
Transformasi Data Solarman Untuk Pengungkapan Informasi dan Pola PLTS dengan Metode Semi-Supervised Learning

M. Adhit Dwi Yuda

**Business Intelligence, Universitas Raharja
Modern, Jl. Jenderal Sudirman No.40, Cikokol, Kec. Tangerang, Kota Tangerang, Banten 15117**

adhit@raharja.info

***Corresponding Author:** adhit@raharja.info

ABSTRAK

Saat ini EBT (Energy Baru Terbarukan) merupakan peluang sekaligus tantangan bagi berbagai elemen, baik masyarakat sebagai prosumer, pemerintah sebagai regulator, hingga pengelola sistem tenaga listrik terutama PLTS (pembangkit listrik tenaga surya) di masa mendatang. Penyebaran informasi saat ini bisa melalui berbagai platform. SolarmanPV sebagai penyimpan data iot gratis yang berlokasi di Huishan district, wuxi, jiangsu, china adalah satu platform logging PLTS dunia. Permasalahan bagaimana cara menyampaikan informasi dalam meningkatkan minat penggunaan PLTS. Tujuan dari penelitian ini adalah memberikan informasi berupa Visualisasi Peta, dan Ranking penggunaan PLTS yang akurat mengenai pola informasi teselubung dari data solarmanPV. Metode data analisis terdiri dari data cleansing, data transformasi, dan visualisasi peta dengan folium, serta tambahan visualisasi peta menggunakan perangkat lunak google earth pro dan visualisasi ranking dengan perangkat lunak tableau. Pengungkapan informasi dan pola sebaran datanya secara informatif dapat meningkatkan kualitas informasi terkait penggunaan dan kapasitas PLTS terpasang di berbagai benua dan negara yang dapat di zoom out/zoom in sebagai proses rollup/drilldown.

Kata Kunci: SolarmanPV, Data cleansing, Data analisis, Visualisasi peta

ABSTRACT

Currently, NRE (New Renewable Energy) is an opportunity as well as a challenge for various elements, both the community as prosumers, the government as a regulator, to the manager of the electric power system, especially SPP (Solar Power Plant) in the future. Dissemination of information today can be through various platforms. SolarmanPV as a free IoT data store located in Huishan district, wuxi, jiangsu, china is one of solar PV logging platforms. The problem is how to convey information in increasing interest in using solar power. The purpose of this research is to provide information in the form of map visualization, and an accurate SPP usage rankings regarding the pattern of hidden information from solarmanPV data. Data analysis methods consist of data cleansing, data transformation, and map visualization with folium, as well as additional map visualization using google earth pro software and ranking visualization with tableau software. Proper disclosure of information and data distribution patterns can improve the quality of information related to the use of SPP capacity installed in various continents and countries that can be zoom out / zoom in as a rollup / drillup process.

Keywords: SolarmanPV, Data cleansing, Data analysis, Map visualization.



Adhit Yuda, M. D. Y. (2023). Transformasi Data Solarman Untuk Pengungkapan Informasi dan Pola PLTS dengan Metode Semi-Supervised Learning. *Jurnal MENTARI: Manajemen, Pendidikan Dan Teknologi Informasi*, 1(2), 100–112.
<https://doi.org/10.34306/mentari.v1i2.145>

Retrieved from <https://journal.pandawan.id/mentari/article/view/145>

Notifikasi Penulis: 19 September 2022

Akhir Revisi: 30 November 2022

Terbit: 05 Januari 2023

1. PENDAHULUAN

Indonesia mulai saat ini hingga dimasa yang akan datang akan menghadapi era EBT dimana produksi energi listrik akan banyak memanfaatkan pembangkit listrik tenaga (PLT) EBT khususnya yang dari energi surya [1]. Penetrasi PLTS kedalam sistem tenaga listrik kedepan akan memberikan beberapa peluang sekaligus tantangan di berbagai elemen, baik kepada masyarakat sebagai prosumer, pemerintah sebagai regulator, hingga pengelola sistem tenaga listrik [2].

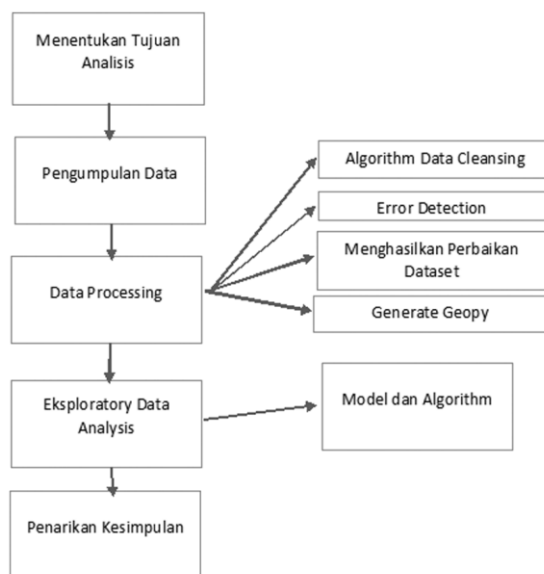
Melalui penerapan data cleaning dan transformasi , Pembersihan kualitatif selanjutnya dibagi dua langkah, yaitu deteksi kesalahan dan perbaiki kesalahan, mekanisme deteksi untuk mencari informasi dan data terselubung, kemudian digunakan untuk mengkategorikan data dalam status valid atau status tidak valid [3]. Penelitian ini akan fokus pada penggunaan langkah-langkah dalam pendekatan kualitatif ini untuk mendeteksi kesalahan pada kumpulan data terdistribusi dan kemudian menggunakan program anaconda python untuk menyelesaikan fase perbaikan kesalahan dalam proses pembersihan data[4].

Kualitas data sangat penting untuk keberhasilan setiap proyek analisis data, karena hasil analisis sebagian besar akan bergantung pada data awal yang dimasukkan ke dalam sistem analisis yang digunakan untuk mencapai hasil [5]. Kesalahan sering ditemui dalam proses akuisisi data, ini berkisar dari kesalahan yang relatif seperti nilai duplikat, kesalahan penempatan nama tabel, yang mungkin mudah untuk diatasi, hingga kesalahan yang lebih kompleks seperti nilai yang hilang atau data dalam format yang tidak dikenal [6].

Pada penelitian ini, digunakan teknik analisis data sebagai prosedur untuk menganalisis data analisis [7]. Mengacu pada pemecahan keseluruhan menjadi komponen tersendiri untuk pemeriksaan individual [8]. Analisis data adalah proses untuk menggunakan data mentah dan mengubahnya menjadi informasi yang berguna untuk pengambilan keputusan oleh pengguna[9].

2. METODOLOGI PENELITIAN

Metode atau tahapan analisis data penelitian ini dilakukan melalui lima tahapan eksperimen. Seperti dapat dilihat pada gambar 1, tahap awal analisis adalah penentuan tujuan analisis [10]. Analisis dilakukan untuk mendapatkan data pengguna PLTS (Pembangkit Tenaga Listrik) di SolarmanPV yang salah satu penyimpan data pengguna IoT gratis [11].



Gambar 1. Metode Analisis Data

Tahap kedua adalah pengambilan data yang bersumber dari website SolarmanPV yang beralamat web : www.solarmanpv.com/portal/ (repository data IoT gratis) [12]. Dikarenakan banyak faktor yang sangat penting untuk wawasan yang ingin sekali mengungkapkan informasi pada datanya. Dari data yang tersedia terdapat 15.562 baris (list) dan 17 kolom (atribut) untuk kebutuhan analisa, digunakan 16 atribut seperti yang ditunjukkan kepada Tabel 1.

Tabel 1. Daftar Kolom yang digunakan

Status				
Site Name	Dieppsolar I/II/III	Meij-Omnik-Thuis1	Regeltronics B.V.	zonnebaas
Country/State	Netherlands	Netherlands	Netherlands	Netherlands
City	Ederveen	Uithuizermeeden	Ospel	Hoogeveen
Address			Hagelkruisbaan 3	Zanderhof
System Size	7.2	1.44	7320	4255
Your own SIM card	0	0.017	0.049	0.038
Power Now	0.00	0.00	0.00	0.00
Today Energy	36.78	13.16	48.35	27.40
Total Energy	€ 6.252.056	€ 2.764.356	€ 12.499.484.825	€ 633.322.617
Total Income	0.75	2.79	0.00	1.98
Efficiency	12.09	12.36	12.23	12.08
Averaged Working Time	18/08/2012 13:57	26/08/2012 14:58	27/08/2012 13:30	13/09/2012 15:14
Comission Date	Other	Eging	RISEN	JC Solar
Panel Type	Omnik	Omnik	Omnik	Omnik
Inverter Type	my brother son and me :-) and Guppie !		Regeltronics Adviesbureau	carport mdz-oost-redasolar
Installer	Public	Public	Public	Public

Tahap ketiga adalah data processing[13]. Tujuan dari pengolahan data adalah mempersiapkan data agar siap dianalisis[14]. Pengolahan data dilakukan dengan menggunakan perangkat lunak anaconda python, aplikasi Jupyter Notebook dengan bahasa pemrograman *python* [15]. Pada tahapan ini dilakukan pemeriksaan dan penanganan terhadap data untuk kebutuhan analisis, langkah diantaranya algorithm data cleansing, error detection dan menghasilkan perbaikan data [16].

Tahap keempat adalah pemeriksaan exploratory data analysis [17]. Pada tahap ini dilakukan penyajian data secara visual menggunakan aplikasi Jupyter Notebook dengan bahasa pemrograman Python [18]. Berbagai atribut yang relevan dengan tujuan analisa dieksplorasi secara visual untuk mendapatkan pola atau anomali dari data dengan menggunakan library folium untuk membuat peta [19].

Tahap Terakhir adalah penarikan kesimpulan[20]. Pada tahapan ini disimpulkan temuan-temuan yang didapatkan pada tahapan sebelumnya [21].

3. HASIL DAN PEMBAHASAN

Penulis menerapkan sistem pembersihan data sampai visualisasi peta dengan transfer metode pembelajaran bahasa python 3.9 dalam data science menggunakan python library, secara khusus penulis memanfaatkan [22]:

1. Pandas library untuk menyimpan dan memproses data relasional
2. NumPy library untuk melakukan operasi komputasi ilmiah pada array
3. Folium library untuk memvisualisasikan data pada peta selebaran interaktif, library ini memiliki sejumlah tileset built-in dari OpenStreet, Mapbox, dan Stamen.
4. Geopy library untuk mengolah beberapa GEO API, seperti google, openstreet, baidu, ArcGis.
5. ArcGis library untuk menggunakan geoprocessing.
6. MarkerCluster library untuk menampilkan sejumlah besar penanda pada peta.
7. The jupyter notebook environment untuk membangun antarmuka pengguna yang interaktif.

3.1 Implementasi Algoritma Drop / Hapus Kolom Atribut yang tidak Valid

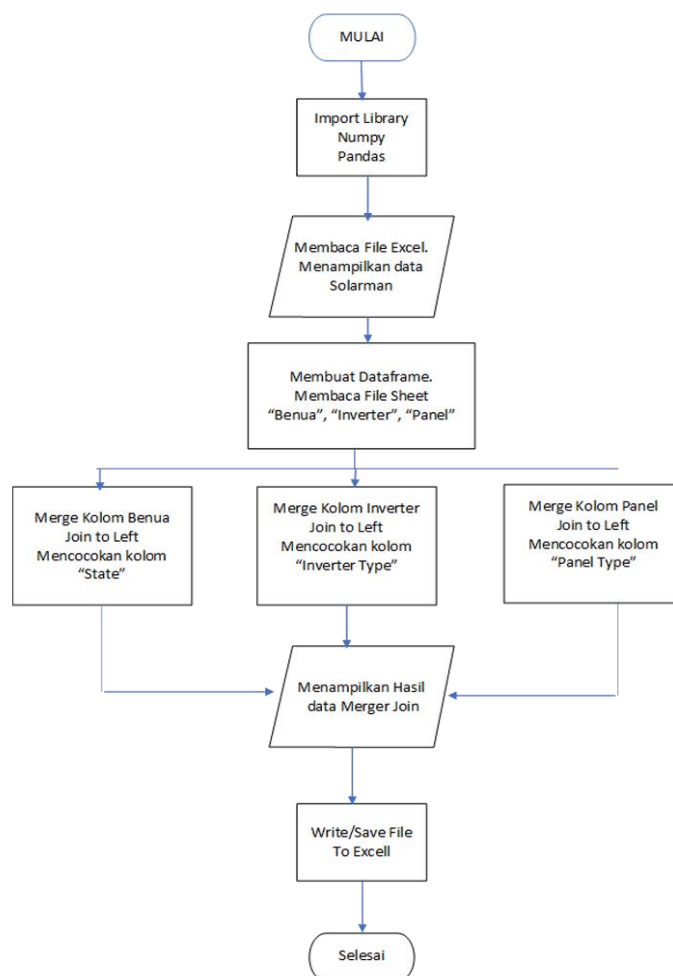
Berdasarkan struktur data pada tabel 1, terdapat pelanggaran dan data yang valid data pada atribut “Status” dan “Your own SIM card”, karena tidak ada ketertarikan informasi apa yang ingin di cari polanya pada atribut tersebut, maka peneliti melakukan pembersihan menggunakan algoritma drop, alur langkah penghapusan kolom atribut dan rename ditunjukkan pada Gambar Flowchart 1, selanjutnya peneliti melakukan rename dengan penyusunan nama atribut agar kolom atribut sesuai dengan informasi pada datanya.



Gambar Flowchart 1. Model Algoritma Drop dan Rename

3.2 Implementasi Algoritma Data Cleaning Merge Item

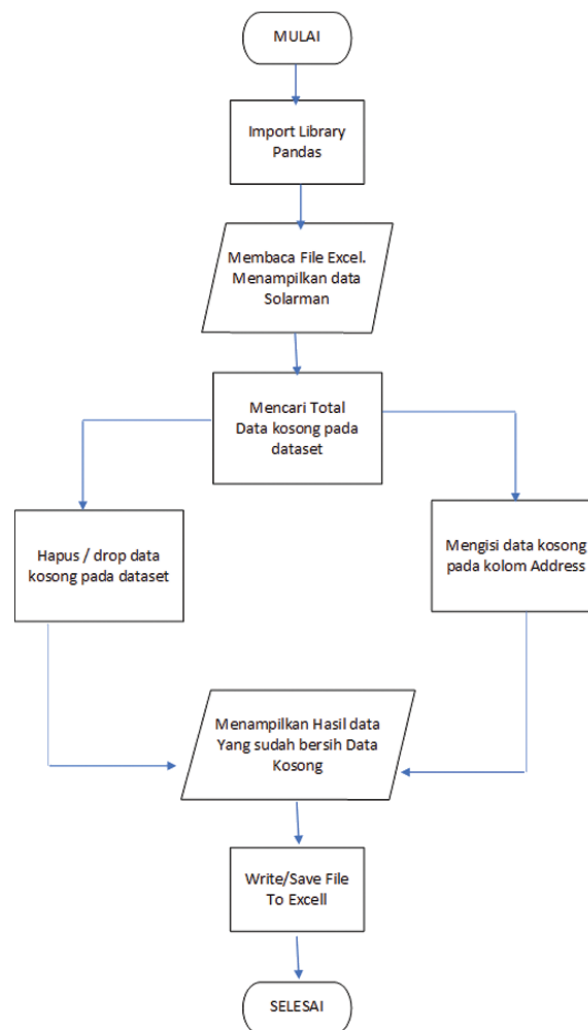
Merge kelompok item data subset dengan menambah atribut kolom baru menggunakan algoritma merge join [23]. Karena banyaknya data yang terdapat pada dataset default solarman, berjumlah 15.562 data, maka beberapa item/subset pada kolom atribut dikelompokkan. Seperti atribut state dikelompokkan menjadi atribut benua, dan atribut panel inverter type dikelompokkan menjadi jenis item yang sama untuk peringkasan, alur langkah ditunjukkan pada Gambar Flowchart 2, dan hasil merge join ditunjukkan pada Gambar 6.



Gambar Flowchart 2. Model Algoritma Merge Join Item Kelompok

3.3 Implementasi Algoritma Data Cleaning Deteksi Data Kosong dan Pembersihan Data Kosong.

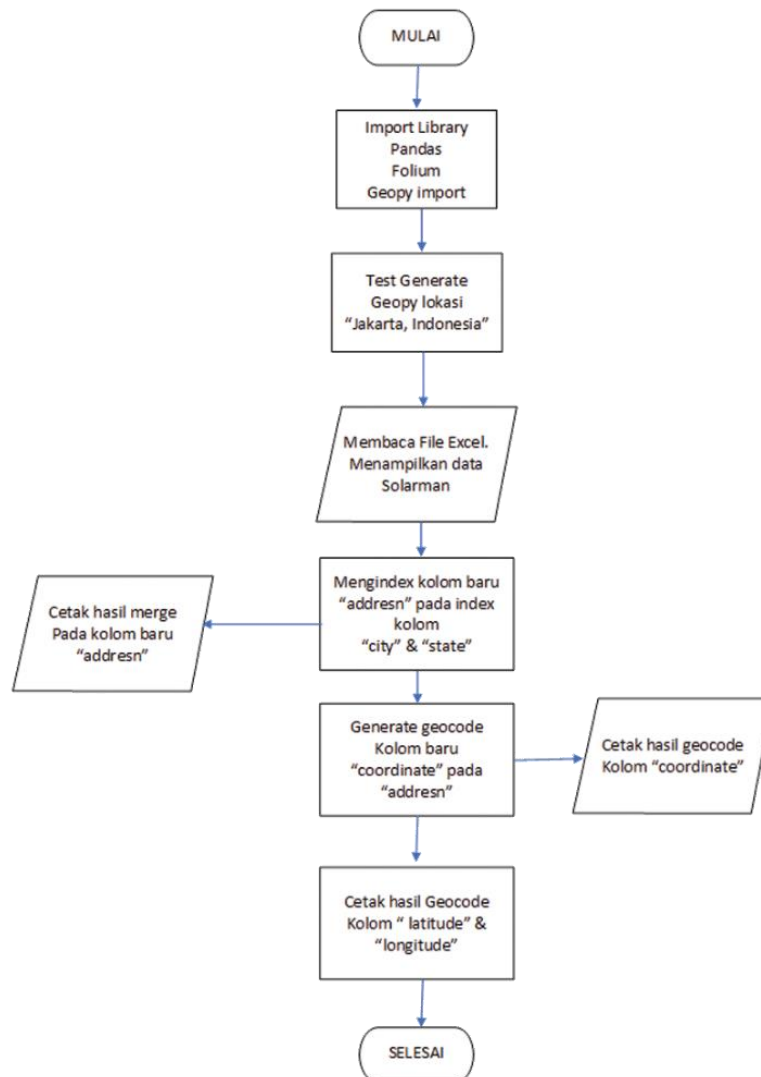
Terdapat beberapa kasus yang mengakibatkan data kosong, seperti user lupa mengisi data, data hilang tidak terekam, program error, dan lain sebagainya [24]. Maka peneliti mencari dan menangani data kosong dengan model algoritma deteksi data kosong dan pembersihan data kosong [25]. Alur langkah ditunjukkan pada Gambar Flowchart 3, algoritma deteksi data kosong dapat dilihat pada Gambar 7, Algoritma isi data kosong pada baris atribut adres ditunjukkan pada Gambar 8 [26].



Gambar Flowchart 3. Model Algoritma Deteksi dan Perbaikan Data Kosong

3.4 Implementasi Algoritma Generate Geopy

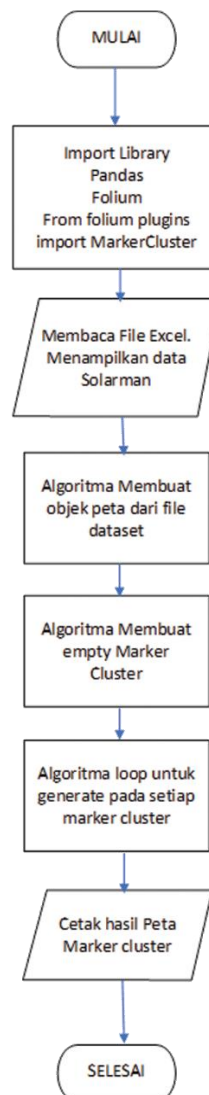
Implementasi mencari coordinate dengan algoritma geopy, pencarian diterapkan pada aplikasi Jupyter Notebook yaitu bahasa Pemrograman Python [27]. Alur langkah ditunjukkan pada Gambar Flowchart 4, dan hasil mencari koordinat latitude dan longitude ditunjukkan pada Gambar 3 [28].



Gambar Flowchart 4. Model Algoritma Generate Geopy

3.5 Implementasi Algoritma Folium untuk Visualisasi Peta Interaktif

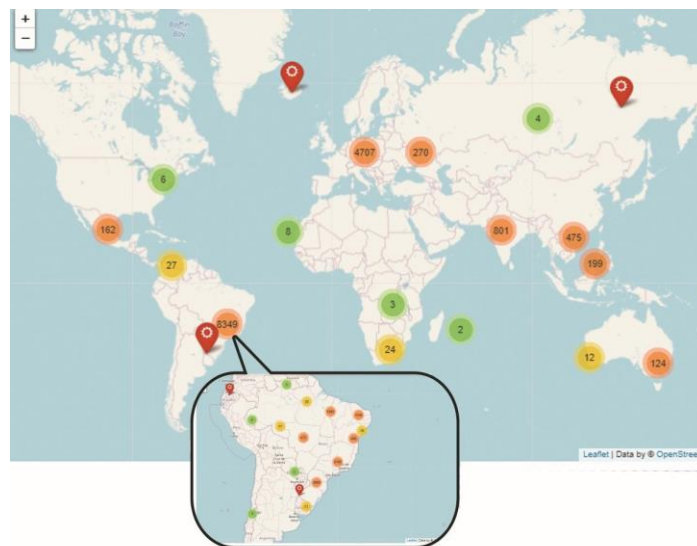
Penelitian ini memberikan pola sebaran wawasan informasi terselubung tentang data solarman berdasarkan benua dan negara [29]. Terdapat 6 Benua yang menggunakan PLTS yaitu, Amerika Utara, Amerika Selatan, Eropa, Afrika, Asia, Australia di tunjukkan pada gambar 4 [30]. Benua yang memiliki PLTS terbanyak yaitu Amerika Selatan di tunjukkan pada gambar 5, sedangkan Indonesia total pengguna yaitu 42 di tunjukkan pada gambar 6 [31]. Alur langkah di tunjukan pada Gambar Flowchart 5.



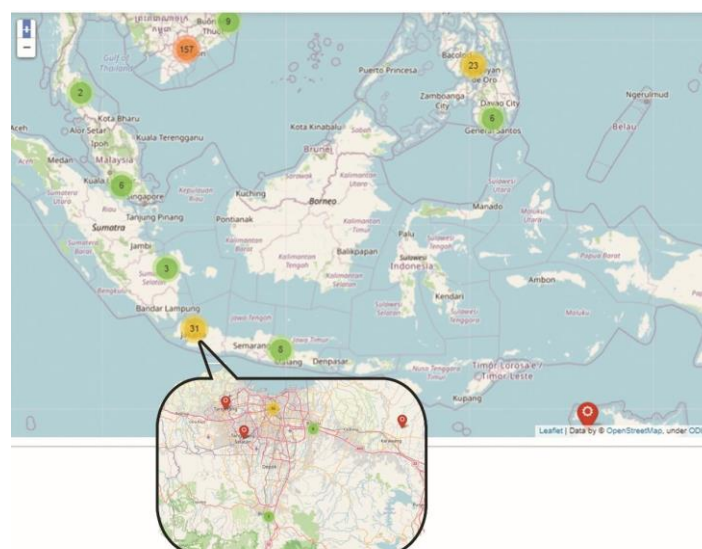
Gambar Flowchart 5. Visualisasi Peta



Gambar 4. Visualisasi Peta Dunia



Gambar 5. Visualisasi Peta Benua Amerika selatan



Gambar 6. Visualisasi Peta Indonesia

4. KESIMPULAN

Telah dilakukan data cleansing pembersihan terhadap kualitas data pada dataset solarman. dan telah dilakukan analisis menggunakan model visual untuk sebaran informasi visualisasi peta dengan bahasa pemrograman python, visualisasi peta kedua dengan software google earth pro, dan sebaran informasi ranking pengguna kapasitas plts seluruh dunia berdasarkan tahun. Dari hasil yang didapatkan, maka disimpulkan bahwa analisis data dengan menggunakan model visual, khususnya peta geospasial, sangat membantu dalam menemukan pola-pola informasi terselubung pada data solarman.

SARAN

Menurut hasil penelitian, metode data cleansing, data transformasi, dan visualisasi peta dapat mengungkapkan informasi pola sebaran data secara informatif terkait penggunaan dan kapasitas PLTS terpasang di berbagai benua dan negara. Sehingga langkah tersebut dapat dilakukan oleh peneliti lain untuk menggunakan metode tersebut dengan ruang lingkup yang berbeda.

UCAPAN TERIMAKASIH

Penulis ucapkan terimakasih kepada Universitas Raharja dan dosen yang telah mendukung penelitian ini.

DAFTAR PUSTAKA

- [1] S. Ayu Arsita, G. Eko Saputro, and S. Susanto, "Perkembangan Kebijakan Energi Nasional dan Energi Baru Terbarukan Indonesia," *Jurnal Syntax Transformation*, vol. 2, no. 12, pp. 1779–1788, Dec. 2021, doi: 10.46799/jst.v2i12.473.
- [2] M. A. Julianto *et al.*, "How to cite: LISTRIK TENAGA SURYA (PLTS) SOLAR FOTOVOLTAIK".
- [3] N. W. Utami *et al.*, "PENERAPAN DATA MINING UNTUK MENGETAHUI POLA PEMILIHAN PROGRAM STUDI DI STMIK PRIMAKARA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING."
- [4] R. I. Borman and M. Wati, "Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes."
- [5] S. Riyanto, E. Marlina, H. Subagyo, H. Triasih, and A. Yaman, "METODE PENILAIAN KUALITAS DATA SEBAGAI REKOMENDASI SISTEM REPOSITORI ILMIAH NASIONAL," *BACA: JURNAL DOKUMENTASI DAN INFORMASI*, vol. 41, no. 1, p. 11, May 2020, doi: 10.14203/j.baca.v41i1.544.
- [6] R. Takdirillah, "Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan," *Edumatic : Jurnal Pendidikan Informatika*, vol. 4, no. 1, pp. 37–46, Jun. 2020, doi: 10.29408/edumatic.v4i1.2081.
- [7] K. Nisa, D. Sugiarto, and T. Siswanto, "Perancangan Data Warehouse Harga Pangan Di Wilayah Perumda Pasar Jaya," vol. 12, 2021.
- [8] M. Faeshol Umam *et al.*, "Performance Analysis of 120 kWp Grid-Connected Rooftop Solar Photovoltaic System in Central Java", doi: 10.37525/ABSTRACT.
- [9] A. R. Uin and A. Banjarmasin, "Analisis Data Kualitatif," 2018.
- [10] A. Prasetyo *et al.*, "IMPLEMENTASI DATA MINING UNTUK ANALISIS DATA PENJUALAN DENGAN MENGGUNAKAN ALGORITMA APRIORI (STUDI KASUS DAPOERIN'S) Program Studi Sistem Informasi [2]," vol. VIII, no. 2.
- [11] I. G. N. D. Paramartha, I. N. H. Kurniawan, G. B. Subiksa, and A. S. Kartika, "Arsitektur Internet of Things (IoT) Berskala Industri dengan fitur Over The Air Update," *TIERS Information Technology Journal*, vol. 2, no. 2, pp. 31–36, Dec. 2021, doi: 10.38043/tiers.v2i2.3311.
- [12] Y. Karyadi, R. Eko Indrajit, and E. Dazki, "Penerapan Arsitektur Enterprise Berkelanjutan Untuk Perusahaan Sistem Integrator Teknologi Informasi".
- [13] D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer, and B. Lakshminarayanan, "AugMix:

- A Simple Data Processing Method to Improve Robustness and Uncertainty,” Dec. 2019, [Online]. Available: <http://arxiv.org/abs/1912.02781>
- [14] H. Mao, M. Schwarzkopf, S. B. Venkatakrishnan, Z. Meng, and M. Alizadeh, “Learning scheduling algorithms for data processing clusters,” in *SIGCOMM 2019 - Proceedings of the 2019 Conference of the ACM Special Interest Group on Data Communication*, Aug. 2019, pp. 270–288. doi: 10.1145/3341302.3342080.
- [15] M. Sholeh, S. #2, and D. Andayati, “JEPIN (Jurnal Edukasi dan Penelitian Informatika) Machine Linear untuk Analisis Regresi Linier Biaya Asuransi Kesehatan dengan Menggunakan Python Jupyter Notebook,” 2022, [Online]. Available: www.data.jakarta.go.id.
- [16] A. Yuniarti, A. Yasin, and Y. A. Nugroho, “Efektifitas Algoritma Data Mining dalam Menentukan Pendorong Darah Potensial,” 2022.
- [17] N. Verbeeck, R. M. Caprioli, and R. van de Plas, “Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry,” *Mass Spectrometry Reviews*, vol. 39, no. 3. John Wiley and Sons Inc., pp. 245–291, May 01, 2020. doi: 10.1002/mas.21602.
- [18] J. Yandi, T. B. Kurniawan, E. S. Negara, and M. Akbar, “InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan Prediksi Lokasi Titik Panas Kebaran Hutan Menggunakan Model Regresion SVM (Support Vector Machine) pada Data Kebakaran Hutan Daops Manggala Agni Oki Provinsi Sumatera Selatan Tahun 2019,” vol. 6, no. 1, 2021, doi: 10.30743/infotekjar.v6i1.4101.
- [19] D. Ramandhany and Y. N. Kunang, “VISUALISASI HEAT MAP DATA KECELAKAAN DI KOTA PALEMBANG,” *Bina Darma Conference on Computer Science*.
- [20] “Visualisasi Data Dari Dataset COVID-19 Menggunakan Pemrograman Python”.
- [21] M. Waskom, “seaborn: statistical data visualization,” *J Open Source Softw*, vol. 6, no. 60, p. 3021, Apr. 2021, doi: 10.21105/joss.03021.
- [22] N. Hug, “Surprise: A Python library for recommender systems,” *J Open Source Softw*, vol. 5, no. 52, p. 2174, Aug. 2020, doi: 10.21105/joss.02174.
- [23] Y. Suhanda, I. Kurniati, and S. Norma, “Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik,” *Jurnal Teknologi Informatika dan Komputer*, vol. 6, no. 2, pp. 12–20, Sep. 2020, doi: 10.37012/jtik.v6i2.299.
- [24] Y. Bao, Z. Chen, S. Wei, Y. Xu, Z. Tang, and H. Li, “The State of the Art of Data Science and Engineering in Structural Health Monitoring,” *Engineering*, vol. 5, no. 2, pp. 234–242, Apr. 2019, doi: 10.1016/j.eng.2018.11.027.
- [25] M. P. Fernando, F. César, N. David, and H. O. José, “Missing the missing values: The ugly duckling of fairness in machine learning,” *International Journal of Intelligent Systems*, vol. 36, no. 7, pp. 3217–3258, Jul. 2021, doi: 10.1002/int.22415.
- [26] H. M. Marin-Castro and E. Tello-Leal, “Event log preprocessing for process mining: A review,” *Applied Sciences (Switzerland)*, vol. 11, no. 22. MDPI, Nov. 01, 2021. doi: 10.3390/app112210556.
- [27] U. Dereci and M. E. Karabekmez, “The applications of multiple route optimization heuristics and meta-heuristic algorithms to solid waste transportation: A case study in Turkey,” *Decision Analytics Journal*, vol. 4, Sep. 2022, doi: 10.1016/j.dajour.2022.100113.
- [28] Y. Ouali, C. Hudelot, and M. Tami, “An Overview of Deep Semi-Supervised Learning,” Jun. 2020, [Online]. Available: <http://arxiv.org/abs/2006.05278>
- [29] R. A. Widyanto, M. Hadi Avizenna, N. A. Prabowo, K. Alfata, and A. Ismanto, “Data Mining Predicts the Need for Immunization Vaccines Using the Naive Bayes Method,” *Journal of Applied Data Sciences*, vol. 2, no. 3, pp. 93–101, 2021.
- [30] T. Hidayat, “Teknologi Deteksi dan Diagnosis Kerusakan pada PLTS: Review,” *Jurnal Teknik Elektro ITP*, vol. 9, no. 1, pp. 11–18, Jan. 2020, doi: 10.21063/jte.2020.3133903.
- [31] B. Winardi, A. Nugroho, and E. Dolphina, “Perencanaan Dan Analisis Ekonomi Pembangkit Listrik Tenaga Surya (PLTS) Terpusat Untuk Desa Mandiri,” vol. 16, no. 2, p. p-ISSN, 2019.