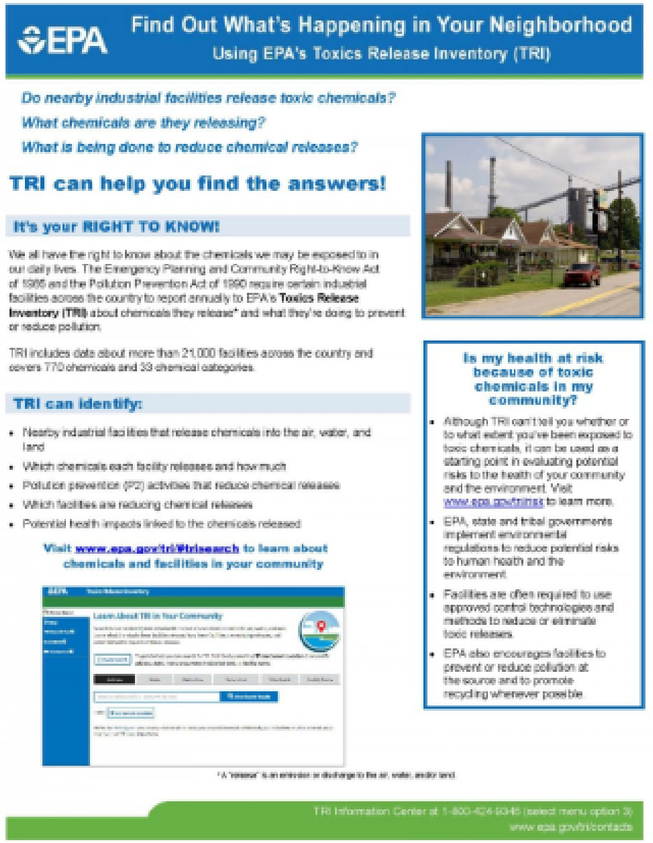
- **Kesehatan Publik**
- **Efisiensi Operasional**
- **Kepatuhan Regulasi**
- **Pengawasan Efektif**

Pembahasan: Landasan Teori

Toxic Release Inventory

Toxic Release Inventory (TRI) merupakan program dari agensi keamanan lingkungan Amerika Serikat (EPA) yang bertujuan untuk **melaporkan limbah kimia** yang dilepaskan ke lingkungan untuk setiap fasilitas industri. Program terbentuk dari kesadaran terhadap limbah kimia yang dilepaskan ke lingkungan [4].

Tujuan utama TRI adalah untuk **meningkatkan transparansi dan kesadaran masyarakat** mengenai penggunaan dan pelepasan bahan kimia berbahaya di lingkungan sekitar mereka. Dengan menyediakan informasi tentang jumlah dan jenis bahan kimia yang dilepaskan ke udara, air, dan tanah, TRI membantu masyarakat dan pemerintah dalam mengidentifikasi potensi risiko kesehatan dan lingkungan [5].


EPA

Find Out What's Happening in Your Neighborhood

Using EPA's Toxics Release Inventory (TRI)

Do nearby industrial facilities release toxic chemicals?

What chemicals are they releasing?

What is being done to reduce chemical releases?

TRI can help you find the answers!

It's your RIGHT TO KNOW!

We all have the right to know about the chemicals we may be exposed to in our daily lives. The Emergency Planning and Community Right-to-Know Act of 1986 and the Pollution Prevention Act of 1990 require certain industrial facilities across the country to report annually to EPA's **Toxics Release Inventory (TRI)** about chemicals they release* and what they're doing to prevent or reduce pollution.

TRI includes data about more than 21,000 facilities across the country and covers 770 chemicals and 33 chemical categories.

TRI can identify:

- Nearby industrial facilities that release chemicals into the air, water, and land
- Which chemicals each facility releases and how much
- Pollution prevention (P2) activities that reduce chemical releases
- Which facilities are reducing chemical releases
- Potential health impacts linked to the chemicals released

Visit www.epa.gov/tri/#triesearch to learn about chemicals and facilities in your community

Is my health at risk because of toxic chemicals in my community?

- Although TRI can't tell you whether or to what extent you've been exposed to toxic chemicals, it can be used as a starting point in evaluating potential risks to the health of your community and the environment. Visit www.epa.gov/trihlias to learn more.
- EPA, state and tribal governments implement environmental regulations to reduce potential risks to human health and the environment.
- Facilities are often required to use approved control technologies and methods to reduce or eliminate toxic releases.
- EPA also encourages facilities to prevent or reduce pollution at the source and to promote recycling whenever possible.

TRI Information Center at 1-800-424-9346 (select menu option 3)

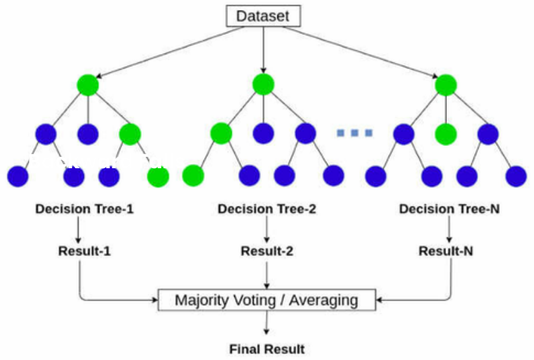
www.epa.gov/tri/contacts

[4] M. M. Jobe, "The power of information: The example of the u.s. toxics release inventory**The author thanks Patricia McClure of the Government Publications Library, University of Colorado at Boulder, for her editorial assistance.," Journal of Government Information, vol. 26, no. 3, hlm. 287–295, Mei 1999, doi: 10.1016/S1352-0237(99)00030-1.

[5] O. US EPA, "Toxics Release Inventory (TRI) Program." Diakses: 8 Januari 2025. [Daring]. Tersedia pada: https://www.epa.gov/toxics-release-inventory-tri-program

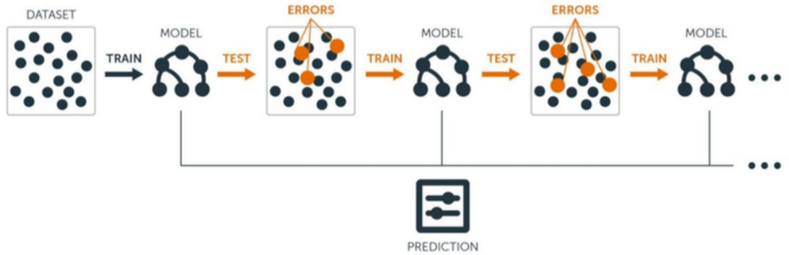
Model Machine Learning

Random Forest [6]



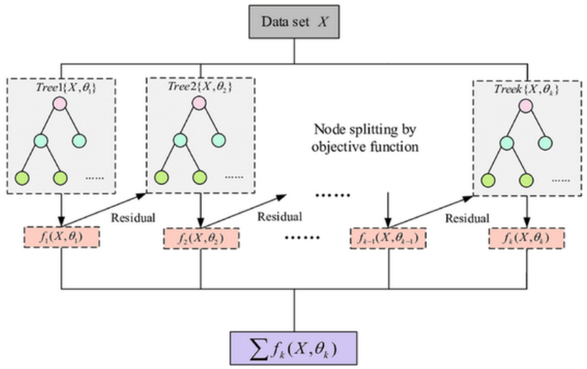
Voting beberapa pohon keputusan

LightGBM [7]



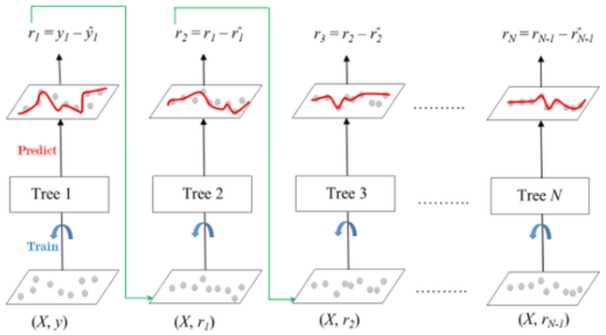
membangun pohon berdasarkan strategi leaf-wise

XGBoost [8]



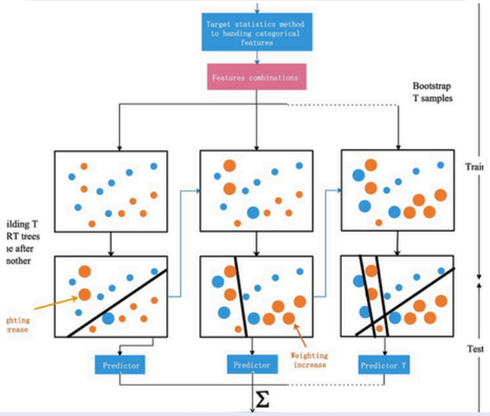
Paralelisme Gradient Boosting

Gradient Boosting [9]



membangun model secara iteratif

CatBoost [10]



dibangun supaya dapat menangkap fitur kategorikal lebih baik

[6] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, hlm. 5–32, Okt 2001, doi: 10.1023/A:1010933404324.

[7] G. Ke dkk., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” dalam Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017. Diakses: 8 Januari 2025. [Daring]. Tersedia pada: https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

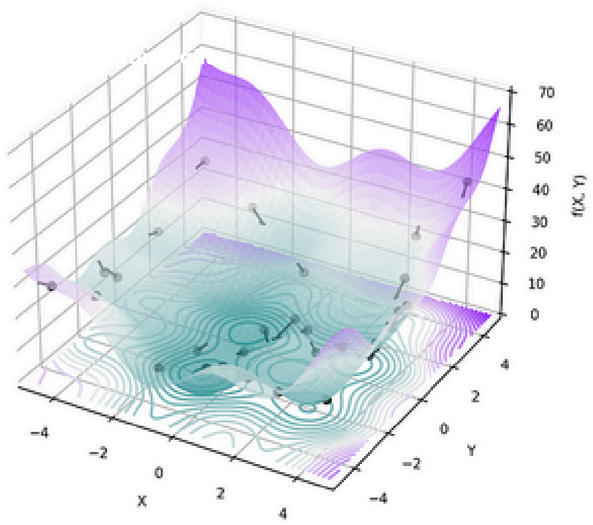
[8] G. Ke dkk., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” dalam Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017. Diakses: 8 Januari 2025. [Daring]. Tersedia pada: https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

[9] Y. Freund dan R. E. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,” J. Comput. Syst. Sci., vol. 55, no. 1, hlm. 119–139, Agu 1997, doi: 10.1006/jcss.1997.1504.

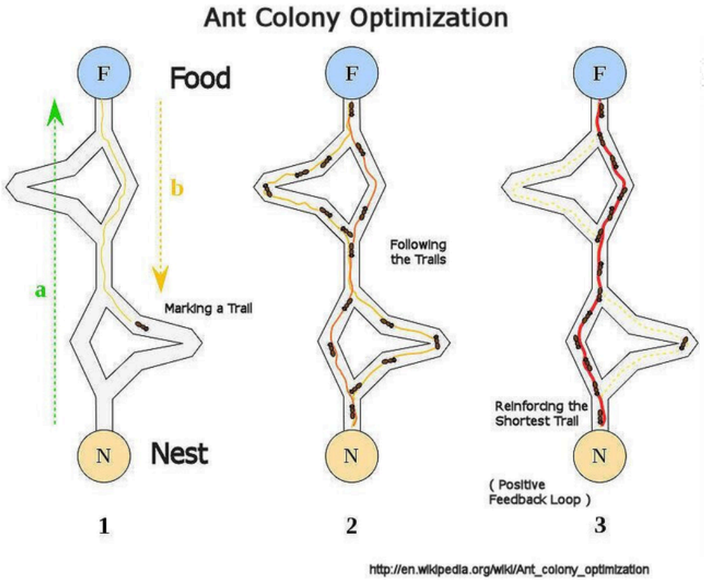
[10] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, dan A. Gulin, “CatBoost: unbiased boosting with categorical features,” 20 Januari 2019, arXiv: arXiv:1706.09516. doi: 10.48550/arXiv.1706.09516.

Metode Optimisasi

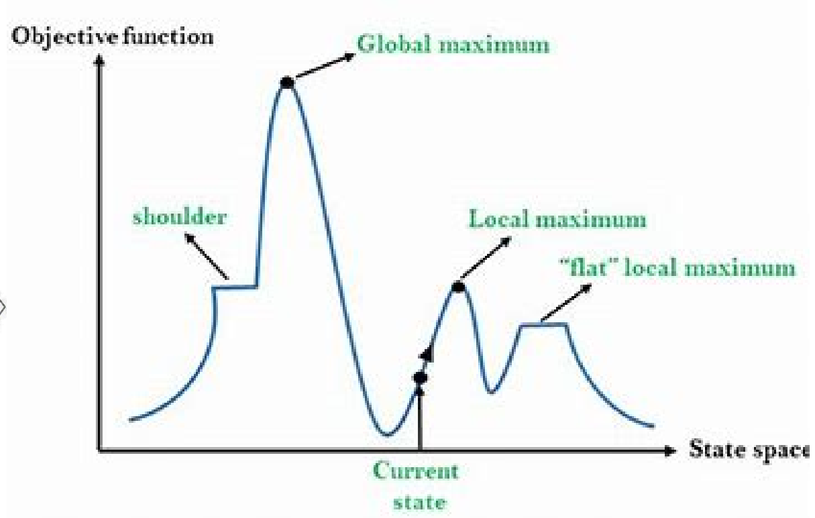
Particle Swarm Optimization [11]



Ant Colony Optimization [12]



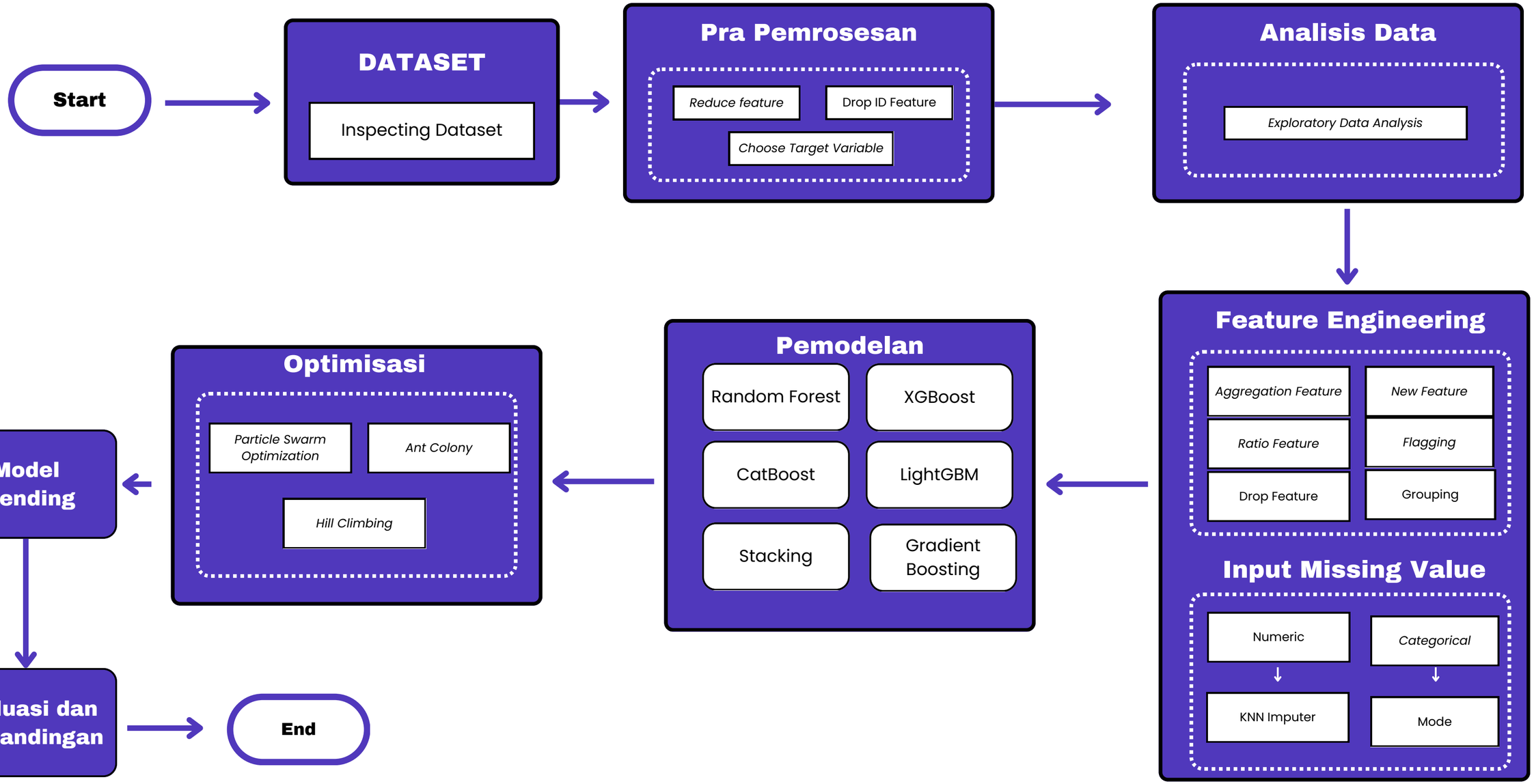
Hill Climbing [13]



[11] J. Kennedy dan R. Eberhart, "Particle swarm optimization," dalam Proceedings of ICNN'95 - International Conference on Neural Networks, Nov 1995, hlm. 1942–1948 vol.4. doi: 10.1109/ICNN.1995.488968.
[12] M. Dorigo, M. Birattari, dan T. Stutzle, "Ant colony optimization," IEEE Computational Intelligence Magazine, vol. 1, no. 4, hlm. 28–39, Nov 2006, doi: 10.1109/MCI.2006.329691.
[13] M. A. Al-Betar, "$$\beta$$-Hill climbing: an exploratory local search," Neural Comput & Applic, vol. 28, no. 1, hlm. 153–168, Des 2017, doi: 10.1007/s00521-016-2328-2.

Pembahasan:
Metode Penelitian

Diagram Alir Penelitian: TRI



Dataset

1 Initial Cleaning

Total Fitur
122 Fitur

- **Numerik** — 93 Fitur
- **Kategorikal** — 30 Feature

Feature ID
- *2. TRIFD*
- *3. FRS ID*
- *41. SRS ID*
- *39. TRI CHEMICAL/COMPOUND ID*

Features with Missing Value >30%
- 25. SIC 2
- 24. PRIMARY SIC
- 27. SIC 4
- 28. SIC 5
- 29. SIC 6
- 26. SIC 3
- 35. NAICS 6
- 34. NAICS 5
- 10. BIA
- 11. TRIBE
- .....
- 16. PARENT CO DB NUM

Features With Unique Value = 1
- *54. 5.4 - UNDERGROUND*
- .....
- *108. 8.1 - RELEASES*

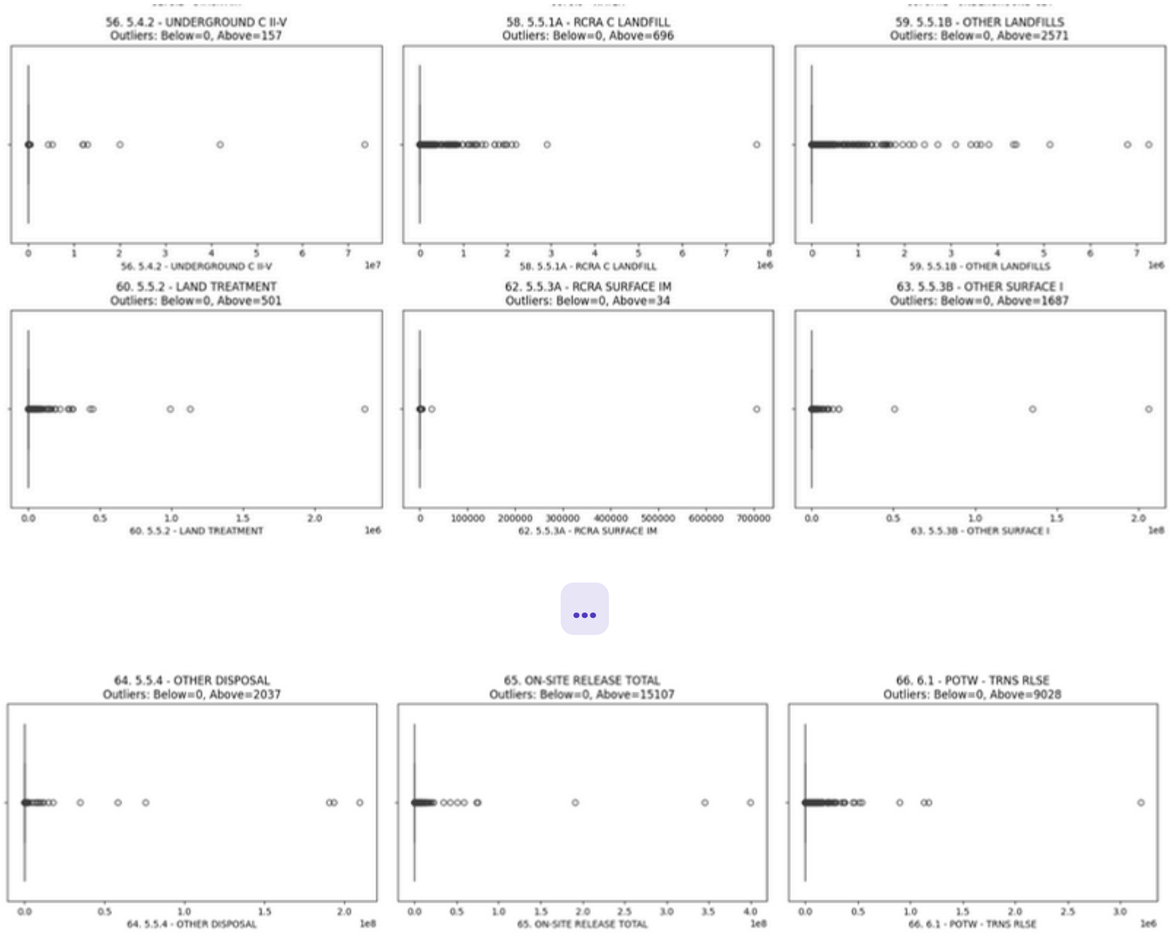
Drop
30 Fitur

Total Fitur
92 Fitur

Target Feature
119. PRODUCTION WSTE (8.1-8.7)

Analisis Data

Analisis Data

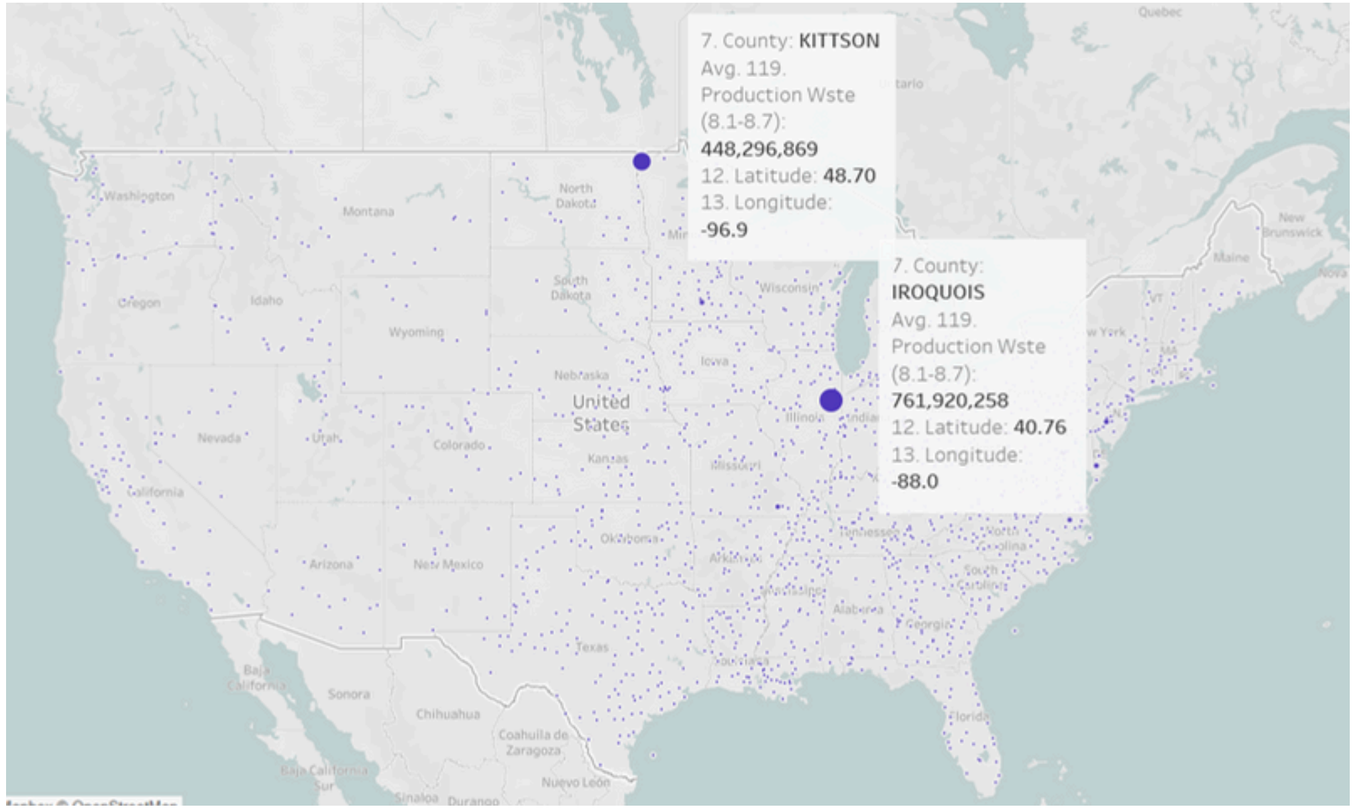


Boxplot Beberapa Fitur Numerikal

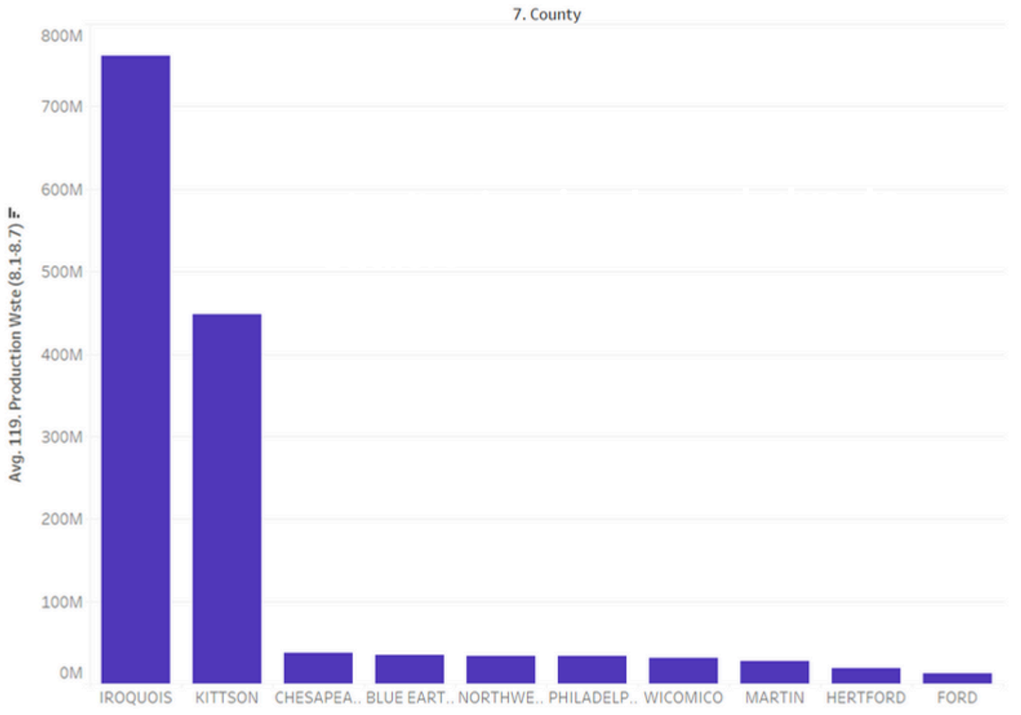
Column	Skewness	Kurtosis	Skew
62. 5.5.3A - RCRA SURFACE IM	278.662188	77748.745942	Right
79. 6.2 - M66	276.205821	76805.381032	Right
102. 6.2 - M69	272.206998	75270.849848	Right
103. 6.2 - M95	263.810544	72042.323218	Right
36. DOC_CTRL_NUM	257.714928	70059.357552	Right
99. 6.2 - M50	255.230281	68871.969347	Right
122. 8.9 - PRODUCTION RATIO	246.895126	65431.991366	Right
51. 5.1 - FUGITIVE AIR	218.642440	55137.748910	Right
117. 8.6 - TREATMENT ON SITE	207.447434	51596.538275	Right
85. 6.2 - M90	193.044943	40812.818163	Right

Skewness dan Kurtosis

1 Analisis Terhadap County

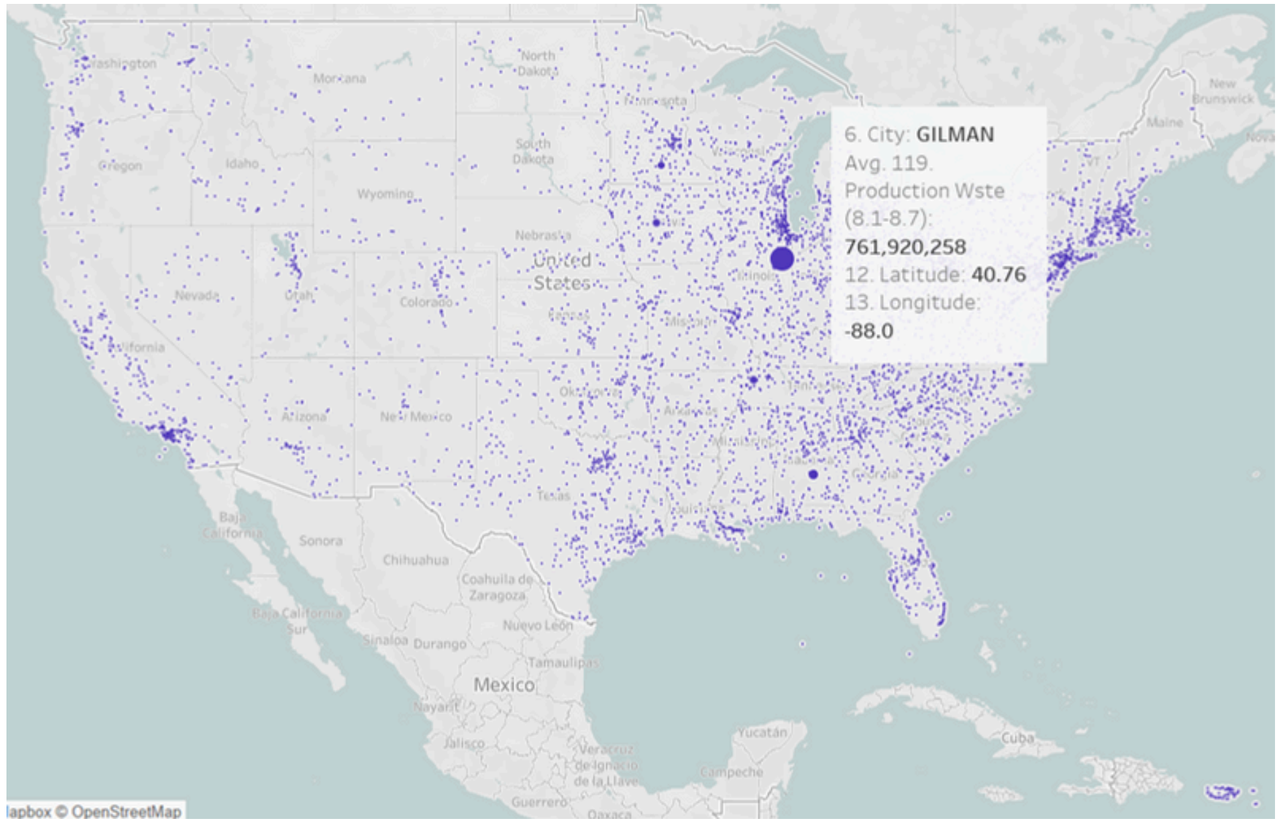


Rata-Rata Production Waste tiap County

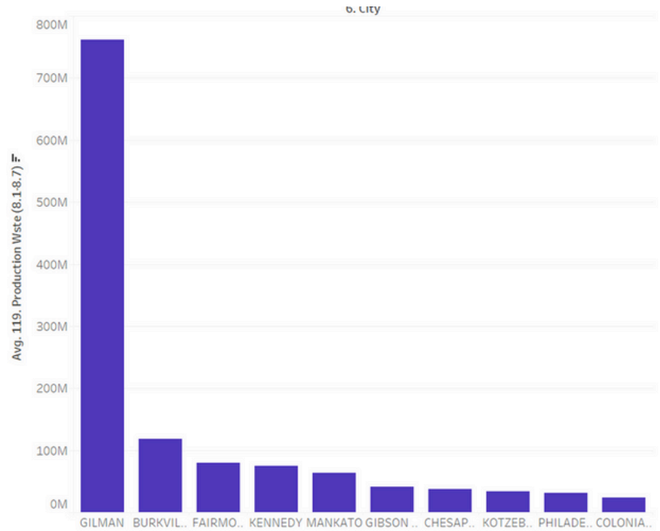


10 County dengan Rata-Rata Production Waste Tertinggi

1 Analisis Terhadap City

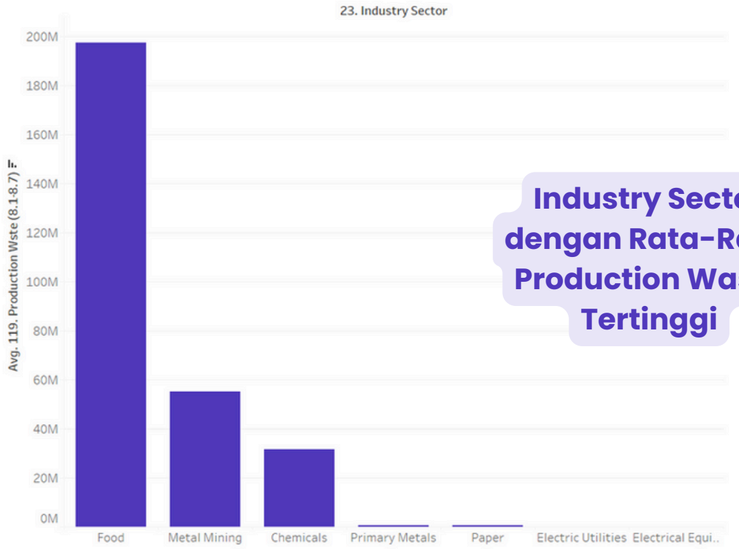


Rata-Rata Production Waste tiap City

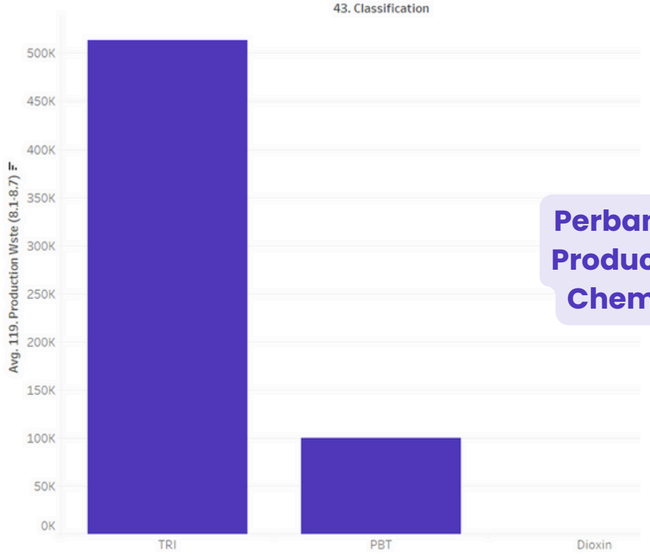


10 City dengan Rata-Rata Production Waste Tertinggi

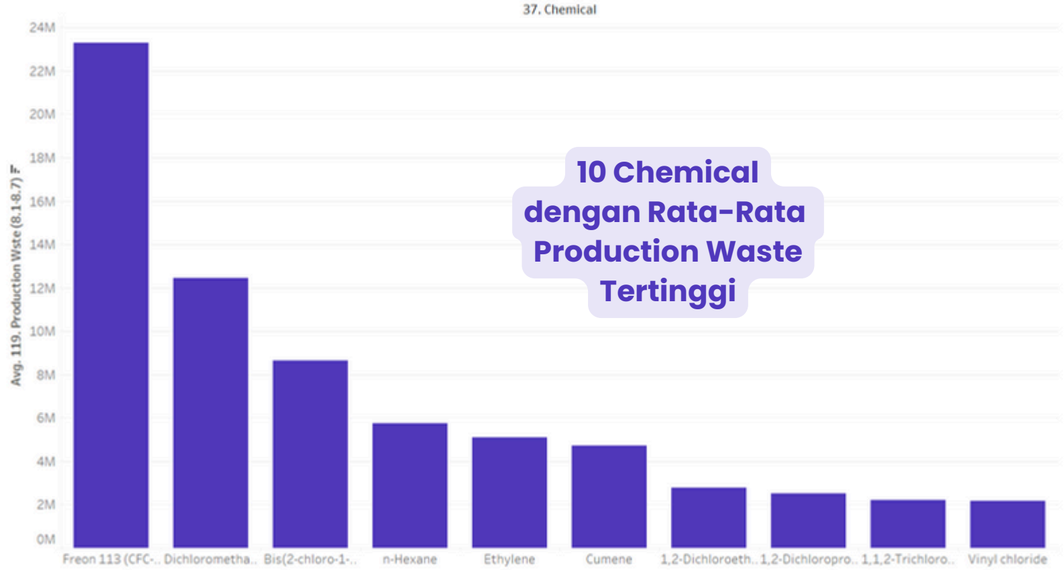
1 Analisis terhadap Fitur Kategorikal



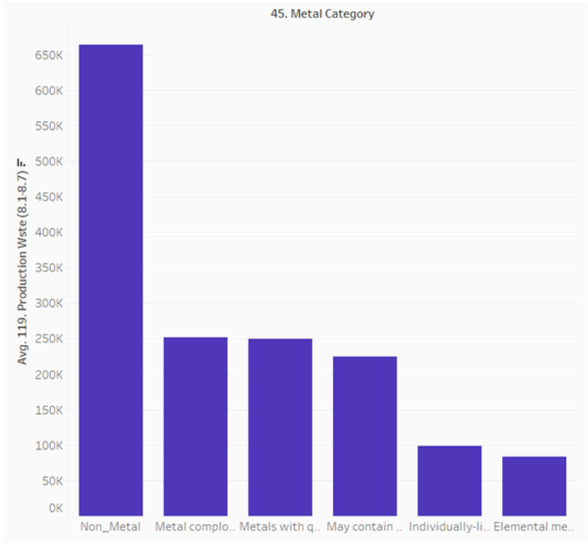
Industry Sector dengan Rata-Rata Production Waste Tertinggi



Perbandingan Rata-Rata Production Waste dengan Chemical Classification

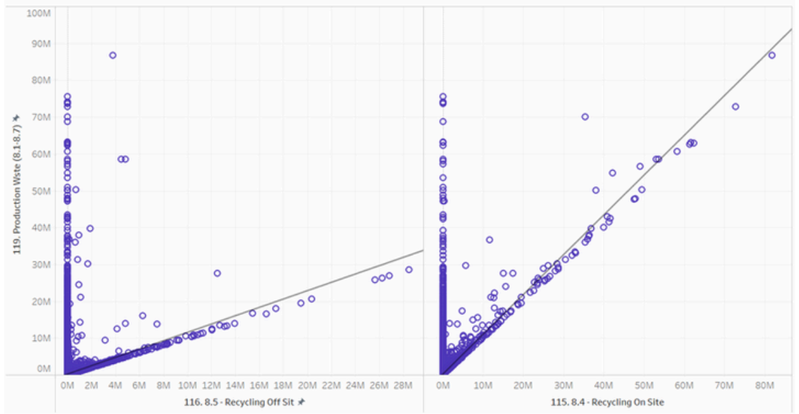


10 Chemical dengan Rata-Rata Production Waste Tertinggi

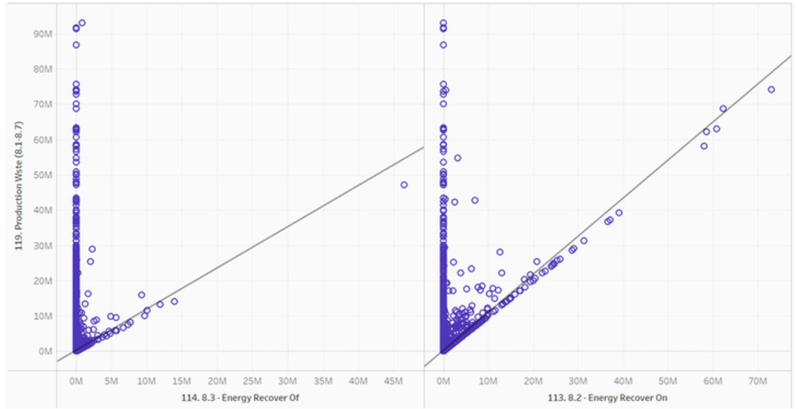


Perbandingan Production Waste dengan Metal Category

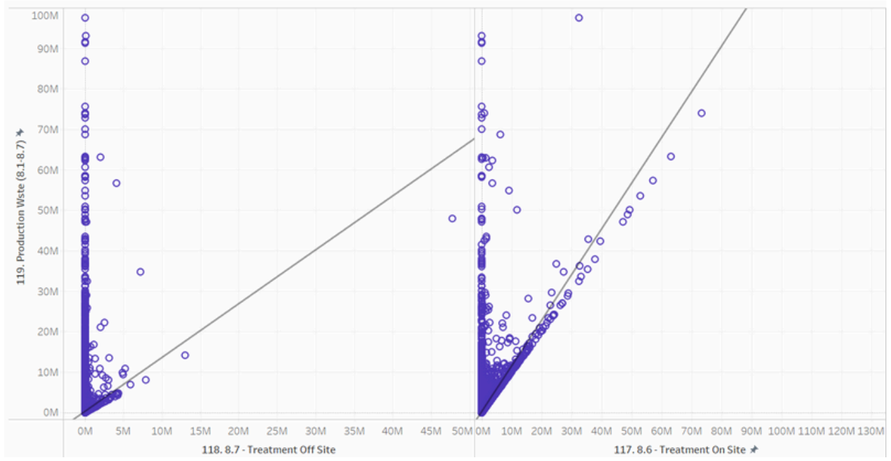
1 Analisis terhadap Kolom Numerikal



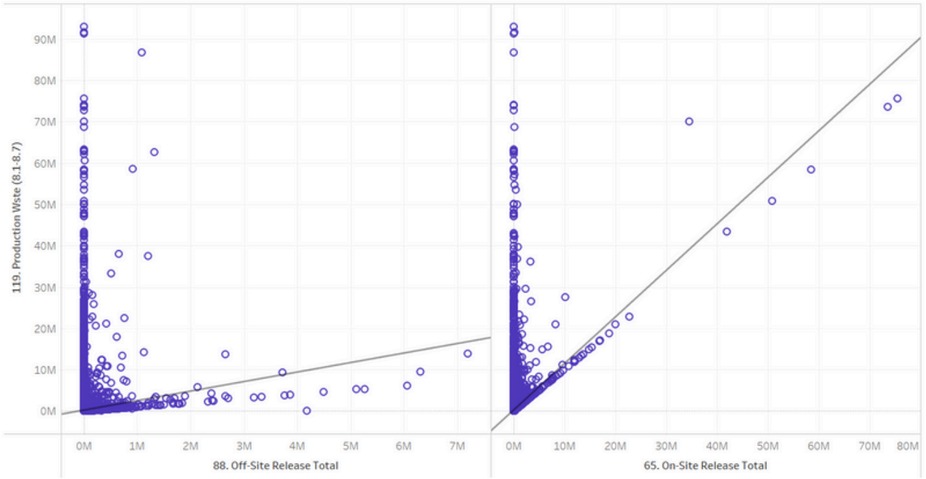
Perbandingan Recycling Site dengan Production Waste



Perbandingan Energy Recover dengan Production Waste

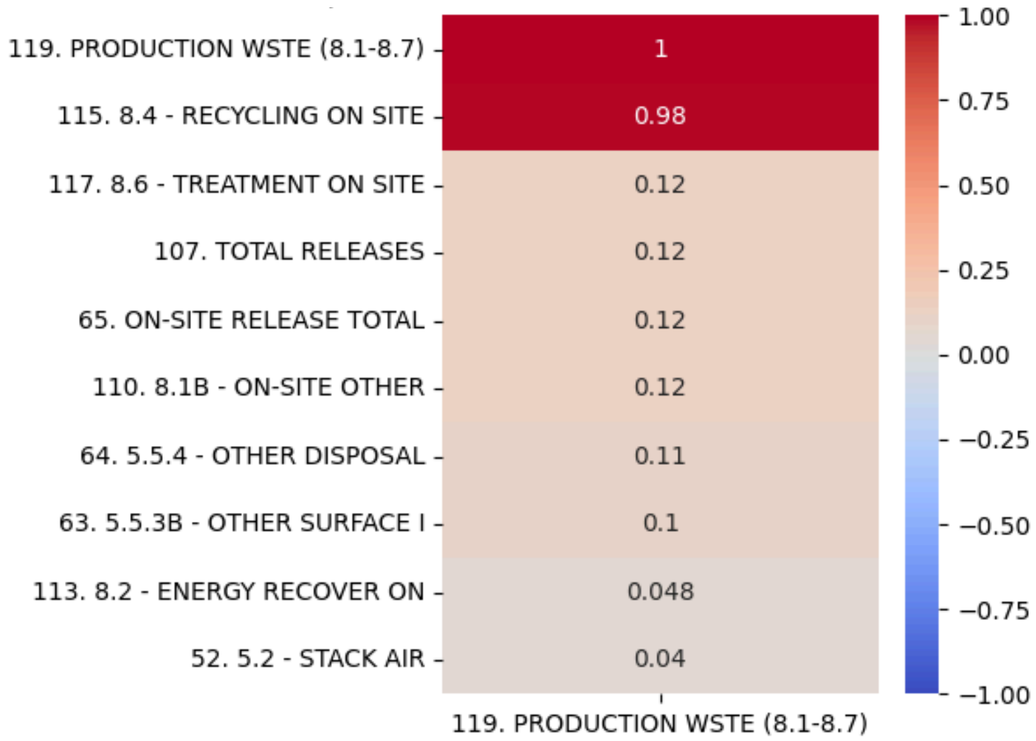


Perbandingan Site Treatment dengan Production Waste

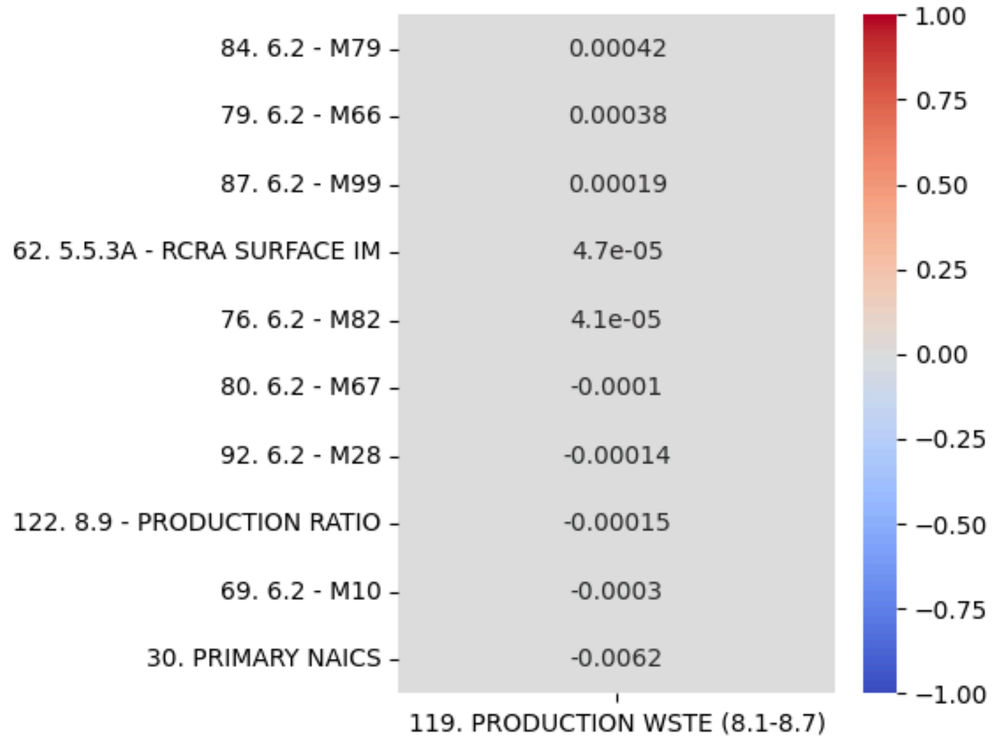


Perbandingan Site Release dengan Production Waste

Korelasi Terhadap Target



10 Fitur dengan Korelasi Tertinggi terhadap Production Waste

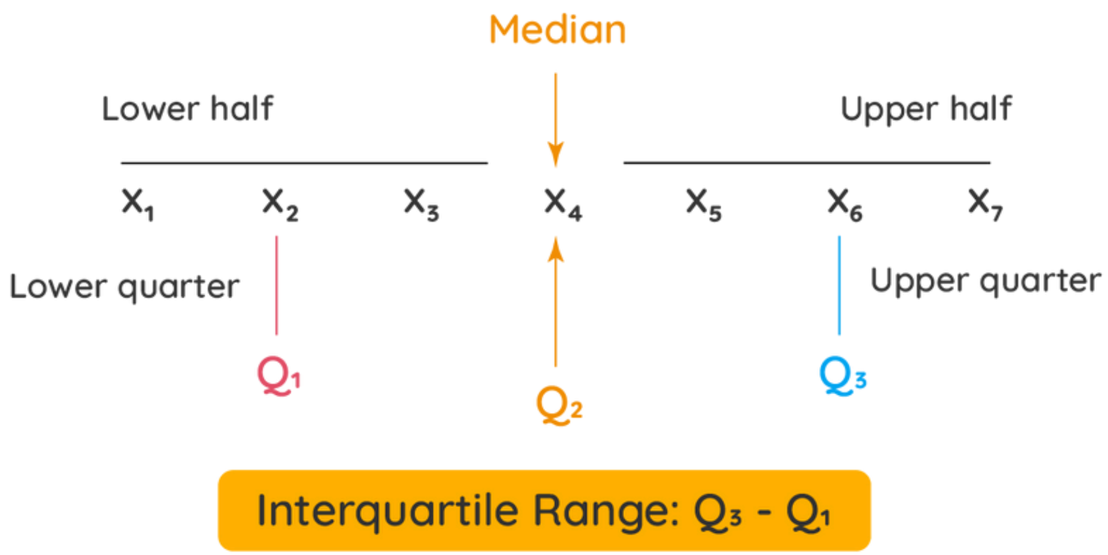


10 Fitur dengan Korelasi Terendah terhadap Production Waste

Data Cleaning

1 Data Cleaning

Outliers



Skewness

Standard Scaler

$$x' = \frac{x - \bar{x}}{\sigma}$$

Missing Value

Feature	Metode
Numerik	KNN, Mean, Median
Kategorikal	Modus

Feature Engineering

1 Feature Engineering

Penggabungan

Feature	Deskripsi
Total Waste	Jumlah Seluruh *waste*
Total Emission	Jumlah Seluruh emisi
On-Site Waste Disposal	Jumlah Seluruh *Onsite Waste*
Off-Site Waste Disposal	Jumlah Seluruh Offsite Waste
Total POTW Release	Jumlahan POTW
Total Releases	POTW + All Releases
Energy Recovery to Reclycing	Jumlahan pada Energy Recovery
Total M-Series	Jumlah Seluruh M-series
Average M Series	Rata-rata M-Series

Feature	Description
Total Off-Site Releases	Sum of All Off Site Releases
Energy Recovery to Total Releases	(Energy Recovery on + Energy Recovery off)/Total Releases
Total Transfers to Waste	Sum of All Transfer
Total Off-Site Treatment	Sum of All Offside Releases
Total POTW Release	Sum of POTW
Mean Waste	Mean of All Waste
Median Emission	Mean of All Emission
Cumullative Waste	Cummulative of Total Waste
Cummulative Emission	Cummulative of Emission

1 Feature Engineering

Fitur Rasio

Feature	Description
Landfill to Surface Ratio	Rata-rata seluruh *Waste*
On-Site Treatment Ratio	Rata-rata seluruh *Emission*
Emissions to Waste per Carcinogen	Emission/Carcinogen
Cummulative Emission	Cummulative of Emission

Feature	Description
On-Site to Off-Site Ratio	Offsite/Onsite
Emissions to Waste Ratio	Total Emission/ Total Waste
Emissions to Waste per Carcinogen	Total Emission/Total Waste * Carcinogen
Emissions to Waste per PBT	Total Emission/Total Waste * PBT
Emissions to Waste per PFAS	Total Emission/Total Waste * PFAS

Transformasi Data

Feature	Description
Semua fitur Numerik	Transformasi Standard Scaler
Semua Fitur Kategorikal	Encoding dengan Label Encoder

1 Feature Engineering

Grouping

Feature	Description
Category Chemical	Kelompok Zat Kimia berdasarkan Senyawa
Hazard Chemical	Kelompok Zat Kimia Berdasarkan Bahaya
Risk Level	Kelompok Zat Kimia Berdasarkan Tingkat Resiko
Risk Score	Nilai Tingkat Resiko
Hazard Score	Nilai Tingkat Bahaya

Fitur Baru

Feature	Description
+	Terkandung
Metal Waste Disposal	Jika Zat yang Dibuang Mengandung Logam
Non-Metal Waste Disposal	Jika Zat yang Dibuang adalah Non-Logam
Clean Air Act Chemical	Jika Udara Bersih
Risk Factor	Jika Zat Karsinogen/PBT/PFAS ada
Elemental Metal	Jika Terdapat Zat Logam

Feature Engineering

Flagging

Feature	Description
Zero Flag	if M Series Contain 0
High Values	If Emission and Waste is in High Volume
Emissions to Waste per PFAS	Emission/PFAS

Binning

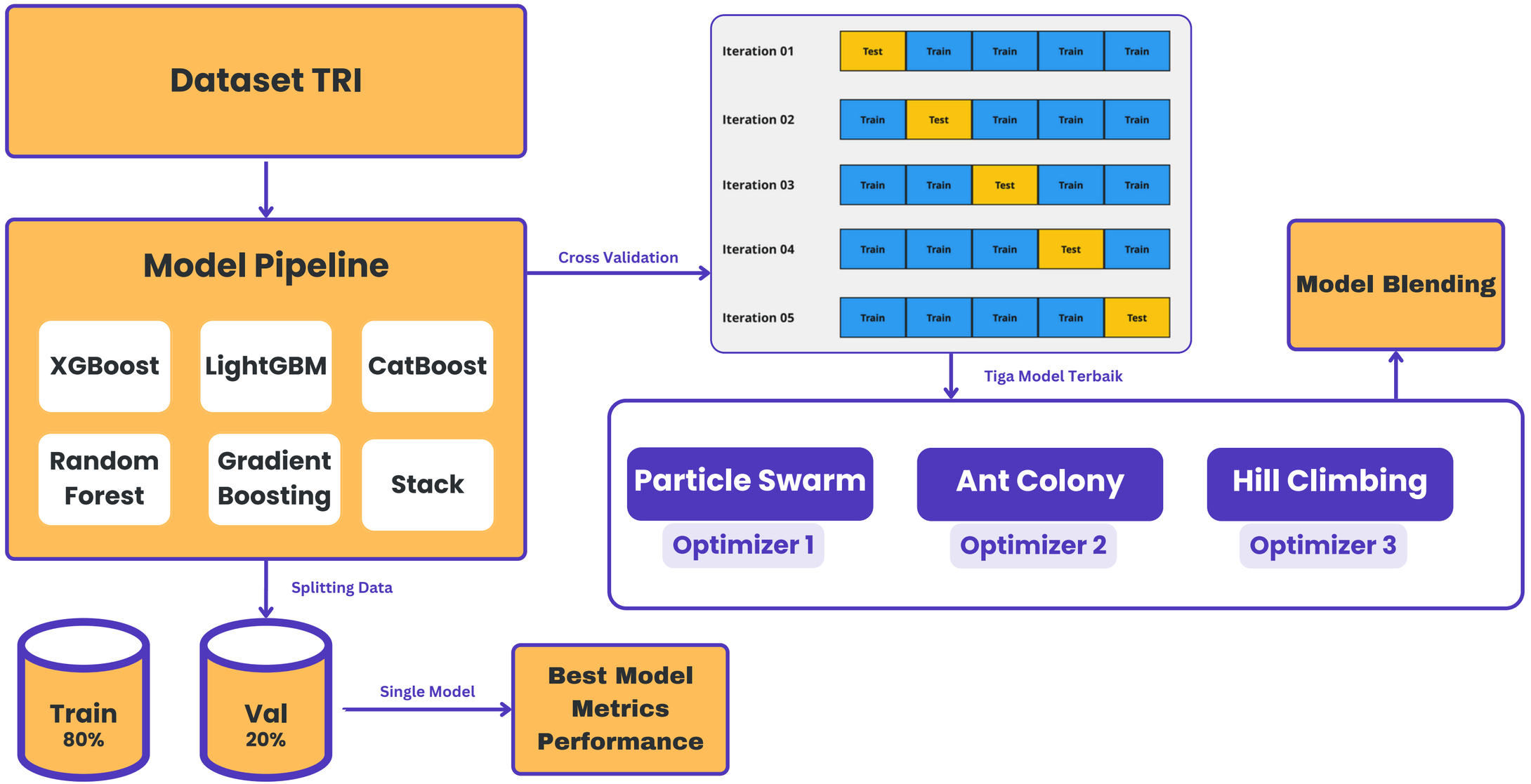
Feature	Description
Waste Binning	Binning Waste
Emission Binning	Binning Emission

Flagging

Feature	Description
+	Contained
Metal Waste Disposal	If Disposal is Metal
Non-Metal Waste Disposal	If Disposal is Not Metal
Clean Air Act Chemical	If air is Clean
Risk Factor	If Carcinogen or PBT or PFAS Present
Elemental Metal	If Elemental Metal Present
Risk Factor x Total Waste	Risk Factor * Total Waste

Pemodelan

Model TRI


Dataset TRI
Model Pipeline
XGBoost
LightGBM
CatBoost
Random Forest
Gradient Boosting
Stack
Cross Validation
Iteration 01
Iteration 02
Iteration 03
Iteration 04
Iteration 05
Test
Train
Tiga Model Terbaik
Model Blending
Particle Swarm
Optimizer 1
Ant Colony
Optimizer 2
Hill Climbing
Optimizer 3
Splitting Data
Train
80%
Val
20%
Single Model
Best Model
Metrics
Performance

Hasil dan Pembahasan

1 Kinerja Model

Sebelum FE

Model	MAE	RMSE	R2
XGBoost	112,32	552,93	0,998
Random Forest	108,77	685,28	0,997
LightGBM	208,41	912,22	0,995
Gradient Boosting	**120,45**	**500,02**	**0,996**
CatBoost	**113,45**	**765,10**	**0,996**
Stacking	**105,55**	**505,5**	**0,99**

Setelah FE

Model	MAE	RMSE	R2
XGBoost	91,24	510,52	0,997
Random Forest	73,14	601,10	0,997
LightGBM	180,52	700,52	0,995
Gradient Boosting	**160,52**	**470,51**	**0,998**
CatBoost	**105,52**	**650,10**	**0,997**
Stacking	**88,66**	**450,55**	**0,998**

Kinerja Optimisasi Model

PSO

Model	MAE	RMSE	R2
XGBoost	100,52	420,52	0,998
Gradient Boosting	**90,52**	**380,5**	**0,997**
Stacking	**100,5**	**470,5**	**0,99**

ANT COLONY

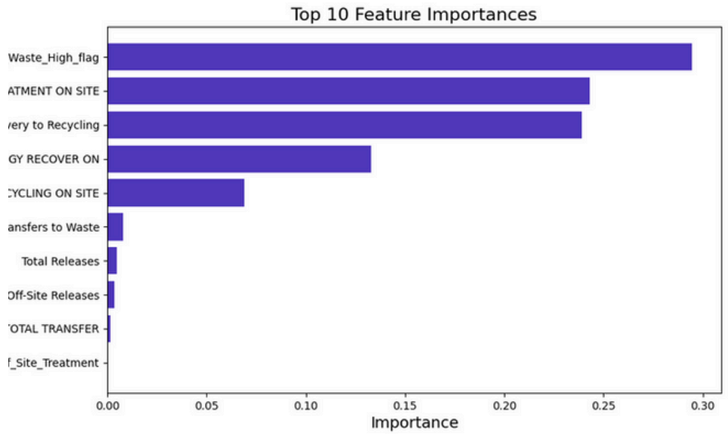
Model	MAE	RMSE	R2
XGBoost	95,56	460,52	0,997
Gradient Boosting	**86,5**	**430,44**	**0,998**
Stacking	**88,66**	**420,4**	**0,999**

Hill Climb

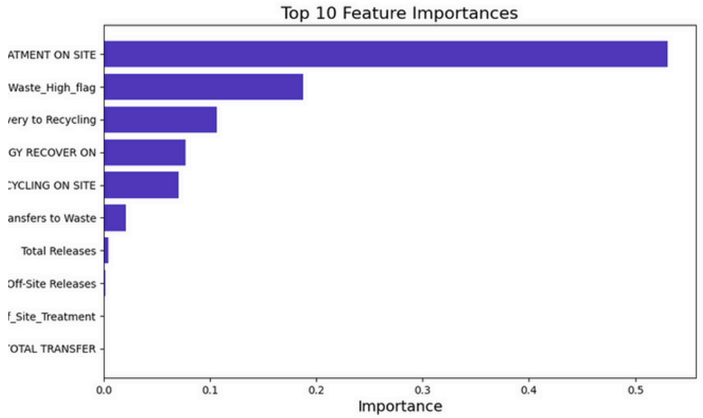
Hill Climbing Mengalami Penurunan Kinerja sangat tajam, dengan MAE 1089 dan RMSE 2164 sehingga kami memutuskan tidak menggunakan optimisasi ini

Model Visualization

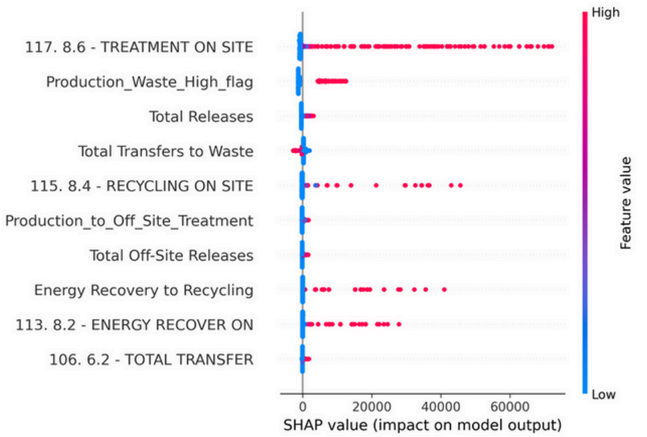
XGB + PSO

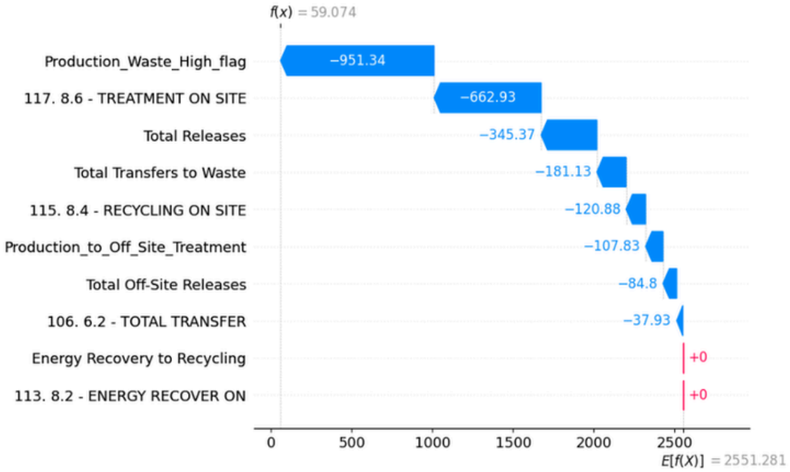


GRADIENT BOOSTING + AC



SHAP VISUALIZATION





Model Blending

	WEIGHT	RMSE
Gradient Boosting + PSO	0,6	380,5
XGBoost + PSO	0,2	420,52
STACKING + ANT COLONY	0,2	420,4

RMSE

METODE YANG DIUSULKAN (BLENDED MODEL GRADIENT BOOSTING PSO + XGBOOST PSO + STACKING AC)	MAE	RMSE
	55,0	255,12
Coefficient of Variation	1,9%	9,2%

Analisis Opini Publik Terhadap Limbah

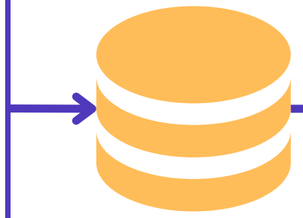
AMERICAN OPINION

APA PENDAPAT ORANG AMERIKA MENGENAI LIMBAH?

1 Scrapping X

Kata kunci yang digunakan

- *production waste*
- *toxic waste*
- *US waste problem*
- *US environment waste problem*
- *bad waste system problem*
- *reason of waste*
- *chemical waste*



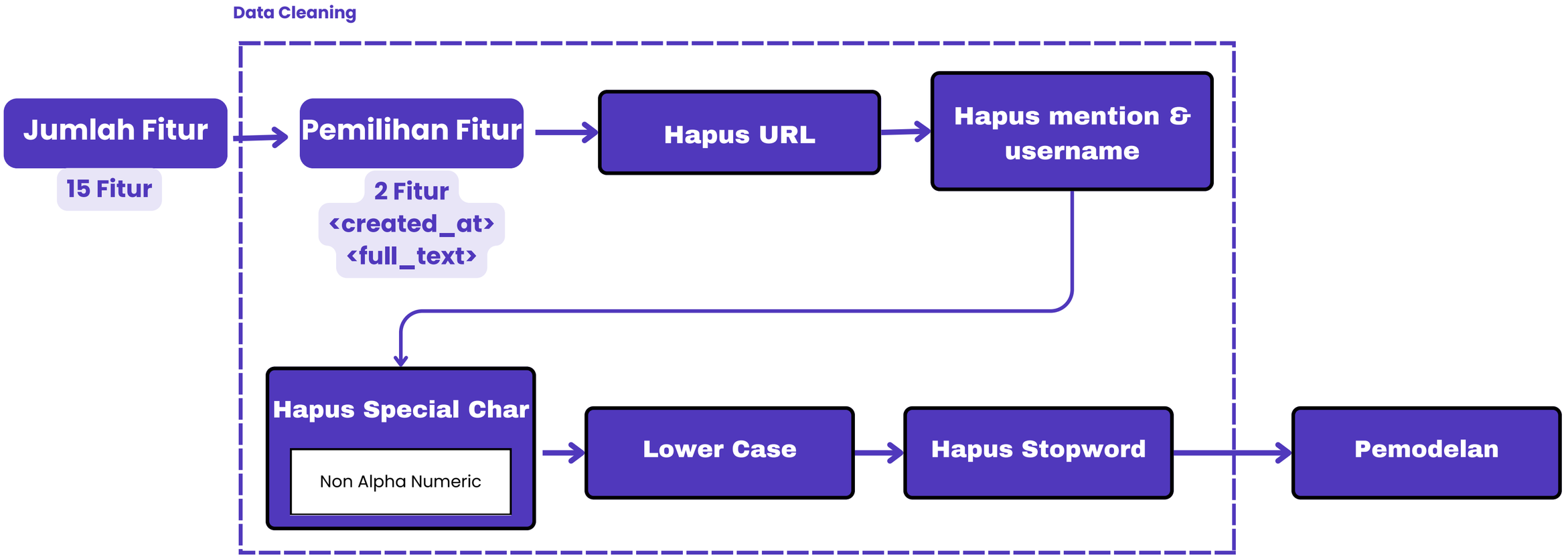
Drop Duplicate

Analisis Sentimen	*Topic Modelling*
BERTweet	BERTopic

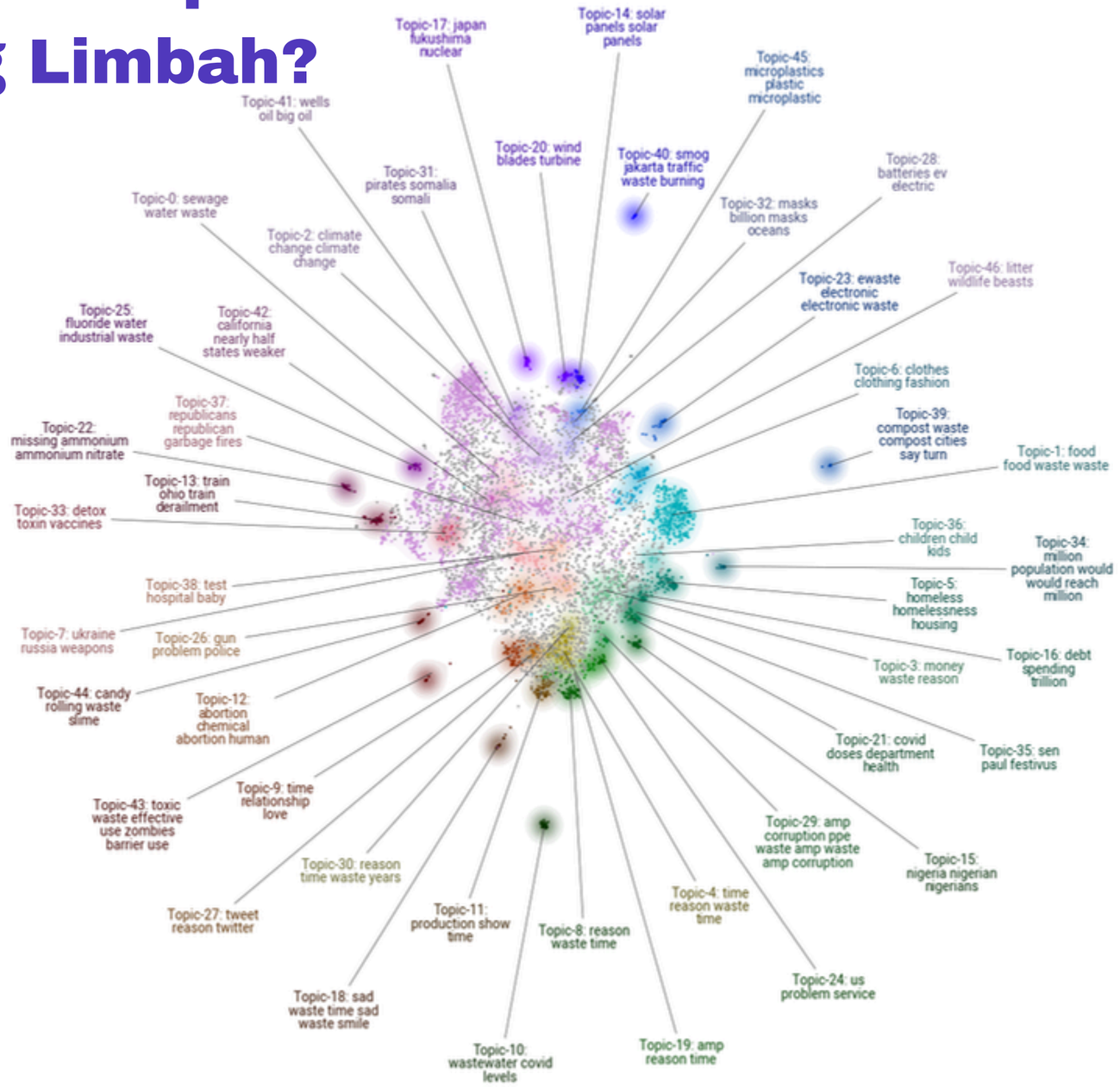
9885 Data

Rentang waktu: Jan - Des 2023

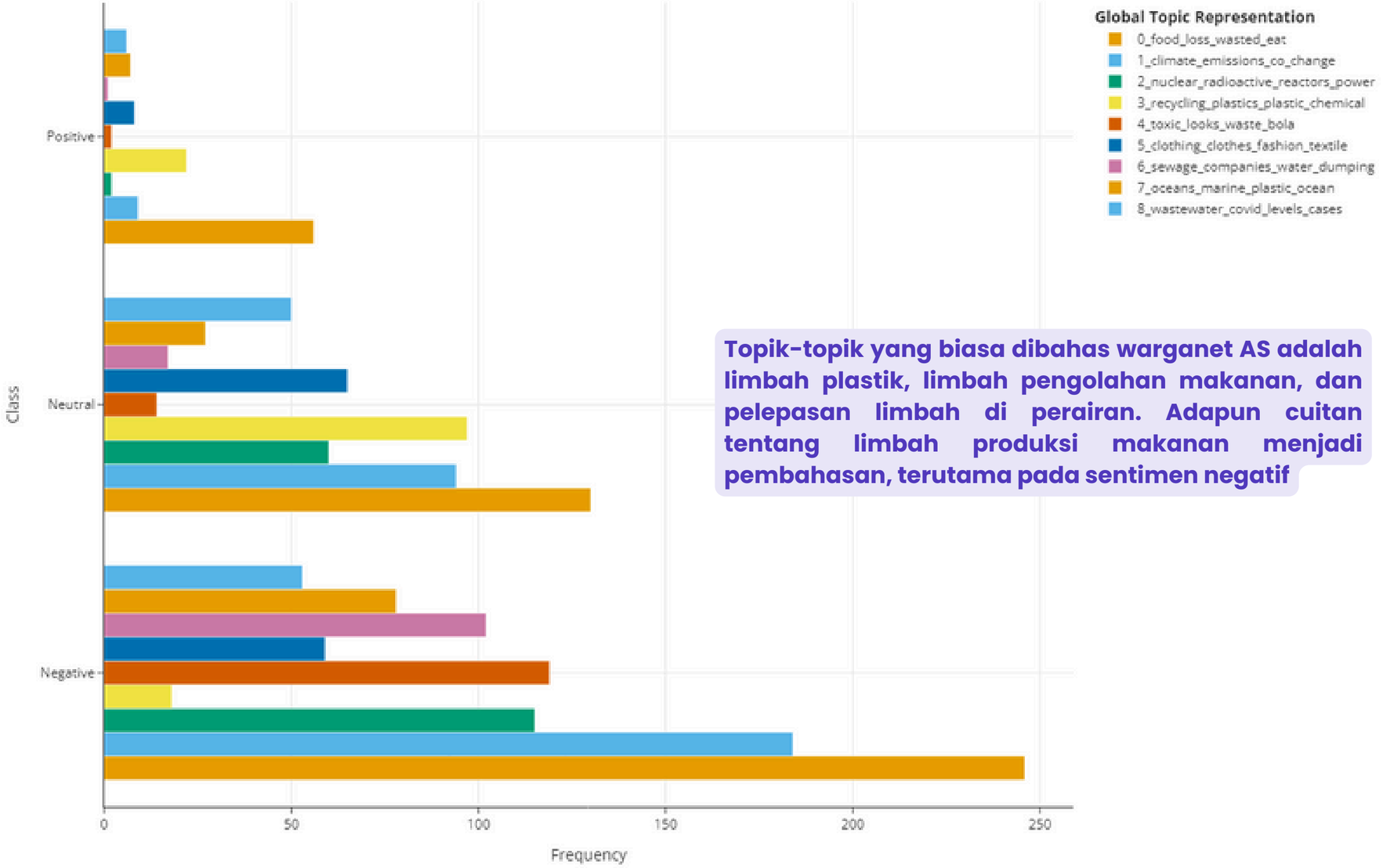
Cleaning Data Teks



Bagaimana Opini Penduduk Amerika Tentang Limbah?

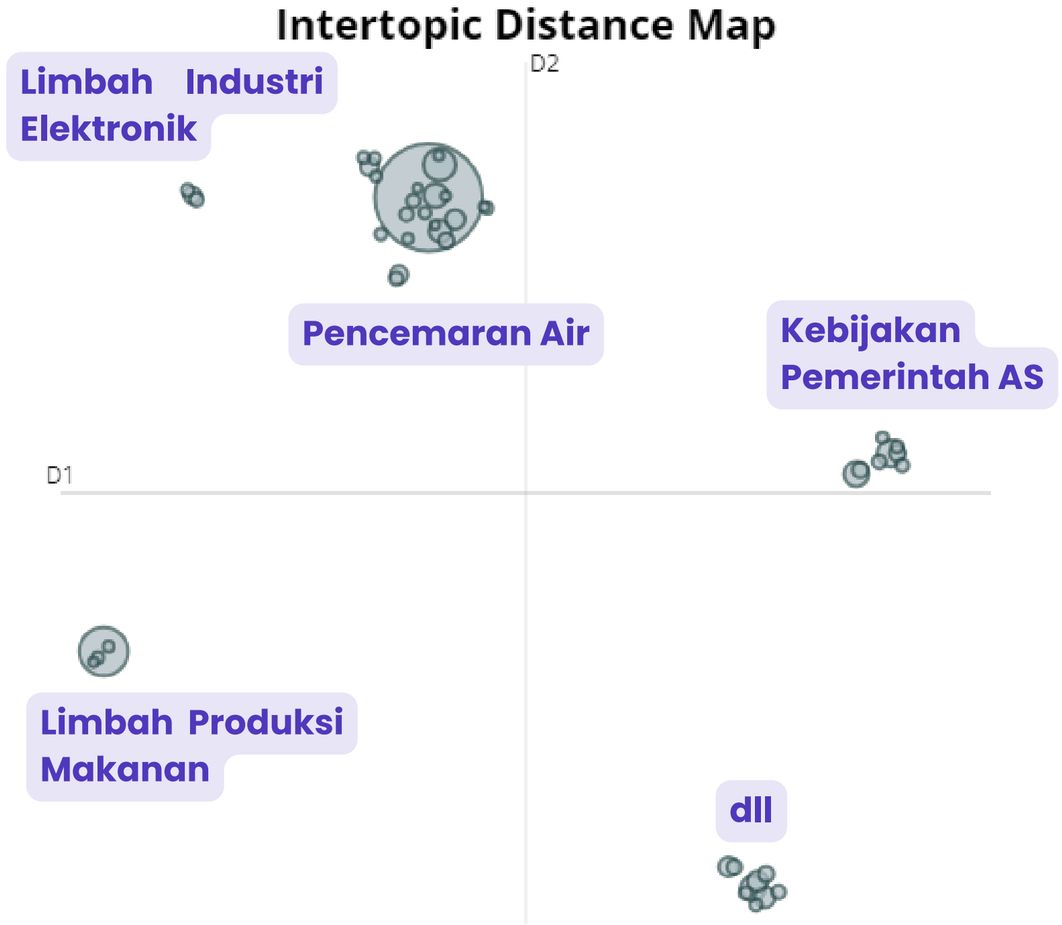


1 Bagaimana Opini Penduduk Amerika Tentang Limbah?



Topik-topik yang biasa dibahas warganet AS adalah limbah plastik, limbah pengolahan makanan, dan pelepasan limbah di perairan. Adapun cuitan tentang limbah produksi makanan menjadi pembahasan, terutama pada sentimen negatif

1 Bagaimana Opini Penduduk Amerika Tentang Limbah?



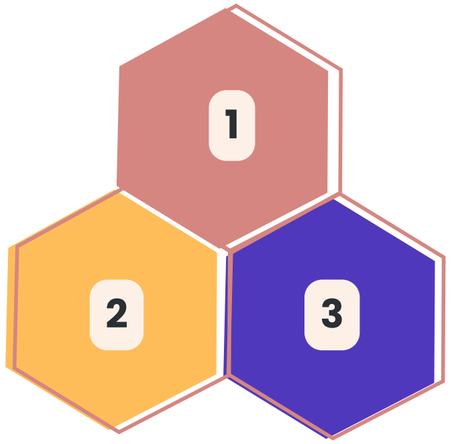
Cuitan sentimen negatif fokus dalam penyampaian keluhan terkait air yang tercemar. Beberapa limbah seperti limbah industri elektronik dan limbah produksi makanan juga disampaikan. Kebijakan pemerintah AS perlu ditinjau lagi untuk permasalahan kesehatan lingkungan.

Penutup

Kesimpulan

Metode yang diusulkan berhasil memperkirakan produksi limbah yang dihasilkan, dengan MAE=55,0

Cuitan yang dikeluhkan kebanyakan tetang ***food production waste***, sesuai dengan laporan TRI



Feature Engineering yang dilakukan berhasil meningkatkan performa model

Rekomendasi

Fokus **mengurangi** limbah pada Iroquois, Kittson, dan Gilman, limbah yang paling banyak diproduksi

Gunakan **wawasan sentimen** untuk mengomunikasikan dan melibatkan publik dalam inisiatif lingkungan, mengatasi kekhawatiran mereka dan meningkatkan kesadaran.

Mengevaluasi dan meningkatkan sistem pengelolaan limbah di luar lokasi untuk meningkatkan efektivitasnya.

Perkuat regulasi untuk material yang diklasifikasikan, terutama terkait limbah makanan

Terima Kasih

[1] A. T. Charette, M. B. Collins, dan J. E. Mirowsky, “Assessing residential socioeconomic factors associated with pollutant releases using EPA’s Toxic Release Inventory,” J Environ Stud Sci, vol. 11, no. 2, hlm. 247–257, Jun 2021, doi: 10.1007/s13412-021-00664-7.
[2] A. Marvuglia, M. Kanevski, dan E. Benetto, “Machine learning for toxicity characterization of organic chemical emissions using USEtox database: Learning the structure of the input space,” Environment International, vol. 83, hlm. 72–85, Okt 2015, doi: 10.1016/j.envint.2015.05.011.
[3] S.-R. Lim, C. W. Lam, dan J. M. Schoenung, “Quantity-based and toxicity-based evaluation of the U.S. Toxics Release Inventory,” Journal of Hazardous Materials, vol. 178, no. 1, hlm. 49–56, Jun 2010, doi: 10.1016/j.jhazmat.2010.01.041.
[4] M. M. Jobe, “The power of information: The example of the u.s. toxics release inventory**The author thanks Patricia McClure of the Government Publications Library, University of Colorado at Boulder, for her editorial assistance.,” Journal of Government Information, vol. 26, no. 3, hlm. 287–295, Mei 1999, doi: 10.1016/S1352-0237(99)00030-1.
[5] O. US EPA, “Toxics Release Inventory (TRI) Program.” Diakses: 8 Januari 2025. [Daring]. Tersedia pada: https://www.epa.gov/toxics-release-inventory-tri-program
[6] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, hlm. 5–32, Okt 2001, doi: 10.1023/A:1010933404324.
[7] G. Ke dkk., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” dalam Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017. Diakses: 8 Januari 2025. [Daring]. Tersedia pada: https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
[8] G. Ke dkk., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” dalam Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017. Diakses: 8 Januari 2025. [Daring]. Tersedia pada: https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
[9] Y. Freund dan R. E. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,” J. Comput. Syst. Sci., vol. 55, no. 1, hlm. 119–139, Agu 1997, doi: 10.1006/jcss.1997.1504.
[10] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, dan A. Gulin, “CatBoost: unbiased boosting with categorical features,” 20 Januari 2019, arXiv: arXiv:1706.09516. doi: 10.48550/arXiv.1706.09516.
[11] J. Kennedy dan R. Eberhart, “Particle swarm optimization,” dalam Proceedings of ICNN’95 - International Conference on Neural Networks, Nov 1995, hlm. 1942–1948 vol.4. doi: 10.1109/ICNN.1995.488968.
[12] M. Dorigo, M. Birattari, dan T. Stutzle, “Ant colony optimization,” IEEE Computational Intelligence Magazine, vol. 1, no. 4, hlm. 28–39, Nov 2006, doi: 10.1109/MCI.2006.329691.
[13] M. A. Al-Betar, “$$\beta$$-Hill climbing: an exploratory local search,” Neural Comput & Applic, vol. 28, no. 1, hlm. 153–168, Des 2017, doi: 10.1007/s00521-016-2328-2.
[14] J. M. Pérez dkk., “pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks,” 13 Juli 2024, arXiv: arXiv:2106.09462. doi: 10.48550/arXiv.2106.09462.
[15] D. Q. Nguyen, T. Vu, dan A. Tuan Nguyen, “BERTweet: A pre-trained language model for English Tweets,” dalam Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Q. Liu dan D. Schlangen, Ed., Online: Association for Computational Linguistics, Okt 2020, hlm. 9–14. doi: 10.18653/v1/2020.emnlp-demos.2.
[16] M. Grootendorst, “BERTopic: Neural topic modeling with a class-based TF-IDF procedure,” 11 Maret 2022, arXiv: arXiv:2203.05794. doi: 10.48550/arXiv.2203.05794.