

Prediksi Penyakit Paru Menggunakan Algoritma Deep Neural Network

Venia Restreva Danestiara, Agung Nur Fadhilah

Informatika, Teknologi dan Informatika, Universitas Informatika dan Bisnis Indonesia

Email: veniarestreva@unibi.ac.id; syaddad.master@gmail.com

Abstrak

Penyakit paru-paru, termasuk tuberculosis (TB) dan kanker paru, merupakan masalah kesehatan utama di Indonesia dengan angka kejadian dan kematian yang tinggi. Diagnosa dini dan akurat sangat penting untuk meningkatkan peluang kesembuhan pasien. Penelitian ini bertujuan untuk mengembangkan model prediksi penyakit paru-paru menggunakan algoritma Deep Neural Network (DNN). Dataset yang digunakan dalam penelitian ini terdiri dari 30.000 data riwayat kesehatan yang diambil dari situs Kaggle, dengan 52% data menunjukkan adanya penyakit paru-paru. Proses penelitian meliputi pengumpulan data, pre-processing data, pembuatan model DNN, dan evaluasi model menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa model prediksi yang dikembangkan memiliki akurasi sebesar 94%, dengan nilai precision dan recall yang tinggi untuk kelas positif dan negatif. Evaluasi model menunjukkan bahwa model mampu mengidentifikasi kasus penyakit paru-paru dengan tingkat kesalahan yang rendah.

Kata Kunci: Prediksi Penyakit Paru-paru, Deep Neural Network, Deep Learning

Abstract

Lung diseases, including tuberculosis (TB) and lung cancer, are major health issues in Indonesia with high incidence and mortality rates. Early and accurate diagnosis is crucial to increase the chances of patient recovery. This research aims to develop a predictive model for lung diseases using the Deep Neural Network (DNN) algorithm. The dataset used in this study consists of 30,000 health records obtained from the Kaggle website, with 52% of the data indicating the presence of lung diseases. The research process includes data collection, data pre-processing, DNN model development, and model evaluation using a confusion matrix. The results show that the developed predictive model achieved an accuracy of 94%, with high precision and recall values for both positive and negative classes. The model evaluation indicates that it is capable of identifying lung disease cases with a low error rate.

Keywords: Lung Disease Prediction, Deep Neural Network, Machine Learning

1 PENDAHULUAN

Menurut Kementerian Kesehatan (Kemenkes) Indonesia, negara ini berada di peringkat ketiga setelah India dan China dalam hal jumlah penderita tuberculosis (TB), dengan 824 ribu kasus dan 93 ribu kematian per tahun, atau setara dengan 11 kematian per jam. Penyakit paru-paru dapat disebabkan

oleh berbagai faktor seperti kebiasaan merokok, paparan polusi udara, atau bahan kimia, yang menyebabkan organ paru-paru tidak dapat berfungsi dengan normal (Gumelar, 2022).

Faktor utama penyebab penyakit paru di Indonesia adalah merokok, termasuk paparan asap rokok yang dapat meningkatkan

risiko kanker paru. Faktor lain yang berperan meliputi faktor genetik, riwayat kanker dalam keluarga, konsumsi kopi lebih dari 6 gelas per hari, penyakit paru kronis, konsumsi alkohol, konsumsi daging yang digoreng atau dipanggang, polusi udara, dan paparan bahan kimia (Sari, Romadloni, & Listyaningrum, 2022).

Penyakit paru, salah satunya kanker paru sering kali terdiagnosis pada tahap lanjut karena pasien sering menganggap batuk dan sesak napas sebagai hal biasa dan tidak segera memeriksakan diri ke fasilitas kesehatan. Selain itu, banyak petugas kesehatan hanya mengatasi gejala tanpa melakukan pemeriksaan lanjutan. Oleh karena itu, diperlukan metode yang dapat mendeteksi kanker paru dengan akurat melalui prediksi berdasarkan berbagai faktor dan riwayat kesehatan (Sari, Romadloni, & Listyaningrum, 2022). Oleh karena itu, metode prediksi seperti Deep Neural Network menjadi salah satu metode yang digunakan untuk memprediksi secara akurat.

Deep Neural Network (DNN) adalah jenis model machine learning yang terdiri dari banyak lapisan yang saling terhubung. Setiap lapisan memiliki banyak unit komputasi yang disebut neuron. Dalam DNN, data input diproses secara berurutan melalui setiap lapisan, dengan setiap lapisan melakukan transformasi pada data input dan menghasilkan representasi semantik yang semakin kompleks (Irpandudin, et al., 2022). Deep Neural Network (DNN) termasuk algoritma dalam sub-bidang machine learning yang dapat digunakan untuk prediksi, klasifikasi, dan regresi. DNN merupakan salah satu algoritma Deep Learning yang berbasis jaringan saraf dan digunakan untuk pengambilan keputusan. Algoritma ini memiliki struktur berlapis-lapis yang meniru model sirkuit saraf otak manusia dan hewan, dirancang untuk mengenali pola yang disebut Neural Network (Rizky, Pramuntadi, Prastowo, & Gutama, 2022).

Beberapa penelitian sebelumnya telah dilakukan menggunakan Deep Learning untuk memprediksi. Muhammad Rizky, dkk (2022) dengan implementasi Metode Deep Neural

Network pada Klasifikasi Penyakit Diabetes Melitus Tipe 2 dengan mendapat akurasi sebesar 95%. Kristina Ronalita Nainggolan, dkk (2022) dengan Prediksi Penyakit Parkinson Melalui Dataset Rekam Suara Dengan Menggunakan Algoritma Deep Neural Network dengan hasil akurasi sebesar 97%. Rudianto, dkk (2022) dengan prediksi tingkat kelulusan mahasiswa menggunakan algoritma neural network mendapat akurasi sebesar 98,27%. Sukri Syafudin, dkk (2021) dengan prediksi status pinjaman bank dengan deep learning neural network mendapat akurasi sebesar 82,27%. (Susanty & Wardana, 2021)

Berdasarkan pendahuluan dan penelitian sebelumnya, penelitian ini bertujuan untuk melakukan prediksi sakit paru berdasarkan berbagai faktor dan riwayat aktivitas menggunakan metode Deep Neural Network. (Fafaza, et al., 2022)

2 KAJIAN PUSTAKA

2.1 Machine Learning

Pembelajaran mesin (Machine Learning) adalah bidang studi yang terus berkembang dalam kecerdasan buatan, berfokus pada konsep pengenalan pola dan pembelajaran komputasi. Bidang ini menggunakan berbagai algoritma pembelajaran, baik yang diawasi maupun yang tidak diawasi, untuk memprediksi dan mendukung pengambilan keputusan otomatis berdasarkan sekumpulan data. Pembelajaran mesin ini dibangun dengan kerangka kerja persiapan data (Data Preparation) dan pemodelan data (Data Modelling) yang dapat dijalankan secara otomatis (Wardhana, Wang, & Sibuea, 2023).

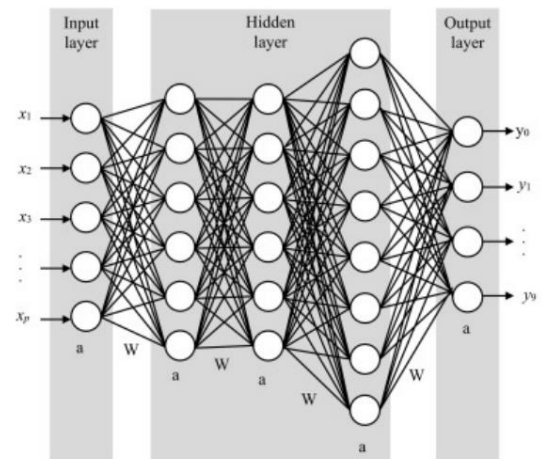
Machine learning adalah cabang AI yang paling banyak digunakan yang memungkinkan mesin untuk belajar menggunakan algoritma dari data dan membuat keputusan berdasarkan pola-pola yang ditemukan dalam data tersebut. Machine learning menggunakan model algoritma untuk bekerja dengan baik. Algoritma dalam machine learning dibagi menjadi tiga tipe yaitu supervised learning, unsupervised learning, dan reinforcement learning, yang

masing-masing memiliki proses berbeda. Salah satu terobosan terbesar dalam machine learning adalah deep learning. Deep learning adalah metode dalam kecerdasan buatan (AI) yang mengajarkan komputer untuk memproses data dengan cara yang terinspirasi oleh otak manusia. Model deep learning dapat mendeteksi pola kompleks dalam gambar, teks, audio, dan data lainnya untuk menghasilkan wawasan dan prediksi yang akurat (Ananto, et al., 2023).

2.2 Deep Neural Network (DNN)

Metode DNN (Deep Neural Network) adalah salah satu algoritma Deep Learning berbasis jaringan saraf yang digunakan untuk pengambilan keputusan. Algoritma ini menggunakan struktur berlapis-lapis yang meniru model sirkuit saraf kranial manusia dan hewan untuk mengenali pola yang dikenal sebagai neural network (NN) (Fafaza, et al., 2024). DNN menggunakan teknik Supervised Learning yang juga dikenal sebagai "programming by example". Dalam teknik ini, algoritma machine learning dilatih menggunakan dataset yang sudah diberi label, memungkinkan pengenalan pola yang ada di dalamnya. Supervised Learning melibatkan dataset dengan variabel target yang sudah diketahui, yang digunakan untuk mengatur dan mengkonfigurasi data. Tujuan utama dari supervised learning adalah mengeksplorasi hubungan antara variabel independen dan variabel dependen dalam konteks seperti klasifikasi dan regresi (Rizky, Pramuntadi, Prastowo, & Gutama, 2024). DNN (Deep Neural Network) dibangun dengan banyak lapisan yang setiap lapisannya saling terhubung ke lapisan berikutnya terkecuali lapisan input dan output. Gambar dibawah

merupakan struktur DNN (Deep Neural Network).

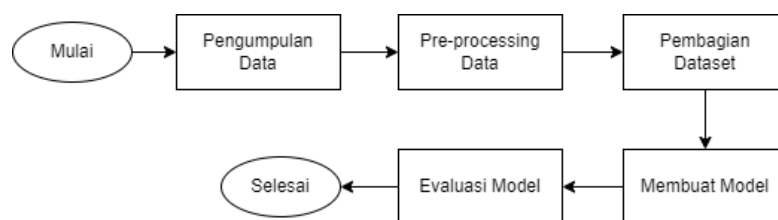


Gambar 1. Struktur DNN (Lestari & Halim, 2022)

Dalam Deep Neural Network (DNN), setiap neuron di setiap lapisan terhubung ke masing-masing neuron di lapisan berikutnya, seperti yang terlihat pada Gambar 1. Jumlah neuron di lapisan input harus sesuai dengan dimensi fitur. Sementara itu, jumlah neuron di lapisan output bergantung pada jumlah label berbeda yang dimiliki dataset. Misalnya, jika hanya ada satu neuron di lapisan output, DNN dapat digunakan untuk klasifikasi biner, karena 0 mewakili false dan 1 mewakili true (Lestari & Halim, 2022).

3 METODE PENELITIAN

Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 2.



Gambar 2. Tahapan Penelitian

3.1 Pengumpulan Data

Metode pengumpulan data riwayat kesehatan yang dilakukan yaitu dengan cara melakukan studi literatur. Mempelajari data-data yang sudah di publish oleh website Kaggle.com (Bsrc, 2024). Dataset Predic

Terkena Penyakit Paru-Paru ini memiliki 30.000 record data dan dari jumlah record 52% diantaranya menunjukkan hasil memiliki penyakit paru dan 48% diantaranya tidak memiliki penyakit paru. Tabel 1 merupakan dataset yang belum dikelola.

Tabel 1. Dataset yang belum dikelola

No	Usia	Jenis_Kelamin	Merokok	Bekerja	Rumah_Tangga	Aktivitas_Begadang	Aktivitas_Olahraga	Asuransi	Penyakit_Bawaan	Hasil
1	Tua	Pria	Pasif	Tidak	Ya	Ya	Sering	Ada	Tidak	Ya
2	Tua	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Ada	Tidak
3	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
4	Tua	Pria	Aktif	Ya	Tidak	Tidak	Jarang	Ada	Ada	Tidak
5	Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada	Ya
6	Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada	Tidak
7	Tua	Wanita	Pasif	Tidak	Ya	Tidak	Sering	Tidak	Tidak	Ya
8	Muda	Pria	Aktif	Tidak	Ya	Ya	Sering	Tidak	Tidak	Tidak
9	Tua	Wanita	Aktif	Ya	Ya	Ya	Jarang	Ada	Ada	Ya
10	Muda	Wanita	Pasif	Ya	Tidak	Ya	Jarang	Ada	Ada	Ya

3.2 Pre-Processing Data

Tahap pre-processing data merupakan salah satu tahapan penting sebelum data digunakan untuk membangun model prediksi. Pada tahap pre-processing dilakukan proses encoder untuk mengubah data mentah menjadi format yang dapat digunakan oleh algoritma deep learning (Arfan & Lussiana,

2019). Berikut adalah proses yang dilakukan pada tahap pre-processing.

a. Menghapus Kolom yang Tidak Relevan

Pada tahap ini kolom yang tidak diperlukan untuk analisis akan dihapus seperti pada Tabel 2.

Tabel 2. Dataset setelah kolom yang tidak relevan dihapus

Usia	Jenis_Kelamin	Merokok	Bekerja	Rumah_Tangga	Aktivitas_Begadang	Aktivitas_Olahraga	Asuransi	Penyakit_Bawaan	Hasil
Tua	Pria	Pasif	Tidak	Ya	Ya	Sering	Ada	Tidak	Ya
Tua	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Ada	Tidak
Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
Tua	Pria	Aktif	Ya	Tidak	Tidak	Jarang	Ada	Ada	Tidak
Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada	Ya
Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada	Tidak
Tua	Wanita	Pasif	Tidak	Ya	Tidak	Sering	Tidak	Tidak	Ya
Muda	Pria	Aktif	Tidak	Ya	Ya	Sering	Tidak	Tidak	Tidak
Tua	Wanita	Aktif	Ya	Ya	Ya	Jarang	Ada	Ada	Ya
Muda	Wanita	Pasif	Ya	Tidak	Ya	Jarang	Ada	Ada	Ya

Berdasarkan Tabel 1. ditemukan bahwa kolom “No” tidak mendukung dalam pembuatan model prediksi sehingga dilakukan penghapusan kolom dengan hasil yang didapat seperti pada Tabel 2.

b. Label Encoding

Label Encoding digunakan untuk mengubah nilai kategori ke dalam bentuk angka. Setiap kategori akan diubah menjadi angka unik (Herdian, Kamila, & Budidarma, 2024).

Tabel 3. Mengubah nilai kategori kedalam bentuk angka

Usia	Jenis_Kelamin	Merokok	Bekerja	Rumah_Tangga	Aktivitas_Begadang	Aktivitas_Olahraga	Asuransi	Penyakit_Bawaan	Hasil
1	0	1	0	1	1	1	0	1	1
1	0	0	0	1	1	0	0	0	0
0	0	0	0	1	1	0	0	1	0
1	0	0	1	0	0	0	0	0	0
0	1	1	1	0	0	1	1	0	1
0	1	1	1	0	0	1	1	0	0
1	1	1	0	1	0	1	1	1	1
0	0	0	0	1	1	1	1	1	0
1	1	0	1	1	1	0	0	0	1
0	1	1	1	0	1	0	0	0	1
1	1	1	1	1	0	1	0	0	1

3.3 Pembagian Dataset

Pada tahap ini dilakukan proses pembagian dataset menjadi 80% data training dan 20% data testing.

3.4 Membuat Model

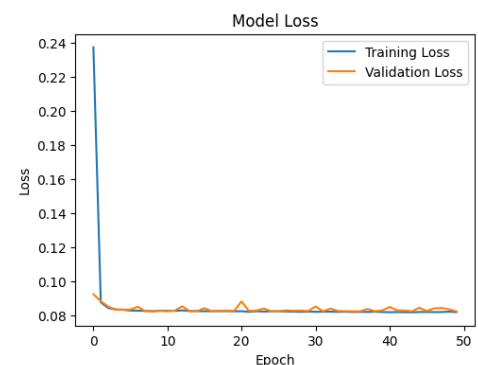
Pada tahap ini, model prediksi menggunakan DNN dikembangkan. Tujuan dari pembuatan model prediksi ini adalah untuk mencapai akurasi yang lebih tinggi serta menghindari masalah underfitting dan overfitting (Lestari & Halim, 2022). Data yang digunakan untuk membangun model adalah data pelatihan yang terdiri dari 80% data.

3.5 Evaluasi Model

Model prediksi yang telah dibangun menggunakan data pelatihan kemudian dievaluasi dengan menggunakan confusion matrix. Evaluasi ini bertujuan untuk menghasilkan nilai akurasi dan presisi, menggunakan 20% data uji yang dipakai dalam proses evaluasi model.

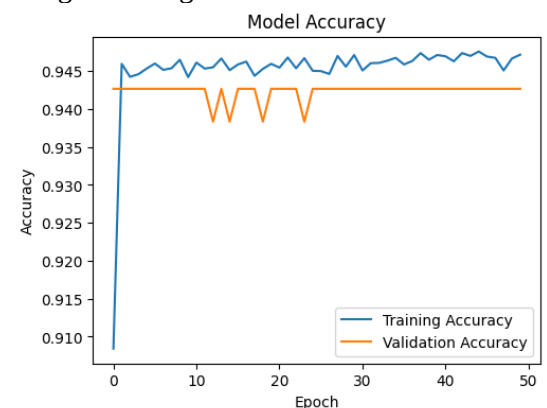
4 HASIL DAN PEMBAHASAN

4.1 Hasil Pembuatan Model



Gambar 3. Model Loss

Pada epoch pertama, training loss cukup tinggi, kemudian menurun drastis dan stabil di nilai yang rendah. Validation loss juga cukup rendah dan stabil, menandakan model tidak mengalami overfitting atau underfitting secara signifikan.

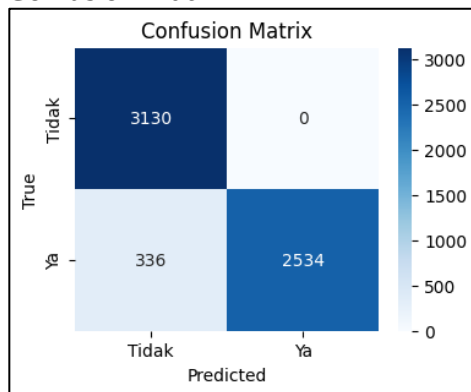


Gambar 4. Model Accuracy

Training accuracy meningkat tajam di epoch pertama dan kemudian stabil di sekitar 94.5%. Validation accuracy cukup stabil di sekitar 94.0%, yang menunjukkan bahwa model memiliki performa yang baik pada data yang tidak terlihat dan tidak overfitting.

4.2 Hasil Evaluasi Model

a. Confusion Matrix



Gambar 5. Confusion Matrix

Model berhasil mengidentifikasi dengan benar 2534 instance yang sebenarnya adalah "Ya," yang dikenal sebagai True Positif. Di sisi lain, model juga berhasil mengidentifikasi dengan benar 3075 instance yang sebenarnya adalah "Tidak," yang disebut sebagai True Negatif. Tidak ada instance yang salah diprediksi sebagai "Ya" ketika sebenarnya "Tidak" (False Positif), yang menunjukkan bahwa model sangat baik dalam menghindari kesalahan tipe ini. Namun, terdapat 344 instance yang salah diprediksi sebagai "Tidak" ketika sebenarnya "Ya," yang dikenal sebagai False Negatif, menunjukkan bahwa model masih melewatkan beberapa instance yang seharusnya diklasifikasikan sebagai "Ya."

b. Classification Report

Classification Report:				
	precision	recall	f1-score	support
0	0.90	1.00	0.95	3130
1	1.00	0.88	0.94	2870
accuracy			0.94	6000
macro avg	0.95	0.94	0.94	6000
weighted avg	0.95	0.94	0.94	6000

Gambar 6. Classification Report

Model memiliki akurasi keseluruhan sebesar 94%, yang menunjukkan bahwa model mampu memprediksi dengan benar sebagian besar waktu.

a. Kinerja untuk kelas 0 :

Precision sebesar 90%, ketika model memprediksi kelas 0, 90% dari prediksi tersebut benar. Recall sebesar 100%, model berhasil menemukan semua contoh yang benar-benar merupakan kelas 0. F1-score sebesar 95% menunjukkan keseimbangan yang baik antara ketepatan dan kemampuan model untuk menemukan semua contoh kelas 0.

b. Kinerja untuk Kelas 1:

Precision sebesar 100%, ketika model memprediksi kelas 1, semua prediksi tersebut benar. Recall sebesar 88%, model berhasil menemukan 88% dari semua contoh yang benar-benar merupakan kelas 1. F1-score sebesar 94% menunjukkan keseimbangan yang sangat baik antara ketepatan dan kemampuan model untuk menemukan sebagian besar contoh kelas 1.

c. Cross Validation

```

188/188 [=====] - 0s 1ms/step
188/188 [=====] - 0s 1ms/step
188/188 [=====] - 1s 2ms/step
188/188 [=====] - 0s 2ms/step
188/188 [=====] - 0s 1ms/step
Cross-Validation Accuracies: [0.9458333333333333, 0.9466666666666667, 0.9538333333333333, 0.944]
Mean Accuracy: 0.9465999999999999
Standard Deviation of Accuracies: 0.0036645448486164858
Overall Confusion Matrix:
[[15648  0]
 [1682 12750]]

```

Gambar 7. Cross Validation

Dalam pengujian ini, terdapat 5 fold dengan akurasi sebagai berikut: Fold 1 mencapai akurasi sebesar 94.58%, Fold 2 dan Fold 3 masing-masing memiliki akurasi sebesar 94.47%, Fold 4 memperoleh akurasi tertinggi sebesar 95.38%, dan Fold 5 memiliki akurasi sebesar 94.40%. Rata-rata akurasi dari kelima fold ini adalah sekitar 94.66%. Standar deviasi dari akurasi adalah sekitar 0.0037, yang menunjukkan bahwa variasi akurasi antar fold sangat kecil, sehingga performa model stabil di berbagai fold yang diuji.

5 SIMPULAN

Deep Neural Network (DNN) berhasil digunakan untuk memprediksi penyakit paru-

paru dengan tingkat akurasi yang tinggi, mencapai 94%. Model ini menunjukkan performa yang baik dalam mengidentifikasi pasien dengan dan tanpa penyakit paru, dengan nilai precision dan recall yang tinggi pada kedua kelas. Selain itu, evaluasi melalui matriks kebingungan memperlihatkan bahwa model memiliki tingkat kesalahan yang rendah, terutama dalam meminimalkan false positives, yang menunjukkan kemampuannya dalam menghindari prediksi yang salah. Dengan demikian, model DNN yang dikembangkan dapat diandalkan sebagai alat bantu dalam diagnosis dini penyakit paru-paru, membantu meningkatkan peluang kesembuhan pasien dengan deteksi yang lebih akurat.

DAFTAR PUSTAKA

- Gumelar, M. A. (2022). Sistem Deteksi Penyakit Paru-paru Melalui Citra X-Ray Thorax Dengan Algoritma Convolutional Neural Network (CNN) dan Deep Hybrid Learning (DHL) Berbasis Aplikasi Website.
- Ananto, D. T., Mahardewantoro, D. D., Mustafa, F., Ardianto, M. G., Rafi, M. M., Zein, R. A., . . . Adharani, Y. (2022). Edukasi dan Pelatihan Pengenalan Machine Learning dan Computer Vision Untuk Mengeksplorasi Potensi Visual. *Prosiding Seminar Nasional LPPM UMJ*.
- Arfan, A., & Lussiana. (2019). Prediksi Harga Saham Di Indonesia Menggunakan Algoritma Long Short-Term Memory. *Seminar Nasional Teknologi Informasi dan Komunikasi STI&K*.
- Bsrc, A. (2022, Mei 7). *DATASET PREDIC TERKENA PENYAKIT PARU-PARU*. Diambil kembali dari Kaggle: <https://www.kaggle.com/datasets/andot03bsrc/dataset-predic-terkena-penyakit-paruparu>
- Fafaza, S. A., Rohman, M. S., Pramunendar, R. A., Winarsih, N. A., Saraswati, G. W., Saputra, F. O., . . . Shidik, G. F. (2022). Prediksi Banjir Berdasarkan Indeks Curah Hujan Menggunakan Deep Neural Network (Dnn). *Jurnal Media Informatika Budidarma*, 173-184.
- Herdian, C., Kamila, A., & Budidarma, I. G. (2022). Perbandingan Label Encoding Dan One-Hot Encoding Pada Metode Linear Regresi. *Technologia : Jurnal Ilmiah*.
- Irpanudin, Reka, Anggraeni, R. N., Pratama, P., Sujjada, A., & Fergina, A. (2022). Prediksi Penyakit Jantung Menggunakan Metode Deep Neural Network Dengan Memanfaatkan Internet Of Things. *Jurnal Informasi Dan Teknologi*, 45-55.
- Lestari, W. S., & Halim, A. (2022). Prediksi Kesuksesan Startup Menggunakan Deep Neural Network. *Jurnal Sifo Mikroskil*.
- Nainggolan, K. R., Purnamasari, F., & Pulungan, A. F. (2022). Prediksi Penyakit Parkinson melalui Dataset Rekam Suaradengan Menggunakan Algoritma Deep Neural Network. *Jurnal Minfo Polgan*.
- Rizky, M., Pramuntadi, A., Prastowo, W. D., & Gutama, D. H. (2022). Implementasi Metode Deep Neural Network Pada Klasifikasi Penyakit Diabetes Melitus Tipe 2. *Institut Riset Dan Publikasi Indonesia(Irpi)*.
- Rudianto, Kania, R., & Solihati, T. I. (2022). Prediksi Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika Universitas Banten Jaya Menggunakan algoritma Neural Network. *Jurnal Sistem Informasi Dan Informatika (Simika)*.

- Sari, L., Romadloni, A., & Listyaningrum, R. (2022). Penerapan Data Mining Dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest. *Infotekmesin*.
- Susanty, M., & Wardana, R. S. (2021). Prediksi Tekanan Pori Berdasarkan Data Logging Sumur Menggunakan Deep Neural Network. *Jurnal Pengkajian Dan Penerapan Teknik Informatika*.
- Syafudin, S., Nugraha, R. A., Handayani, K., Gata, W., & Linawati, S. (2021). Prediksi Status Pinjaman Bank Dengan Deep Learning Neural Network (Dnn). *Jurnal Teknik Komputer Amik Bsi*.
- Wardhana, R. G., Wang, G., & Sibuea, F. (2022). Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia. *Journal Of Information System Management (Joism)*.