

PREDIKSI PASIEN KANKER PARU-PARU MENGGUNAKAN METODE K-NEAREST NEIGHBORS, DECISION TREE, DAN NAIVE BAYES

Yosia Amadeus Ishak¹⁾ Manatap Dolok Lauro²⁾

¹⁾ Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

email : yosia.535210037@stu.untar.ac.id

²⁾ Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara

Jl. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

email : manataps@fti.untar.ac.id

ABSTRACT

This research aims to determine the application performance of the K-Nearest Neighbors (KNN), Decision Tree, and Naive Bayes methods in predicting lung cancer patients. Lung cancer is a deadly disease that is often difficult to detect in its early stages. Therefore, the development of accurate and efficient prediction models has a significant impact in early diagnosis and improving patient survival rates. This research yields a deeper understanding of the performance of these methods in the context of lung cancer prediction. The dataset used includes information such as age, gender, and other medical history of the lung cancer patients observed. Experimental results show that Decision Tree has the highest accuracy, followed by KNN and Naive Bayes. However, these three methods provide valuable contributions in the context of lung cancer prediction. These findings can be a basis for further development in the field of cancer diagnosis and provide valuable insights for medical practitioners and researchers in efforts to improve the effectiveness of early lung cancer detection.

Key words

Decision Tree, K-Nearest Neighbors, Lung Cancer Patients, Naive Bayes, Prediction

1. Pendahuluan

Dewasa ini, teknologi informasi telah berkembang secara pesat dan mengubah cara kita memandang dan mendekati berbagai fenomena, termasuk fenomena seputar kesehatan contohnya seperti melakukan deteksi dan pengobatan penyakit kronis seperti kanker paru-paru. Kanker paru-paru merupakan suatu keganasan yang terjadi pada organ paru-paru yang disebabkan oleh perubahan genetika dalam sel epitel saluran nafas, sehingga terjadi proliferasi sel yang tidak terkendali [1]. Kanker paru-paru adalah kanker paling umum nomor 2 di dunia [2]. Menurut data Global Burden of Cancer Study dari World Health Organization (WHO), kasus kanker di

Indonesia berada pada angka 396.914 kasus dengan angka kematian sebesar 234.511 kasus [3]. Paparan atau inhalasi berkepanjangan terhadap suatu zat yang bersifat karsinogenik merupakan faktor penyebab utama disamping adanya faktor lain seperti kekebalan tubuh, genetik dan lain-lain [4]. Oleh karena itu, pengembangan metode deteksi dini untuk penyakit kanker terutama kanker paru-paru menjadi sangat penting dengan tujuan untuk meningkatkan angka kelangsungan hidup setiap pasien dan merupakan pertahanan terbaik melawan kanker [5].

Dengan adanya tantangan tersebut, kecerdasan buatan khususnya dalam bentuk pembelajaran mesin dapat menjadi solusi atas permasalahan yang dihadapi. Penggunaan algoritma pembelajaran yang diawasi seperti algoritma K-Nearest Neighbors (KNN), algoritma Decision Tree, dan algoritma Naive Bayes dapat digunakan dalam konteks klasifikasi dan prediksi penyakit sebagai pendekatan baru terhadap upaya deteksi dini pasien kanker paru-paru. K-Nearest Neighbors adalah sebuah algoritma sederhana dimana titik data akan dikelompokkan kedalam kelompok tertentu bersama dengan titik lain yang berdekatan berdasarkan kemiripan karakteristik [6]. Algoritma Decision Tree adalah metode klasifikasi berupa struktur pohon dimana setiap atribut diwakili dengan node, nilai suatu atribut diwakili dengan cabang, dan setiap kelas diwakili dengan daun [7]. Kemudian, algoritma Naive Bayes adalah sebuah metode klasifikasi menggunakan sekumpulan probabilitas sederhana dan menjumlahkan frekuensi dan kombinasi nilai dari dataset [8].

Tujuan penelitian ini adalah untuk melakukan perbandingan akurasi dan mengetahui tingkat efektifitas penggunaan algoritma K-Nearest Neighbors, algoritma Decision Tree, dan algoritma Naive Bayes dalam memprediksi pasien kanker paru-paru berdasarkan data klinis dan histopatologi. Pada penelitian ini, masing algoritma akan dibahas secara lebih mendalam, termasuk konsep dasar algoritma terkait, dan penerapan mereka dalam konteks triase pasien kanker paru-paru. Selain itu, kinerja ketiga algoritma tersebut akan dievaluasi dan

dibandingkan untuk mengetahui algoritma yang terbaik dalam memprediksi pasien kanker paru-paru. Melalui penelitian ini, diharapkan dapat memberikan wawasan berharga mengenai penggunaan pembelajaran mesin untuk membantu memudahkan layanan kesehatan dan mempercepat pengambilan keputusan klinis serta meningkatkan pemahaman mengenai potensi penerapan algoritma KNN, algoritma Decision Tree, dan algoritma Naive Bayes dalam meningkatkan akurasi diagnostik dan meningkatkan angka kesembuhan dan kualitas hidup pasien kanker paru-paru.

2. Dasar Teori

2.1 Kerangka Penelitian

Kerangka dari penelitian ini dimulai dari permasalahan analisa penyakit pasien kanker paru-paru, kemudian dibuat sebuah model dalam bentuk algoritma K-Nearest Neighbors, algoritma Decision Tree, dan algoritma Naive Bayes. Tools yang digunakan untuk mendukung penelitian ini adalah Jupyter Notebook atau Visual Studio Code. Data penelitian diambil dari situs Kaggle dengan judul "Lung Cancer Predictions" yang menampung informasi para pasien kanker paru-paru seperti umur, jenis kelamin, dan 23 kelas kategori riwayat medis lainnya [9]. Tujuan dari dataset ini adalah untuk memprediksi kemungkinan pasien terkena kanker paru-paru, dan mengidentifikasi faktor risiko kanker paru-paru supaya dapat menentukan pengobatan yang paling efektif untuk mereka.

2.2 Data Preprocessing

Data preprocessing adalah proses transformasi raw data menjadi data dengan bentuk yang lebih mudah dipahami untuk memperbaiki kesalahan yang terdapat pada data raw yang biasanya tidak lengkap atau memiliki format yang tidak teratur [10]. Berikut adalah tahap-tahap pre-processing data yang digunakan pada penelitian ini:

1. Pembersihan data (Data cleansing)
Data cleaning atau pembersihan data adalah sebuah proses dalam pembelajaran mesin yang dilakukan dengan tujuan untuk mempersiapkan data analisis dengan mengubah format data yang salah, tidak lengkap, terduplikasi, dan tidak relevan [11]. Pada penelitian ini, tahap data cleaning tidak dilakukan karena data sudah terlebih dahulu diproses oleh pemilik dataset pada situs Kaggle.
2. Transformasi Data (Data Transformation)
Data transformation atau transformasi data adalah upaya yang dilakukan dengan tujuan utama untuk mengubah skala pengukuran pada

data asli menjadi bentuk lain sehingga data tersebut memenuhi asumsi-asumsi yang mendasari analisis ragam [12], seperti mengonversi tipe pada data, menormalisasi suatu nilai, atau melakukan ekstraksi pada fitur. Pada penelitian ini, data tidak melalui tahap transformasi karena data sudah terlebih dahulu diproses dan diransformasi oleh pemilik dataset pada situs Kaggle.

3. Pemilihan Fitur (Feature Selection)

Feature selection atau pemilihan fitur adalah teknik dalam Data Mining (atau dalam hal ini Machine Learning) untuk melakukan pemilihan terhadap kolom fitur yang relevan dan mengurangi atribut yang berlebihan dan tidak dibutuhkan dari kumpulan data [13]. Tujuan dari hal ini adalah untuk memilih fitur yang paling relevan dari banyaknya fitur dan mengurangi dimensi model. Pada penelitian ini, ditambahkan kolom baru yaitu kolom "Level" sebagai kolom target untuk 1000 data yang merupakan hasil pemeriksaan terhadap 1000 orang pasien kanker paru-paru. 30% dimasukkan sebagai kategori 'Low', 33% dimasukkan sebagai kategori 'Medium', 37% dimasukkan sebagai kategori 'High'.

2.3 K-Nearest Neighbors

Algoritma K-Nearest Neighbor atau disingkat KNN adalah sebuah algoritma pengklasifikasian yang didasarkan dengan jarak kedekatan antara suatu data dengan data baru lainnya [14]. Data yang memiliki ciri karakteristik yang sama akan dikelompokkan hingga terbentuk beberapa kelompok data. Terdapat beberapa metode penghitungan jarak pada algoritma KNN, tetapi yang paling umum digunakan adalah rumus Euclidean Distance seperti pada persamaan berikut [15]:

$$d(x, y) = \sum_{i=1}^n (x_i - y_i)^2$$

Keterangan:

x_i = sampel data

y_i = data uji atau data testing

i = variabel data

$d(x, y)$ = jarak Euclidean distance

n = dimensi data

2.4 Decision Tree

Decision tree merupakan algoritma pengklasifikasian yang menggunakan struktur pohon yang terdiri dari kumpulan node dimana masing-masing node

merepresentasikan keputusan dan cabang-cabang merepresentasikan konsekuensi dari keputusan tersebut [16]. Kelebihan dari algoritma Decision Tree adalah memudahkan menginterpretasikan solusi dari permasalahan dengan kemampuan Decision Tree dalam menjabarkan setiap proses pengambilan keputusan yang kompleks [17]. Untuk melakukan klasifikasi menggunakan algoritma Decision Tree, digunakan persamaan dalam melakukan perhitungan nilai entropy dan gain yang ditunjukkan pada persamaan sebagai berikut [18]:

$$Entropy(S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i$$

Keterangan:

S = himpunan kasus
 n = jumlah partisi S
 pi = proporsi Si terhadap S

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

S = himpunan kasus
 n = jumlah partisi S
 pi = proporsi Si terhadap S
 |Si| = Proporsi Si terhadap S
 |S| = jumlah kasus dalam S

2.5 Naive Bayes

Algoritma Naive Bayes adalah algoritma klasifikasi atau prediksi peluang di masa yang akan datang berdasarkan pengalaman yang sudah terjadi sebelumnya menggunakan probabilitas yang dicetuskan oleh ilmuwan Inggris yaitu Thomas Bayes [19]. Rumus umum dasar algoritma Naive Bayes ditunjukkan sebagai berikut [20]:

$$P(H|X) = \frac{P(X|H) * P(H)}{P(X)}$$

Keterangan:

X = data tanpa kelas
 H = Hipotesis data X merupakan suatu kelas spesifik
 P(X) = Probabilitas hipotesis X
 P(H) = Probabilitas hipotesis H
 P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H
 P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X

3. Hasil dan Pembahasan

Tujuan utama dari penelitian ini adalah untuk memprediksi dan mengklasifikasikan kategori pasien kanker paru-paru ke dalam tiga kelas, dimana setiap kelas

merepresentasikan tingkat keseriusan kondisi pasien masing-masing. Ketiga kelas tersebut merupakan “Low” yang direpresentasikan dengan angka 1, “Medium” yang direpresentasikan dengan angka 2, dan “High” yang direpresentasikan dengan angka 3. Penelitian dilakukan menggunakan algoritma K-Nearest Neighbors, algoritma Decision Tree, dan algoritma Naive Bayes. Hasil dari model ketiga algoritma akan dievaluasi dan kemudian akan dibandingkan menggunakan confusion matrix untuk menentukan algoritma terakurat dalam menangani data pasien kanker paru-paru. Dataset yang digunakan memiliki jumlah sampel sebanyak 1000 dan fitur sebanyak 24, **Tabel 1** berikut merupakan penjelasan mendalam mengenai masing-masing fitur:

Tabel 1. Penjelasan Fitur

| Fitur | Penjelasan |
|--------------------------|--|
| Age | Umur pasien |
| Gender | Jenis kelamin pasien |
| Air Pollution | Tingkat paparan polusi udara pada pasien |
| Alcohol use | Tingkat penggunaan alkohol pasien |
| Dust Allergy | Tingkat alergi debu pasien |
| Occupational Hazards | Tingkat bahaya kerja |
| Genetic Risk | Risiko Genetik |
| Chronic Lung Disease | Penyakit paru-paru kronis |
| Balanced Diet | Diet seimbang |
| Obesity | Tingkat obesitas pasien |
| Smoking | Tingkat merokok pasien. |
| Passive Smoker | Tingkat perokok pasif pasien |
| Chest Pain | Tingkat nyeri dada pasien |
| Coughing of Blood | Tingkat batuk darah pasien |
| Fatigue | Tingkat kelelahan pasien |
| Weight Loss | Tingkat penurunan berat badan pasien |
| Shortness of Breath | Tingkat sesak nafas pasien |
| Wheezing | Tingkat mengi pasien |
| Swallowing Difficulty | Tingkat kesulitan menelan pasien |
| Clubbing of Finger Nails | Tingkat pembengkakan kuku jari pasien. |
| Frequent Cold | Tingkat demam pasien |
| Dry Cough | Tingkat batuk kering pasien |
| Snoring | Tingkat dengkur pasien |
| Level | Kolom target kelas |

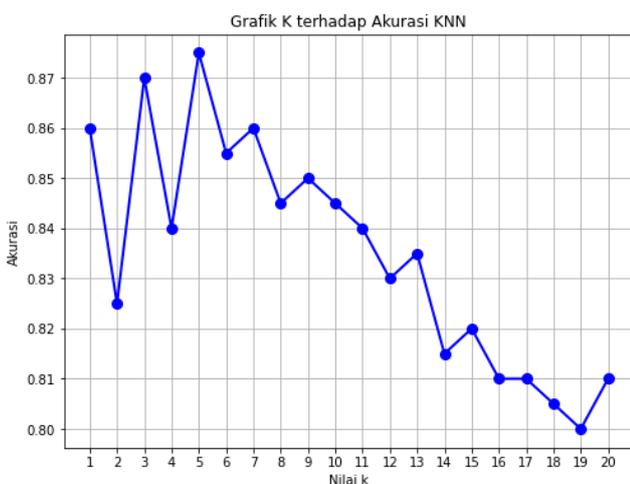
Terdapat 1000 sampel, 24 kolom fitur dan 1 kelas target yang akan digunakan dalam proses prediksi dan klasifikasi. Tidak terdapat nilai null pada data pasien kanker paru-paru tersebut, sehingga data siap dibagi

menjadi data latih dan data uji dengan test_sizes sebesar 0.20 atau rasio pembagian data 80:20 untuk masing-masing data testing dan data training. Kemudian akan dilakukan evaluasi model menggunakan metrik evaluasi yang terdiri dari precision, recall, f1-score, dan support. Hasil kinerja ketiga algoritma akan dibandingkan untuk mengetahui kinerja terbaik dari ketiga algoritma tersebut dalam melakukan prediksi terhadap pasien kanker paru.

```
df.drop(['index', 'Patient Id'], axis =1, inplace = True)
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 24 columns):
# Column Non-Null Count Dtype
---
0 Age 1000 non-null int64
1 Gender 1000 non-null int64
2 Air Pollution 1000 non-null int64
3 Alcohol use 1000 non-null int64
4 Dust Allergy 1000 non-null int64
5 OccuPational Hazards 1000 non-null int64
6 Genetic Risk 1000 non-null int64
7 chronic Lung Disease 1000 non-null int64
8 Balanced Diet 1000 non-null int64
9 Obesity 1000 non-null int64
10 Smoking 1000 non-null int64
11 Passive Smoker 1000 non-null int64
12 Chest Pain 1000 non-null int64
13 Coughing of Blood 1000 non-null int64
14 Fatigue 1000 non-null int64
15 Weight Loss 1000 non-null int64
16 Shortness of Breath 1000 non-null int64
17 Wheezing 1000 non-null int64
18 Swallowing Difficulty 1000 non-null int64
19 Clubbing of Finger Nails 1000 non-null int64
20 Frequent Cold 1000 non-null int64
21 Dry Cough 1000 non-null int64
22 Snoring 1000 non-null int64
23 Level 1000 non-null object
dtypes: int64(23), object(1)
memory usage: 187.6+ KB
```

Gambar 1. Data setelah pre-processing

3.1 Hasil Algoritma KNN



Gambar 2. Grafik K terhadap Akurasi KNN

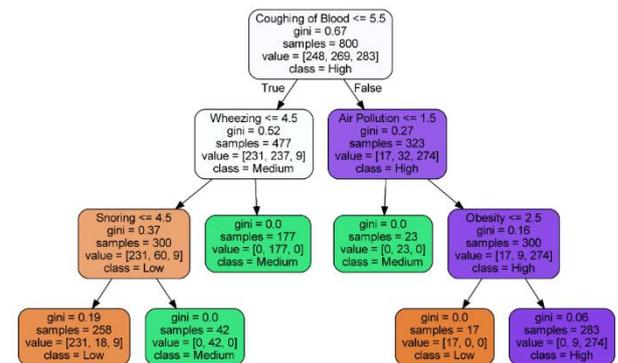
Pada algoritma pertama yaitu algoritma K-Nearest Neighbors, dihitung nilai k dengan akurasi terbaik dari nilai k = 1 hingga k = 20 menggunakan grafik seperti pada

Gambar 2. Didapatkan hasil penemuan dimana nilai k dengan akurasi tertinggi adalah nilai k = 5 dengan akurasi mencapai 88%. Kemudian dilakukan evaluasi model menggunakan metrik evaluasi precision, recall, f1-score, dan support yang dapat dilihat pada Tabel 2 berikut:

Tabel 2. Hasil evaluasi algoritma KNN

| Class | Precision | Recall | F1-score | Support |
|--------------|-----------|--------|----------|---------|
| 1 (Low) | 0.86 | 0.93 | 0.89 | 55 |
| 2 (Medium) | 0.86 | 0.76 | 0.81 | 63 |
| 3 (High) | 0.89 | 0.93 | 0.91 | 82 |
| Accuracy | | | 0.88 | 200 |
| Macro avg | 0.87 | 0.87 | 0.87 | 200 |
| Weighted avg | 0.87 | 0.88 | 0.87 | 200 |

3.2 Hasil Algoritma Decision Tree



Gambar 3. Hasil Decision Tree

Hasil akurasi Decision Tree berada pada angka 98%. Algoritma Decision Tree ini memiliki nilai tertinggi dan kinerja terbaik dalam melakukan prediksi dan klasifikasi pasien kanker paru-paru jika dibandingkan dengan algoritma KNN dan algoritma Decision Tree. Evaluasi algoritma Decision Tree dapat dilihat pada Tabel 3 berikut:

Tabel 3. Hasil evaluasi algoritma Decision Tree

| Class | Precision | Recall | F1-score | Recall |
|--------------|-----------|--------|----------|--------|
| 1 (Low) | 0.95 | 1.00 | 0.97 | 55 |
| 2 (Medium) | 1.00 | 0.95 | 0.98 | 63 |
| 3 (High) | 0.99 | 0.99 | 0.99 | 82 |
| Accuracy | | | 0.98 | 200 |
| Macro avg | 0.98 | 0.98 | 0.98 | 200 |
| Weighted avg | 0.98 | 0.98 | 0.98 | 200 |

3.3 Hasil Algoritma Naïve Bayes

Hasil akurasi dari algoritma Naïve Bayes memiliki nilai yang setara dengan akurasi yang dimiliki oleh algoritma KNN, yaitu sebesar 88%. Hal ini menunjukkan bahwa kedua algoritma memiliki kemampuan klasifikasi yang sama-sama baik dalam penelitian ini. Evaluasi algoritma Naïve Bayes dapat dilihat pada **Tabel 4** berikut:

Tabel 4. Hasil evaluasi algoritma Naive Bayes

| Class | Precision | Recall | F1-score | Support |
|--------------|-----------|--------|----------|---------|
| 1 (Low) | 1.00 | 0.87 | 0.93 | 60 |
| 2 (Medium) | 0.87 | 0.82 | 0.85 | 73 |
| 3 (High) | 0.80 | 0.94 | 0.86 | 67 |
| Accuracy | | | 0.88 | 200 |
| Macro avg | 0.89 | 0.88 | 0.88 | 200 |
| Weighted avg | 0.89 | 0.88 | 0.88 | 200 |

3.4 Perbandingan Kinerja Algoritma

Setelah dilakukan prediksi pasien kanker paru-paru menggunakan 3 algoritma diatas, dapat disimpulkan bahwa algoritma Decision Tree memiliki performa terbaik dalam memprediksi dan mengklasifikasikan pasien kanker paru-paru dengan nilai akurasi sebesar 98%, sementara algoritma KNN dan algoritma Naïve Bayes memiliki nilai akurasi yang sedikit lebih rendah yaitu pada nilai 88% yang menunjukkan bahwa kedua algoritma tersebut tidak seefektif Decision Tree dalam melakukan prediksi dan klasifikasi pada kasus penelitian ini. Perbandingan akurasi antara algoritma KNN, Decision Tree, dan Naïve Bayes dapat dilihat pada **Tabel 5** berikut:

Tabel 5. Perbandingan Akurasi Kinerja Algoritma

| Algoritma | Akurasi |
|---------------|---------|
| KNN | 88% |
| Decision Tree | 98% |
| Naïve Bayes | 88% |

4. Kesimpulan

Hasil penelitian prediksi pasien kanker paru-paru yang telah dilakukan menunjukkan bahwa algoritma Decision Tree menghasilkan kinerja terbaik dengan nilai akurasi sebesar 98%, sedangkan algoritma KNN dan algoritma Naïve Bayes memiliki nilai akurasi yang serupa sebesar 88%.

Algoritma Decision Tree terbukti lebih efektif dalam melakukan prediksi pasien kanker paru-paru dibandingkan dengan algoritma KNN dan algoritma

Naïve Bayes menggunakan dataset ini. Keunggulan algoritma Decision Tree pada penelitian ini disebabkan oleh kemampuannya dalam menangkap relasi non-linear pada fitur-fitur yang terdapat pada dataset pasien kanker paru-paru, dimana algoritma lain tidak dapat melakukan hal yang sama secara optimal. Terdapat faktor lainnya seperti kesensitifan algoritma KNN terhadap nilai k atau nilai kluster serta asumsi independensi fitur pada algoritma Naive Bayes yang memberikan batasan performa pada algoritma KNN dan algoritma Naïve Bayes.

Maka, algoritma Decision Tree dapat dijadikan pilihan terbaik untuk melakukan prediksi kanker paru-paru karena kemampuannya dalam memberikan hasil yang lebih akurat dibandingkan algoritma KNN dan algoritma Naïve Bayes. Namun, disarankan untuk melakukan pengujian lebih lanjut menggunakan dataset dengan jumlah sampel yang lebih besar dan fitur yang beragam untuk memastikan generalisasi hasil yang lebih baik.

Referensi

- [1] I. Buana and D. Agustian Harahap, "ASBESTOS, RADON DAN POLUSI UDARA SEBAGAI FAKTOR RESIKO KANKER PARU PADA PEREMPUAN BUKAN PEROKOK," Aceh, May 2022.
- [2] S. Sugiharto, R. A. P. S. Simanjuntak, and O. Larissa, "KANKER PARU, FAKTOR RISIKO DAN PENCEGAHANNYA," *Seminar Nasional Hasil Penelitian dan Pengabdian Kepada Masyarakat 2021 Pengembangan Ekonomi Bangsa Melalui Inovasi Digital Hasil Penelitian dan Pengabdian Kepada Masyarakat Jakarta*, Oct. 2021.
- [3] N. Handayani, "Kanker dan Serba-Serbinya (Hari Kanker Sedunia 2022)," <https://rsprospira.jogjaprovo.go.id/kanker-dan-serba-serbinya-hari-kanker-sedunia-2022/>.
- [4] J. Joseph and L. W. A. Rotty, "Kanker Paru: Laporan Kasus," *Medical Scope Journal*, vol. 2, no. 1, Jul. 2020, doi: 10.35790/msj.2.1.2020.31108.
- [5] S. S. Bhusare, V. Yatnalli, S. G. Patil, S. Kumar, N. R. Havinal, and C. S. Subhash, "Machine Learning and Digital Image Processing in Lung Cancer Detection," pp. 56–67, May 2023, doi: 10.1007/978-3-031-31164-2_6.
- [6] D. B. Laraswati, "Mengenal K-nearest Neighbor dan Pengaplikasiannya," <https://blog.algoritma.k-nearest-neighbor/>.
- [7] A. M. Alfatah, R. Arifudin, and A. Muslim, "Implementation of Decision Tree and Dempster Shafer on Expert System for Lung Disease Diagnosis," *Scientific Journal of Informatics*, vol. 5, no. 1, pp. 2407–7658, May 2018, [Online]. Available: <http://journal.unnes.ac.id/nju/index.php/sji>

- [8] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec Journal*, vol. 2, no. 3, May 2015.
- [9] T. Devastator, "Lung Cancer Prediction," <https://www.kaggle.com/datasets/thedevastator/cancer-patients-and-air-pollution-a-new-link>.
- [10] A. Adam, "Data Preprocessing: Pengertian, Manfaat, dan Tahapan Kerjanya," <https://accurate.id/teknologi/data-preprocessing/>.
- [11] A. N. Fariza, "Data Cleaning Adalah: Pengertian, Urgensi, Manfaat, dan 3 Contohnya," <https://www.sekawanmedia.co.id/blog/data-cleaning/>.
- [12] F. Effendy and P. Studi, "Klasifikasi Rumah Tangga Miskin Menggunakan Ordinal Class Classifier," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 4, no. 1, 2018, doi: 10.25077/TEKNOSI.v4i1.2018.009-020.
- [13] F. F. Firdaus, H. A. Nugroho, and I. Soesanti, "A Review of Feature Selection and Classification Approaches for Heart Disease Prediction," *IJITEE*, vol. 4, no. 3, Sep. 2020.
- [14] D. Fusvita and F. Hari Utami, "Penerapan Algoritma KNN (K-Nearest Neighbour) Dalam Klasifikasi Data Pinjaman Anggota Koperasi," *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya*, vol. 3, no. 1, 2021.
- [15] Cahya, "Contoh implementasi Data Mining Algoritma k-Nearest Neighbors (k-NN) menggunakan PHP dan MySQL untuk memprediksi kelulusan mahasiswa tepat waktu," <https://extra.cahyadsn.com/knn>.
- [16] D. Feby, "Apa Itu Decision Tree di Machine Learning Model?," <https://dqlab.id/apa-itu-decision-tree-di-machine-learning-model>.
- [17] I. Khoeri and D. I. Mulyana, "Implementasi Machine Learning dengan Decision Tree Algoritma C4.5 dalam Penerimaan Karyawan Baru pada PT. Gitareksa Dinamika Jakarta," *Jurnal Sosial dan Teknologi (SOSTECH)*, vol. 1, no. 7, Jul. 2021, [Online]. Available: <http://sostech.greenvest.co.id>
- [18] A. C. Dewi, "Klasifikasi menggunakan Algoritma Decision Tree," <https://agneschintiadewi.medium.com/klasifikasi-menggunakan-algoritma-decision-tree-446d500ba73c>.
- [19] Syarli and A. A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *Jurnal Ilmiah Ilmu Komputer*, vol. 2, no. 1, Apr. 2016, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
- [20] A. Desiani, M. Akbar, I. Irmeilyana, and A. Amran, "Implementasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular," *Jurnal Teknik Elektro dan Komputasi (ELKOM)*, vol. 4, no. 2, pp. 207–214, Aug. 2022, doi: 10.32528/elkom.v4i2.7691.

Yosia Amadeus Ishak, mahasiswa program studi Teknik Informatika, Fakultas Teknologi Informasi di Universitas Tarumanagara.

Manatap Dolok Lauro, memperoleh gelar S.Kom. dari Universitas Tarumanagara, Jakarta tahun 2006. Kemudian tahun 2010 memperoleh MMSI dari Universitas Bina Nusantara, Jakarta. Saat ini sebagai Staf Pengajar program studi Teknik Informatika Universitas Tarumanagara.