

Pemanfaatan Data Mining untuk Analisis Keputusan Perizinan Tenaga Kesehatan

Diterima:

10 Juni 2024

Revisi:

10 Juli 2024

Terbit:

1 Agustus 2024

¹Fikri Ardyansyah, ²Erna Daniati, ³Aidina Ristyawan

¹⁻³*Universitas Nusantara PGRI Kediri*

¹fikriardyansyah885@gmail.com, ²ernadaniati@unpkediri.ac.id,

³adinaristi@unpkediri.ac.id

Abstrak— Proses perizinan tenaga kesehatan merupakan langkah penting dalam menjaga standar kualitas dan keamanan pelayanan kesehatan. Dalam menghadapi tantangan evaluasi aplikasi yang kompleks dan volume yang tinggi, pemanfaatan teknologi Data Mining menjadi semakin relevan. Penelitian ini menyelidiki pemanfaatan Data Mining, khususnya algoritma Random Forest, untuk menganalisis keputusan perizinan tenaga kesehatan. Dataset yang digunakan mencakup data perizinan yang diterima dan ditolak, dengan berbagai atribut relevan. Proses analisis melibatkan pembacaan dan penggabungan data, penanganan nilai hilang, konversi nilai teks ke numerik, pembagian data, pembuatan model, dan evaluasi model. Hasil penelitian menunjukkan bahwa model Random Forest mencapai akurasi 97%, dengan presisi 95%, recall 97%, dan F1-score 96%. Selain itu, analisis pentingnya fitur mengungkap kontribusi masing-masing fitur dalam membuat prediksi. Dengan demikian, penelitian ini memberikan wawasan baru yang mendukung pengambilan keputusan yang lebih baik dalam proses perizinan tenaga kesehatan, serta memastikan pelayanan kesehatan yang berkualitas dan aman bagi masyarakat.

Kata kunci : *Random forest, data mining, perizinan kesehatan*

Abstract— The licensing process for health workers is an important step in maintaining quality and safety standards for health services. In facing the challenges of evaluating complex applications and high volumes, the use of Data Mining technology is becoming increasingly relevant. This research investigates the use of Data Mining, specifically the Random Forest algorithm, to analyze health worker licensing decisions. The dataset used includes data on accepted and rejected permits, with various relevant attributes. The analysis process involves reading and merging data, handling missing values, converting text to numeric values, data partitioning, model building, and model evaluation. The research results show that the Random Forest model achieves 97% accuracy, with 95% precision, 97% recall, and 96% F1-score. Additionally, feature importance analysis reveals the contribution of each feature in making predictions. Thus, this research provides new insights that support better decision making in the health worker licensing process, as well as ensuring quality and safe health services for the community.

Keywords: *Random forests, data mining, health licensing*

This is an open access article under the CC BY-SA License.



Penulis Korespondensi:

Erna Daniati,
Sistem Informasi,
Universitas Nusantara PGRI Kediri,

Email: ernadaniati@unpkediri.ac.id
ID Orcid: [<https://orcid.org/0009-0008-9471-4421>]
Handphone: 081335242202

I. PENDAHULUAN

Proses perizinan tenaga kesehatan adalah langkah penting dalam menjaga standar kualitas dan keamanan dalam pelayanan kesehatan. Dalam konteks ini, perizinan tidak hanya sekadar menjadi proses administratif, tetapi juga menjadi mekanisme yang sangat vital untuk mengontrol praktik medis dan memastikan bahwa mereka yang terlibat dalam praktik kesehatan memiliki kualifikasi dan kompetensi yang memadai [1]. Proses perizinan ini melibatkan evaluasi aplikasi yang diajukan oleh para profesional kesehatan, baik itu dokter, perawat, maupun tenaga medis lainnya, serta penilaian terhadap tempat praktek mereka.

Namun, proses perizinan tenaga kesehatan seringkali dihadapkan pada berbagai tantangan, terutama karena kompleksitas dari aplikasi yang diajukan serta volume yang tinggi. Setiap aplikasi harus dievaluasi dengan cermat untuk memastikan bahwa standar yang telah ditetapkan terpenuhi dan bahwa praktik kesehatan yang dijalankan aman dan berkualitas tinggi. Selain itu, aspek etis dan hukum juga menjadi pertimbangan penting dalam pengambilan keputusan perizinan [2]

Dalam menghadapi tantangan tersebut, pemanfaatan teknologi menjadi semakin relevan. Salah satu teknologi yang mendapat perhatian adalah Data Mining, yang menjanjikan dalam menganalisis faktor-faktor yang memengaruhi keputusan perizinan tenaga kesehatan. Dengan teknik ini, pola-pola tersembunyi dalam data perizinan dapat diidentifikasi, dan prediksi dapat dibuat tentang hasil perizinan dengan tingkat akurasi yang lebih tinggi [3]

Beberapa penelitian terdahulu telah menyoroti pentingnya pemanfaatan Data Mining dalam konteks perizinan tenaga kesehatan. [4] menyoroti potensi besar teknik Data Mining dalam menganalisis keputusan perizinan, khususnya dalam industri keuangan. Studi ini menyoroti signifikansi model skor kredit dalam menginformasikan keputusan kredit, yang pada gilirannya dapat diterapkan dalam konteks perizinan tenaga kesehatan untuk mengoptimalkan proses pengambilan keputusan.

Selain itu, penelitian [5] menunjukkan kontribusi Data Mining dalam mengidentifikasi pola-pola dalam data perizinan tenaga kesehatan, memungkinkan pengambilan keputusan yang lebih akurat. Temuan dari penelitian-penelitian tersebut menambah pemahaman yang lebih luas tentang manfaat teknik Data Mining dalam konteks perizinan tenaga kesehatan.

Dengan demikian, penelitian ini bertujuan untuk menyelidiki lebih lanjut pemanfaatan Data Mining untuk menganalisis keputusan perizinan tenaga kesehatan. Melalui pendekatan ini, diharapkan dapat ditemukan wawasan baru yang mendukung pengambilan keputusan yang lebih baik dalam proses perizinan tenaga kesehatan, serta memastikan pelayanan kesehatan yang berkualitas dan aman bagi masyarakat.

II. METODE PENELITIAN

Studi ini menggunakan pendekatan Data Mining untuk menganalisis dataset perizinan tenaga kesehatan. Dataset terdiri dari data perizinan yang diterima dan ditolak, yang mencakup berbagai atribut seperti informasi pemohon, tempat praktek, dan atribut lainnya yang relevan. Langkah-langkah metodologi yang digunakan dalam penelitian ini antara lain:

- 1. Pembacaan dan Penggabungan Data:** Data perizinan yang diterima dan ditolak dibaca dari file CSV terpisah dan digabungkan menjadi satu DataFrame. Proses ini memastikan bahwa semua data yang relevan terkumpul dalam satu struktur yang mudah dianalisis, memungkinkan integrasi informasi dari kedua kategori perizinan untuk analisis yang komprehensif. Serta ditambahkan atribut status untuk menentukan status diterima dan ditolak pada data yang akan diproses.
- 2. Penanganan Nilai Hilang:** Nilai-nilai yang hilang dalam dataset diisi dengan nilai konstan menggunakan SimpleImputer. Pendekatan ini menjaga konsistensi data dan mencegah bias yang mungkin muncul akibat penghapusan data yang tidak lengkap[6], sehingga memungkinkan analisis yang lebih akurat dan lengkap.

```
# Tangani nilai yang hilang dengan SimpleImputer
from sklearn.impute import SimpleImputer
imputer = SimpleImputer(strategy='constant', fill_value=0)
data_encoded_imputed = imputer.fit_transform(data)

# Kembalikan hasil imputasi ke dalam DataFrame
data = pd.DataFrame(data_encoded_imputed, columns=data.columns)
```

Gambar 1. SimpleImputer

- 3. Konversi Nilai Teks ke Numerik:** Nilai-nilai teks dalam dataset dikonversi menjadi nilai numerik menggunakan metode *factorize*[7]. Konversi ini memfasilitasi proses pemodelan dengan algoritma machine learning, yang umumnya membutuhkan input dalam bentuk numerik untuk melakukan perhitungan dan prediksi.

4. **Pembagian Data:** Data dibagi menjadi data latih dan data uji menggunakan `train_test_split[8]`, dengan ukuran data uji sebesar 30%. Pembagian ini penting untuk memastikan bahwa model yang dibangun dapat dievaluasi dengan benar dan memiliki kemampuan generalisasi yang baik terhadap data baru yang tidak terlihat selama pelatihan.
5. **Pembuatan Model:** Model klasifikasi `RandomForestClassifier` digunakan untuk membuat model perizinan tenaga kesehatan. `RandomForest` dipilih karena kemampuannya yang kuat dalam menangani dataset yang kompleks dan beragam, serta kemampuannya dalam mengatasi overfitting dengan melakukan ensemble learning dari beberapa pohon keputusan[9].

Dalam pembentukan Random Forest, setiap pohon keputusan dibangun menggunakan subset acak dari data latih dan fitur. Proses ini disebut bootstrap aggregating atau bagging. Pemilihan fitur terbaik dilakukan di setiap node, menggunakan kriteria pemisahan seperti Gini impurity atau Information Gain[10].

a. Rumus dan Kriteria Pemisahan:

1. **Gini Impurity:** Gini impurity mengukur ketidakhomogenitas suatu node. Semakin rendah nilai Gini, semakin murni node tersebut. Rumus untuk Gini impurity adalah:

$$\text{Gini}(D) = 1 - \sum_{i=1}^C p_i^2$$

di mana p_i adalah proporsi sampel dengan kelas i pada dataset D , dan C adalah jumlah kelas.

2. **Information Gain:** Information Gain mengukur penurunan entropi dari pembagian dataset. Rumusnya adalah:

$$\text{IG}(D, A) = \text{Entropy}(D) - \sum_{v \in \text{Values}(A)} \frac{|D_v|}{|D|} \cdot \text{Entropy}(D_v)$$

di mana D adalah dataset, A adalah atribut, v adalah nilai dari atribut A , dan D_v adalah subset dari D untuk nilai v .

Setiap pohon dalam hutan menghasilkan prediksi. Untuk klasifikasi, hasil prediksi digabungkan melalui voting mayoritas, sedangkan untuk regresi, hasil prediksi digabungkan dengan mengambil rata-rata.

b. Evaluasi dan Metrik:

Setelah model dilatih menggunakan `RandomForestClassifier()` dan metode `fit` dengan data latih (`X_train` dan `y_train`), prediksi dilakukan pada data uji (`X_test`)

menggunakan metode **predict**. Hasil prediksi kemudian dievaluasi menggunakan berbagai metrik kinerja:

1. **Akurasi:** Akurasi mengukur persentase prediksi yang benar. Rumusnya adalah:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

di mana TP adalah True Positives, TN adalah True Negatives, FP adalah False Positives, dan FN adalah False Negatives.

2. **Presisi:** Presisi mengukur proporsi prediksi positif yang benar. Rumusnya adalah:

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. **Recall:** Recall mengukur proporsi kejadian aktual positif yang benar-benar terprediksi positif. Rumusnya adalah:

$$\text{Recall} = \frac{TP}{TP + FN}$$

4. **F1-Score:** F1-Score adalah harmoni rata-rata dari presisi dan recall, memberikan keseimbangan antara keduanya. Rumusnya adalah:

$$F1\text{-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Model dievaluasi menggunakan metrik kinerja seperti akurasi, presisi, recall, dan F1-score[11]. Evaluasi ini memberikan pemahaman yang mendalam tentang kinerja model dalam berbagai aspek, memastikan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga efektif dalam mengidentifikasi perizinan yang diterima dan ditolak dengan benar.

III. HASIL DAN PEMBAHASAN

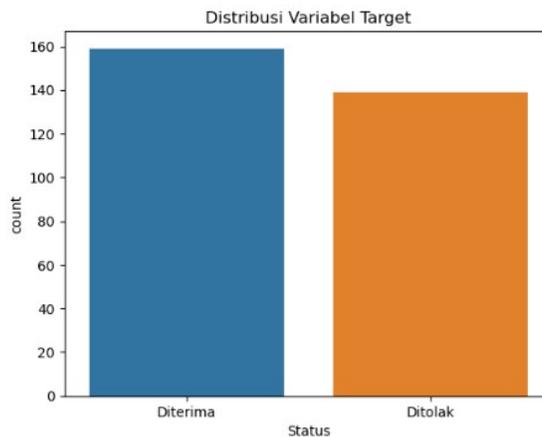
1. Gambaran umum dataset

Data izin praktik tenaga kesehatan yang terdiri dari 17 kolom, Ditunjukkan pada gambar 1.

```
Data columns (total 17 columns):
# Column Non-Null Count Dtype
---
0 No 261 non-null float64
1 Tahun Rekom 136 non-null float64
2 Jabatan 261 non-null object
3 Nama Pemohon 261 non-null object
4 Informasi Tempat Praktek/Kerja (Nama Tempat Praktik) 279 non-null object
5 Informasi Tempat Praktek/Kerja (Alamat) 279 non-null object
6 Informasi Tempat Praktek/Kerja (Praktik Ke-) 278 non-null object
7 Informasi Tempat Praktek/Kerja (No SIP) 116 non-null object
8 Informasi Tempat Praktek/Kerja (Jam Praktik) 281 non-null object
9 Tanggal Rekom OP 31 non-null object
10 Masa Berlaku 48 non-null object
11 No STR 247 non-null object
12 No Telp 247 non-null object
13 Jenis Praktik 247 non-null object
14 Jenis Izin 260 non-null object
15 No SK 146 non-null object
16 Status 298 non-null object
dtypes: float64(2), object(15)
memory usage: 39.7+ KB
```

Gambar 2. Fitur Perizinan

Gambar 2 menunjukkan distribusi variabel target berdasarkan statusnya. Dari diagram batang, terlihat bahwa jumlah variabel target yang diterima lebih banyak daripada yang ditolak.



Gambar 3 distribusi variable target

2. Preprocessing data

Preprocessing yang pertama dilakukan adalah mengisi nilai kosong pada dataset dengan library SimpleImputer, lalu mengubah data yang bertipe object menjadi int menggunakan fungsi *factoryze*.

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 298 entries, 0 to 297  
Data columns (total 17 columns):  
#   Column                                                                                               Non-Null Count  Dtype  
---  ---                                                                                               -  
0   No                                                                                                   298 non-null   object  
1   Tahun Rekom                                                                                         298 non-null   object  
2   Jabatan                                                                                             298 non-null   object  
3   Nama Pemohon                                                                                       298 non-null   object  
4   Informasi Tempat Praktek/Kerja (Nama Tempat Praktik)    298 non-null   object  
5   Informasi Tempat Praktek/Kerja (Alamat)                 298 non-null   object  
6   Informasi Tempat Praktek/Kerja (Praktik Ke-)            298 non-null   object  
7   Informasi Tempat Praktek/Kerja (No SIP)                 298 non-null   object  
8   Informasi Tempat Praktek/Kerja (Jam Praktik)            298 non-null   object  
9   Tanggal Rekom DP                                         298 non-null   object  
10  Masa Berlaku                                             298 non-null   object  
11  No STR                                                    298 non-null   object  
12  No Telp                                                   298 non-null   object  
13  Jenis Praktik                                            298 non-null   object  
14  Jenis Izin                                               298 non-null   object  
15  No SK                                                     298 non-null   object  
16  Status                                                    298 non-null   object  
dtypes: object(17)  
memory usage: 39.7+ KB
```

Gambar 4. Pengisian data

Selanjutnya mengubah data menjadi numerik agar dapat di masukkan dalam pengujian algoritma.

```
# Mengubah tipe data objek menjadi numerik  
for column in data.columns:  
    if data[column].dtype == 'object':  
  
        data[column] = pd.factorize(data[column])[0]
```

3. Pemilihan fitur

Fitur yang digunakan untuk pengujian algoritma ini adalah semua fitur yang ada di dalam dataset. Karena semua fitur memiliki pengaruh terhadap status diterima dan tidak diterima

```
# Memisahkan fitur dan target  
X = data.drop('Status', axis=1)  
y = data['Status']
```

4. Pemilihan *training* dan *testing*

Untuk pembagian data *training* dan *testing* menggunakan library *train tes split* yang membagi variable X dan variable y

```
# Pembagian data menjadi data latih dan data uji  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=12)
```

Jumlah data dalam data latih (Training Data): 208
 Jumlah data dalam data uji (Testing Data): 90

Gambar 5. Pembagian data uji dan tes

5. Random forest

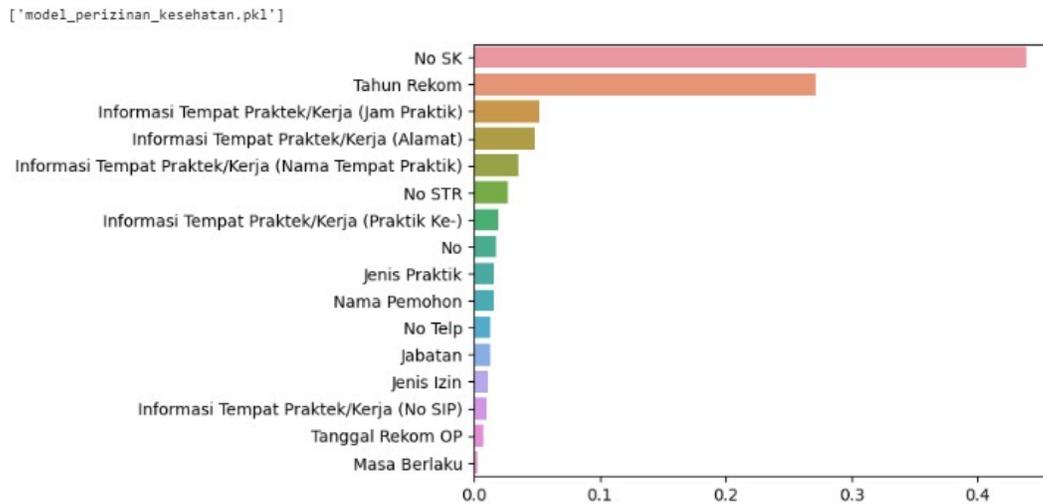
Dalam pembentukan Random Forest, setiap pohon keputusan dibangun menggunakan subset acak dari data latih dan fitur. Hasil prediksi kemudian dievaluasi menggunakan metrik kinerja seperti akurasi, presisi, recall, dan F1-score. Nilai accuracy sebesar 0.97 % dengan precision 0.95, recal 0,97, f1-score 0.96.

| Akurasi: 0.97 | | | | |
|---------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 0.95 | 0.97 | 0.96 | 37 |
| 1 | 0.98 | 0.96 | 0.97 | 53 |
| accuracy | | | 0.97 | 90 |
| macro avg | 0.96 | 0.97 | 0.97 | 90 |
| weighted avg | 0.97 | 0.97 | 0.97 | 90 |

Gambar 6. Clasification Report

Penentuan fitur yang berpengaruh

Dari Gambar 7 didapat informasi mengenai fitur yang factor yang paling penting dalam penerimaan izin tenaga kesehatan.



Gambar 7. Fitur Penting

KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa penerapan Data Mining, khususnya algoritma Random Forest, memiliki manfaat signifikan dalam proses analisis keputusan perizinan tenaga kesehatan. Model yang dibangun mampu mencapai akurasi yang sangat tinggi, serta presisi, recall, dan F1-score yang memuaskan. Hal ini menunjukkan bahwa teknik Data Mining dapat mengidentifikasi pola-pola penting dalam data perizinan dan memberikan prediksi yang akurat mengenai hasil perizinan. Selain itu, analisis pentingnya fitur memberikan wawasan tambahan mengenai faktor-faktor yang paling berpengaruh dalam keputusan perizinan, memungkinkan pihak berwenang untuk mengambil keputusan yang lebih baik dan lebih informatif. Secara keseluruhan, penelitian ini memberikan kontribusi berharga dalam meningkatkan proses perizinan tenaga kesehatan, memastikan bahwa standar kualitas dan keamanan pelayanan kesehatan tetap terjaga. Dengan teknologi Data Mining, pihak berwenang dapat mengoptimalkan proses evaluasi aplikasi perizinan, mengurangi waktu dan usaha yang dibutuhkan, serta meningkatkan akurasi keputusan yang diambil.

UCAPAN TERIMAKASIH

Kami mengucapkan terimakasih yang dan kami sampaikan penghargaan yang begitu tulus kepada Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri atas semua yang telah tcurahkan kepada kami, semoga apa yang kami dapatkan dapat membuat kami menjadi seseorang yang profesional dalam bidang apapun di kemudian hari.

DAFTAR PUSTAKA

- [1] E. Daniati, R. Firliana, A. S. Wardani and A. C. Zarkasi. (2021). Evaluation Framework for Decision Making Based On Sentiment Analysis in Social Media. International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA), Surabaya, Indonesia, pp. 47-51, doi: 10.1109/ICAMIMIA54022.2021.9807790.
- [2] E. Daniati, et al. (2019). Perancangan Jaringan Hotspot Untuk Peningkatan Layanan Teknologi Informasi. ANTIVIRUS: Jurnal Ilmiah Teknik Informatika, 13 (2), pp. 72-79, DOI: <https://doi.org/10.35457/antivirus.v13i2.857>

- [3] Sawabudin, B. ., Andriyanto, T. ., & Ristyawan, A. . (2021). Monitoring Of Scheduled Koi Feeding Through MCU Node And Blynk Application Based Smart Phone. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 5(2), 135–140. <https://doi.org/10.29407/inotek.v5i2.1027>
- [4] Krisnaryoko, E., Andriyanto, T., & Ristyawan, A. (2020). Decision Support System Pemilihan Agen Alat Tulis Kantor. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 4(3), 009–016. <https://doi.org/10.29407/inotek.v4i3.25>
- [5] A. Mawaddah, et al. (2019). Pemanfaatan Data Mining Pada Pengelompokan Provinsi Terhadap Pencemaran Lingkungan Hidup. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 3 (1), pp. 659-666, DOI: 10.30865/komik.v3i1.1675
- [6] S. Nita. (2020). Penerapan Data Mining Asosiasi pada Pola Transaksi dengan Metode Apriori. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 4(2), pp. 728-736, DOI: <http://dx.doi.org/10.30645/j-sakti.v4i2.345>
- [7] P. Astrid Novita, et al. (2022). Pemanfaatan Data Mining untuk Media Pembelajaran di SMK Hidayah Semarang. *E-DIMAS: Jurnal Pengabdian kepada Masyarakat*, 13(3), pp. 487-491, DOI: <https://doi.org/10.26877/e-dimas.v13i3.5572>
- [8] G. Urva, D. Desyanti, I. Albanna, muchamad S. Sungkar, and I. made agus oka Gunawan, *PENERAPAN DATA MINING DI BERBAGAI BIDANG*. 2023.
- [9] D. Cahya Prana Ginting, J. Samuel Parluhutan Sihombing, N. Natalia Aritonang, R. Patricia Sinaga, W. Purba, and F. Sains Dan Teknologi, ‘ANALISIS PEMBERIAN INSENTIF TENAGA MEDIS MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING’, *Jurnal TEKINKOM*, vol. 6, no. 1, 2023, doi: 10.37600/tekinkom.v6i1.858.
- [10] M. Riko Anshori Prasetya and A. Mudi Priyatno, ‘Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining’, doi: 10.37034/jidt.v5i1.324.
- [11] M. A. Latief, Y. Karyanti, M. Manajemen, and S. Informasi, ‘DATA MINING & ANALYTIC FORECASTING INDEKS STANDAR PENCEMAR UDARA JAKARTA MENGGUNAKAN METODE LINEAR REGRESSION (STUDI KASUS: DATASET INDEKS STANDAR PENCEMAR UDARA JAKARTA 2021)’, *JOSR: Journal of Social Research*, vol. 1, no. 10, pp. 1164–1176, 2022, [Online]. Available: <http://https://ijsr.internationaljournallabs.com/index.php/ijsrhttp://ijsr.internationaljournallabs.com/index.php/ijsr>
- [12] R. D. Marzuq, S. A. Wicaksono, and N. Y. Setiawan, ‘Prediksi Kanker Paru-Paru menggunakan Algoritme Random Forest Decision Tree’, 2023. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [13] B. Berdasarkan Tahun Perakitan, A. Suwandhi, J. Putra, and U. Ibbi, ‘Penerapan AI dalam Menentukan Harga Mobil’, *Jurnal Minfo Polgan*, vol. 13, no. 1, 2024, doi: 10.33395/jmp.v13i1.13728.
- [14] M. Ramanda Hasibuan, ‘Pemilihan Fitur dengan Information Gain untuk Klasifikasi Penyakit Gagal Ginjal menggunakan Metode Modified K-Nearest Neighbor (MKNN)’, 2019. [Online]. Available: <http://j-ptiik.ub.ac.id>

- [15] H. Airlangga Maulana, et al. (2021). Pemanfaatan Big Data dalam Bisnis E-commerce OLX. *Informatika dan Teknologi (INTECH)*, 2(2), pp. 13-18, DOI: <https://doi.org/10.54895/intech.v2i2.880>