

PENERAPAN MACHINE LEARNING DALAM PREDIKSI TINGKAT KASUS PENYAKIT DI INDONESIA

Rakha Gusti Wardhana ¹⁾, Gunawan Wang ²⁾, Farida Sibuea ^{3*)}

^{1, 2)} Master of Information Systems Management Bina Nusantara University

³⁾ Pusat Data dan Teknologi Informasi Kementerian Kesehatan Republik Indonesia

email : rakha.wardhana@binus.ac.id ¹⁾, gunawan.wang@binus.ac.id ²⁾, farida.sibuea@kemkes.go.id ³⁾

Abstraksi

Transformasi *digital* selalu menghadirkan upaya pengambilan keputusan yang berlandaskan data. Dalam rangka mendukung transformasi digital di bidang kesehatan, maka diperlukan perangkat yang dapat melakukan proses prediksi untuk mendukung pengambilan keputusan. Penerapan machine learning untuk proses prediksi tingkat persebaran penyakit dapat membantu pengambil kebijakan dalam merumuskan kebijakan pelayanan kesehatan secara cepat. Pemodelan machine learning menggunakan berbagai teknik dalam algoritma klasifikasi dengan *framework* yang dimiliki oleh bahasa pemrograman *python*. Dalam hal ini, data yang digunakan merupakan dataset yang dikoleksi dari Profil Kesehatan Indonesia, yang berisi informasi tingkat pelayanan kesehatan, kebijakan kesehatan, serta tingkat sosio-demografi. Pemodelan dengan teknik *decision tree* memiliki akurasi yang sangat baik dalam memprediksi klasifikasi tingkat kasus penyakit.

Kata Kunci :

Transformasi digital, Prediksi Penyakit, *Machine Learning*, Klasifikasi.

Abstract

Digital transformation always presents data-driven decision-making efforts. In order to support digital transformation in the health sector, a device is needed that can perform predictive processes to support decision making. The application of machine learning for the process of predicting the level of disease spread can assist policy makers in formulating health service policies quickly. Machine learning modeling uses methods by classification algorithms with a framework owned by the Python programming language. In this case, the data used is a dataset collected from the Indonesian Health Profile, which contains information on the level of health services, health policies, and the socio-demographic level. Decision tree modelling gained good accuracy when predicting the classification of disease cases.

Keywords:

Digital Transformation, Disease Prediction, Machine Learning, Classification.

Pendahuluan

Revolusi industri 4.0 telah mendorong diterapkannya transformasi digital bagi semua kalangan, termasuk dalam skala pemerintahan. Pemerintah Indonesia telah menetapkan arah transformasi digital hingga tahun 2024 [1]. Salah satu dari program transformasi digital tersebut berupa proses pelayanan dan kebijakan masyarakat yang berlandaskan data. Dalam hal ini, Kementerian Kesehatan Republik Indonesia turut berpartisipasi dalam transformasi digital dengan mengeluarkan Cetak Biru Strategi Transformasi Digital Kesehatan 2024, dimana salah satunya mengulas permasalahan kebijakan kesehatan yang belum berlandaskan data [2]. Pengambilan kebijakan kesehatan sudah seyogyanya memperhatikan kondisi di masing-masing daerah dalam kurun waktu tertentu, sehingga implementasi kebijakan dapat diterapkan dengan baik

Berkaca dari Cetak Biru Strategi Transformasi Digital Kesehatan 2024 tersebut, pihak Kementerian Kesehatan tentu harus merujuk pada kondisi pelayanan kesehatan, serta kondisi persebaran berbagai penyakit pada masing-masing wilayah di Indonesia. Dalam hal ini, variasi tingkat persebaran penyakit maupun tingkat pelayanan kesehatan layaknya program vaksinasi, sanitasi, serta ketersediaan rumah sakit pada masing-masing periode tersusun secara numerik dalam datasets yang telah disusun oleh tim Pusat Data dan Teknologi Informasi Kementerian Kesehatan. *Datasets* tersebut dikenal dengan program Profil Kesehatan Indonesia, dimana didalamnya terdapat datasets dan infografis yang menyediakan berbagai data dan informasi, mulai dari tingkat demografi, tingkat pendidikan, hingga tingkat vaksinasi dan angka penyakit di seluruh Indonesia [3]. Dalam hal ini, Kementerian

Kesehatan sudah memiliki datasets yang dihimpun secara terpadu.

Dalam mendukung transformasi digital, pengambilan keputusan terkait program vaksinasi, sanitasi, dan program kesehatan lainnya harus didukung dengan perangkat yang dapat mengolah data untuk menyajikan *insight* yang bersifat prediktif. *Machine Learning* dapat mengoptimalkan proses pengambilan keputusan dengan menyajikan *insight* yang bersifat prediktif dan berjalan secara otomatis. Pembelajaran mesin/*Machine Learning* adalah studi berkelanjutan tentang konsep pengenalan pola dan pembelajaran komputasi dalam kecerdasan buatan yang menggunakan algoritma pembelajaran seperti diawasi dan tidak diawasi untuk memprediksi dan mendukung pengambilan keputusan otomatis berdasarkan sekumpulan data [4]. *Machine Learning* tersebut dibangun dengan *framework Data Preparation*, serta *framework Data Modelling* yang dapat diluncurkan, dan berjalan secara otomatis. Dengan model pembelajaran mesin, Pemangku Kepentingan dapat membuat studi prediksi dari kumpulan datasets Profil Kesehatan Indonesia yang membantu mereka membuat keputusan.

Penelitian ini bertujuan untuk menjabarkan implementasi *machine learning* untuk memprediksi tingkat kasus berbagai penyakit menular di Indonesia. Berbagai penelitian mengenai prediksi tingkat persebaran penyakit telah banyak dilakukan. Dalam salah satu *literature review* dari berbagai studi kasus, pengaplikasian *machine learning* dengan pemodelan ARIMA untuk memprediksi penyakit COVID-19 telah menghasilkan prediksi *time series* dengan akurasi yang terbaik [5]. Kemudian, implementasi *machine learning* untuk memprediksi tingkat kasus penyakit *Tuberculosis* di Gambok, Malaysia dengan metode Regresi Linear dan ANN berdasarkan parameter sosio-demografi dan lingkungan telah menghasilkan *margin error* yang rendah [6]. Selanjutnya, pemodelan yang dilakukan untuk memprediksi kasus COVID-19 dari lima negara yang berbeda dalam beberapa kurun waktu menghasilkan kesimpulan bahwa algoritma MLP dan ANFIS memiliki tingkat *margin error* yang rendah, serta tingkat efektifitas yang terbaik dari algoritma lainnya [7].

Dari berbagai penelitian tersebut, penelitian ini berusaha untuk menawarkan implementasi *machine learning* untuk membangun model prediksi tingkat kasus berbagai penyakit dari seluruh provinsi di Indonesia. Pemodelan *machine learning* menggunakan berbagai algoritma klasifikasi dengan atribut berupa tingkat pelayanan kesehatan, serta kondisi lingkungan. Kemudian, tingkat kasus penyakit yang menjadi atribut target prediksi merupakan jumlah kasus dari berbagai penyakit menular dari seluruh provinsi pada masing-masing periode tertentu yang telah dinormalisasi ke dalam lima kelas (1-5), dimana semakin tinggi kelasnya, maka semakin tinggi tingkat kasusnya. Penggunaan

algoritma klasifikasi sesuai dengan keperluan untuk memprediksi tingkat kasus yang telah dinormalisasikan ke dalam kelas tertentu. Algoritma klasifikasi dapat digunakan untuk memprediksi kategorisasi dari berbagai parameter, seperti kasus prediksi klasifikasi kategori seseorang terjangkit COVID-19 dengan parameter umur, jenis kelamin, kondisi spesimen, serta kategori penularan [8]. Penelitian ini menggunakan framework bahasa pemrograman *python*.

Tinjauan Pustaka

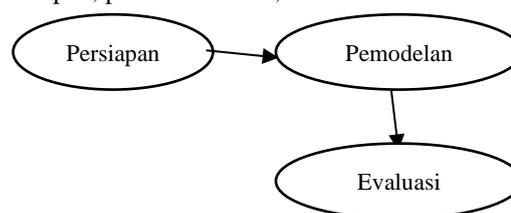
Penelitian terdahulu yang memprediksi jumlah kasus beberapa penyakit di suatu wilayah telah banyak dilakukan dengan pemodelan/algoritma regresi. Akan tetapi, penelitian yang berupaya untuk memprediksi tingkat kasus berdasarkan klasifikasi tinggi/rendahnya persebaran penyakit belum banyak dilakukan. Dalam hal ini, penelitian yang dilakukan merujuk pada penelitian yang menggunakan variabel tingkat sosio-demografi [5]. Dalam hal ini, penelitian kali ini menggunakan variabel sosio-demografi, beserta variabel tingkat pelayanan kesehatan untuk memprediksi tingkat kasus penyakit.

Penelitian dengan algoritma klasifikasi terhadap kasus penyakit telah dilakukan untuk memprediksi kelompok yang rentan berdasarkan kategori penularan dengan pemodelan decision tree, naïve bayes, dan lainnya [8]. Dari pemaparan tersebut, penggunaan algoritma klasifikasi pada penelitian ini dilakukan dengan tujuan yang berbeda, dimana algoritma klasifikasi ditujukan untuk memprediksi klasifikasi penilaian tingkat kasus penyakit mulai dari bilangan 1 (angka kasus rendah) hingga angka bilangan 5 (angka kasus tinggi).

Kemudian, penelitian ini juga membandingkan beberapa algoritma klasifikasi untuk mencari pemodelan yang memiliki akurasi terbaik pada lima kelompok penyakit, yakni Tuberkulosis, Diare, DBD, Campak, Malaria, Kusta, Difteri. Perbandingan ini dijalankan sebagai bagian dari implementasi metode *machine learning*, dimana dalam salah satu tahapan metode *machine learning*, terdapat upaya pemodelan/pembelajaran untuk mencari teknik algoritma yang paling sesuai dan akurat [9].

Metode Penelitian

Proses implementasi *machine learning* dapat dirangkum ke dalam tiga tahapan, yakni tahap persiapan, pemodelan data, dan evaluasi.



Gambar 1. Tahapan Penelitian

Tahap persiapan diawali dengan memilah atribut/parameter yang akan digunakan dalam pemodelan *machine learning*. Korelasi dari masing-masing parameter tersebut dibangun berdasarkan hipotesis awal [10]. Hipotesis awal yang dibangun pada penelitian ini merujuk pada kondisi dimana tingkat kasus penyakit berbanding lurus dengan aspek sosio-demografi, kondisi lingkungan, serta tingkat pelayanan kesehatan yang pernah diujikan pada penelitian terdahulu. Kemudian, tahap persiapan dilanjutkan dengan pengumpulan data dari Profil Kesehatan Indonesia dari tahun 2016 hingga 2021. Dari pengumpulan data tersebut, Variabel/parameter *x* yang digunakan dalam penelitian ini yaitu:

Tahun, Provinsi, Nama Penyakit, Jumlah Kasus, Persentase Penduduk Miskin, Indeks Pembangunan Manusia, Kepadatan Penduduk per KM, Persentase Kabupaten/Kota melaksanakan Kebijakan GERMAS, Tempat Pengolahan Pangan yang Memenuhi Syarat Sesuai Standar, Persentase Tempat dan Fasilitas Umum yang Dilakukan Pengawasan Sesuai Standar, Jumlah Kabupaten/Kota yang Melakukan Pelayanan Terpadu (PANDU) di setiap Puskesmas Menurut Provinsi, Persentase Puskesmas dengan Ketersediaan Vaksin Imunisasi Dasar Lengkap Menurut Provinsi, Persentase Kabupaten/Kota Melaksanakan Pembinaan Posyandu Aktif, Jumlah Sarana Produksi Bidang Kefarmasian dan Alat Kesehatan Menurut Provinsi, Rasio Puskesmas per Kecamatan, Jumlah Klinik Pratama dan Klinik Utama Teregistrasi Menurut Kepemilikan dan Provinsi, Rasio Puskesmas per Kecamatan, Jumlah Rumah Sakit dan Rasio Tempat Tidur Rumah Sakit Per 1000 Penduduk Menurut Provinsi

Kemudian, data-data tersebut dikumpulkan dalam satu dataset yang dipilah dan dibersihkan dengan metode *data cleansing* dan *data formatting* untuk mendapatkan data-data yang sesuai dengan parameter yang dibutuhkan. *Data cleansing* dilakukan dengan menghilangkan baris data yang kurang lengkap serta tidak akurat, sedangkan *data formatting* dilakukan dengan merubah format data agar sesuai dengan karakteristik data yang dibutuhkan untuk pemodelan [11].

PERSENTASE PUSKESMAS DENGAN KETERSEDIAAN VAKSIN IDL (IMUNISASI DASAR LENGKAP) MENURUT PROVINSI	PERSENTASE KABUPATEN/KOTA MELAKSANAKAN PEMBINAAN POSYANDU AKTIF
96,10	21,70
91,90	21,20
100,00	68,40
99,10	75,00
93,60	45,50

Gambar 2. Screenshot Contoh Data Float

Gambar diatas merupakan contoh data float yang belum diproses oleh data formatting. Dalam hal ini, kondisi data yang bernilai *invalid* dan berbentuk *float* dengan koma tidak dapat diproses oleh *pandas*, sehingga perlu untuk dilakukan *data cleansing* dan *data formatting*. Berikut ini *pseudocode* dari proses *data cleansing* dengan *pandas* dan *numpy*.

```

import numpy as np
df.dropna()
df = df.replace('-', np.nan).astype('object')
    
```

Selanjutnya, *data formatting* diimplementasikan dengan menjalankan *pseudocode* kepada variabel/parameter *x* dengan fungsi sebagai berikut

```

Df['parameter/variabel x'] = Df['parameter/variabel x'].str.replace(',', '.').astype(float)
    
```

dengan fungsi tersebut, data numerik yang belum sesuai (angka dengan format ',') akan dinormalisasikan ke dalam angka bertipe *float*. Kemudian, tahap data formatting dilanjutkan dengan proses normalisasi parameter/atribut target (angka jumlah kasus penyakit) menjadi lima kelas (1-5). Proses normalisasi dilakukan dengan fungsi normalisasi dari *scikit-learn* dengan fungsi '*KBinsDiscretizer*'. Fungsi tersebut akan bekerja dengan menerapkan strategi *clustering* (*K-Means*) untuk mengelompokkan nilai bilangan numerik secara unsupervised untuk mengelompokkan nilai bilangan numerik tersebut ke dalam 5 jumlah *cluster* tertentu. Sehingga, fungsi tersebut akan merubah variabel Jumlah Kasus Penyakit yang bersifat *continuous* menjadi bersifat biner dengan rentang 1 hingga 5.

Tabel 1. Penjabaran Fungsi Normalisasi Variabel Target

Parameter KBinsDiscretizer	Input
n_bins (jumlah cluster)	5
Encode (jenis data)	ordinal
Strategy (algoritma)	k-means

Kemudian, penelitian ini dilanjutkan dengan upaya pemodelan. Pemodelan pada penelitian ini dilakukan dengan berbagai teknik algoritma klasifikasi. Secara umum, pemodelan klasifikasi dilakukan dengan memprediksi kemungkinan kelas dari suatu variabel data [12]. Dalam hal ini, pemodelan dilakukan dengan menerapkan algoritma *decision tree*, *random forest*, *logistic regression*, *support vector machine* (*SVM*), dan *xgboost* dengan *framework/library scikit-learn* dari *Python*, yang nantinya akan dilanjutkan dengan tahap perbandingan akurasi dari masing-masing algoritma tersebut. Secara terperinci, algoritma *decision tree* bekerja dengan mengibaratkan struktur pohon yang memiliki atribut/fitur dengan hubungan yang berbeda, sedangkan algoritma *random forest* merupakan

sekumpulan proses *decision tree* yang diproses secara paralel [13].

Tabel 2. Penjabaran Fungsi Algoritma dari Library Scikit-learn

Model	Fungsi Scikit-learn	Parameter
Logistic Regression	LogisticRegression()	Random_state=0
Random Forest	RandomForestClassifier()	n_estimators=1000, oob_score = True, n_jobs=-1 random_state =50 max_features="auto" max_leaf_nodes=30
SVM	SVC()	kernel=linear
Decision Tree	DecisionTreeClassifier	
XGBoost	XGBClassifier()	n_estimators = 400, learning_rate = 0.1, max_depth = 3

Setelah proses pemodelan dilakukan, masing-masing model akan dievaluasi dengan membandingkan masing-masing akurasi dari kelima teknik algoritma tersebut pada masing-masing jenis penyakit. Perhitungan akurasi dilakukan setelah fungsi scikit-learn dari masing-masing algoritma diterapkan. Perhitungan akurasi dilakukan dengan mengimplementasikan rumus scikit-learn berikut

```
sklearn.metrics.accuracy_score(y_true,y_pred,normalize=True, sample_weight=None)
```

Hasil dan Pembahasan

Dari proses implementasi tahapan persiapan, pemodelan, hingga evaluasi, penelitian ini akan menjabarkan hasil dari ketiga proses tersebut. Tahap persiapan yang meliputi proses pengumpulan data, *data cleansing*, serta *data formatting* telah menghasilkan data yang siap digunakan untuk pemodelan. Dalam hal ini, proses data cleansing telah berhasil menghilangkan baris data yang memiliki nilai *null*. Kemudian, proses *data formatting* telah berhasil merubah jenis data numerik pada beberapa baris data dari berbagai variabel x agar sesuai dengan tipe *float*. Selanjutnya, proses data formatting juga berhasil menjalankan normalisasi terhadap variabel/atribut target, dimana proses *clustering* berhasil mengelompokkan seluruh data variabel target (jumlah kasus) yang bernilai *continuous* ke dalam beberapa *cluster*/kelas dengan index 1 hingga index 5, dan membuat variabel baru yang mewakili (jumlah_kasus_clustered).

PERSENTA	PERS	JUMLAH SAR	RASIO	JUMLAH KLIP	JUMLAH RUM	jumlah_kasus_clustered
MENURUT PR						
96.1	21.7	2.0	1.2	2220.0	10611.0	0.0
91.9	21.2	25.0	1.4	2270.0	24358.0	0.0
100.0	68.4	0.0	1.6	1510.0	7558.0	0.0
99.1	75.0	0.0	1.4	1060.0	8438.0	0.0
93.6	45.5	0.0	1.4	750.0	4746.0	0.0
100.0	52.9	3.0	1.4	3320.0	9750.0	0.0
100.0	90.0	0.0	1.4	400.0	2844.0	0.0
95.8	86.7	2.0	1.4	2000.0	8186.0	0.0
100.0	57.1	0.0	1.4	870.0	2427.0	0.0
100.0	71.4	19.0	1.2	1100.0	4036.0	0.0
100.0	83.3	206.0	7.2	6390.0	25222.0	1.0
92.8	44.4	549.0	1.7	1623.0	52298.0	1.0
82.3	88.6	177.0	1.5	114.0	45901.0	0.0
100.0	100.0	27.0	1.6	1330.0	7143.0	0.0
98.2	81.6	191.0	1.5	1111.0	51100.0	1.0
99.6	37.5	214.0	1.6	3620.0	13426.0	0.0
95.8	44.4	5.0	2.1	1520.0	8624.0	0.0
100.0	100.0	4.0	1.5	1260.0	6046.0	0.0
92.5	4.5	0.0	1.4	660.0	5601.0	0.0
100.0	14.3	2.0	1.4	660.0	6556.0	1.0
93.5	78.6	0.0	1.5	560.0	4002.0	0.0
97.5	100.0	0.0	1.5	710.0	6478.0	0.0
100.0	50.0	0.0	1.8	1290.0	6786.0	0.0
100.0	20.0	0.0	1.2	170.0	6833.0	0.0

Gambar 2. Screenshot Hasil Proses Data Formatting

Dari gambar diatas, proses *data formatting* merubah angka numerik yang dilengkapi tanda koma (100,0), menjadi tanda titik (100.0). Kemudian, data-data variabel jumlah kasus dinormalisasi dengan proses *clustering* yang dilakukan oleh fungsi '*KBinsDiscretize*'. Dari proses tersebut, muncul variabel baru (jumlah_kasus_clustered) yang berisi nilai dari proses normalisasi data jumlah kasus kedalam lima kelas dengan index 1.0, 2.0, 3.0, 4.0, dan 5.0. Hal ini sejalan dengan konsep diskritisasi dengan metode *clustering*, konsep tersebut menguraikan proses dimana metode *clustering* dalam diskretisasi menetapkan secara acak titik pusat dari cluster data, dan menetapkan masing-masing nilai dengan titik pusat terdekat [14]. Dengan melakukan tahap persiapan, maka dataset telah siap untuk diolah dalam tahap pemodelan [15]. Pemodelan dilakukan dengan mengimplementasikan seluruh fungsi algoritma *scikit-learn* untuk menjalankan pemodelan yang telah dijabarkan pada metode penelitian. Dalam proses pemodelan, dataset hasil dari proses data cleansing akan diolah dengan proses *training* serta *testing* dengan rumus berikut

```
X2021_train, X2021_test, y2021_train, y2021_test = train_test_split(X2021, y2021, test_size=0.4, random_state=101)
```

Dari rumus diatas, datasets yang telah dikelompokkan berdasarkan penyakit dibagi menjadi *data training* (data yang digunakan untuk pelatihan algoritma pemodelan) dengan komposisi 60% dari total keseluruhan baris data, serta data testing (digunakan untuk uji akurasi algoritma) sebesar 40% dari total keseluruhan baris data. Proses pemodelan dilakukan terhadap dataset yang telah dikelompokkan pada masing-masing kategori penyakit. Untuk penyakit *tuberculosis*, pemodelan dengan teknik *decision tree* dan *XGBoost* menghasilkan akurasi yang paling tinggi dibandingkan pemodelan lainnya.

Tabel 3. Hasil Pemodelan Penyakit Tuberculosis

Pemodelan	Accuracy Score
Logistic Regression	0.75
Random Forest	0.75
Support Vector Machine	0.75
Decision Tree	0.83
XGBoost	0.80

Kemudian, pemodelan terhadap penyakit diare menghasilkan akurasi tertinggi dengan teknik *decision tree* serta *random forest*.

Tabel 4. Hasil Pemodelan Penyakit Diare

Pemodelan	Accuracy Score
Logistic Regression	0.5
Random Forest	0.91
Support Vector Machine	0.52
Decision Tree	0.94
XGBoost	0.88

Pemodelan terhadap penyakit DBD menghasilkan akurasi yang cukup tinggi dengan teknik Decision Tree dan XGBoost.

Tabel 5. Hasil Pemodelan Penyakit DBD

Pemodelan	Accuracy Score
Logistic Regression	0.75
Random Forest	0.81
Support Vector Machine	0.78
Decision Tree	0.83
XGBoost	0.86

Teknik *Decision Tree* terhadap penyakit Malaria juga memiliki akurasi yang cukup tinggi dibandingkan dengan teknik lainnya.

Tabel 6. Hasil Pemodelan Penyakit Malaria

Pemodelan	Accuracy Score
Logistic Regression	0.67
Random Forest	0.78
Support Vector Machine	0.67
Decision Tree	0.81
XGBoost	0.78

Untuk penyakit Kusta, pemodelan dengan teknik random forest, decision tree, dan XGBoost memiliki tingkat akurasi sebesar 0.72, yang merupakan tingkat akurasi tertinggi dibandingkan teknik lainnya.

Tabel 7. Hasil Pemodelan Penyakit Kusta

Pemodelan	Accuracy Score
Logistic Regression	0.69
Random Forest	0.72
Support Vector Machine	0.69
Decision Tree	0.72
XGBoost	0.72

Berbeda dengan sebelumnya, pemodelan dengan teknik decision tree pada penyakit difteri justru tidak memiliki akurasi yang baik. Dalam hal ini, teknik *logistic regression*, *random forest*, dan *support*

vector machine memiliki kesamaan akurasi yang melebihi akurasi decision tree.

Tabel 8. Hasil Pemodelan Penyakit Difteri

Pemodelan	Accuracy Score
Logistic Regression	0.82
Random Forest	0.82
Support Vector Machine	0.82
Decision Tree	0.65
XGBoost	0.80

Dari hasil seluruh proses pemodelan terhadap enam jenis penyakit, proses evaluasi terhadap seluruh pemodelan diatas menyimpulkan bahwa pemodelan dengan teknik *decision tree* dalam memprediksi klasifikasi tingkat kasus penyakit berdasarkan tingkat pelayanan kesehatan, serta tingkat sosio-demografi penduduk dari masing-masing provinsi merupakan pemodelan yang paling akurat. Berdasarkan hal tersebut, maka pihak pengambil keputusan dapat melakukan simulasi untuk menentukan sejauh mana pencapaian tingkat pelayanan kesehatan terhadap tingkat kasus penyakit dengan pemodelan *machine learning* dengan algoritma klasifikasi.

Kesimpulan dan Saran

Hasil dari penelitian ini mengungkapkan bahwa proses implementasi machine learning terhadap prediksi tingkat kasus penyakit dengan algoritma klasifikasi dapat dilakukan dengan melakukan upaya normalisasi angka kasus penyakit, serta menerapkan teknik pemodelan dengan algoritma klasifikasi. Pemodelan dengan teknik decision tree dalam penelitian ini menghasilkan akurasi yang sangat baik. Dalam hal ini, pihak pengambil keputusan dapat melakukan upaya prediksi tingkat kasus penyakit sebagai bahan simulasi untuk menyusun target pencapaian tingkat pelayanan kesehatan pada periode berjalan. Proses pengambilan keputusan yang berlandaskan data tentu sejalan dengan upaya transformasi *digital*.

Untuk penelitian selanjutnya, penggunaan variabel-variabel tingkat pelayanan kesehatan, beserta aspek sosio-demografi masyarakat dapat menjadi bahan penelitian selanjutnya, terutama dengan mengimplementasikan penggunaan algoritma regresi, maupun forecasting dalam memprediksi serta meramalkan angka kasus penyakit secara berkelanjutan.

Daftar Pustaka

- [1] Dewan Teknologi Informasi dan Komunikasi Nasional, "Arah Transformasi Digital Indonesia", 2020. <http://www.wantiknas.go.id/>
- [2] Kementerian Kesehatan Republik Indonesia, "Cetak Biru Strategi Transformasi Digital Kesehatan 2024", 2021

- [3] Pusat Data dan Teknologi Informasi, “Profil Kesehatan Indonesia”, 2021
- [4] R. Abellera, and R. Bulusu, “Oracle Business Intelligence with Machine Learning: Artificial Intelligence Techniques in OBIEE for Actionable BI”, NewYork: Apress, 2018
- [5] Fard, et al., “Application of Machine Learning in the Prediction of Covid-19 daily new cases: A Scoping Review.”, *Heliyon*, 7(10), 2021.
- [6] Mohidem, et al., “Prediction of tuberculosis cases based on socio demographic and environmental factors in gombak, Selangor, Malaysia: A comparative assessment of multiple linear regression and artificial neural network models”, *International journal of mycobacteriology*, 10(4), pp. 442–456, 2021. https://doi.org/10.4103/ijmy.ijmy_182_2
- [7] Ardabili, et al., “COVID-19 Outbreak Prediction with Machine Learning”. *Feature Papers in Evolutionary Algorithms and Machine Learning*, 3(10), 2021. <https://doi.org/10.3390/a13100249>
- [8] J. Kalezhi, et al., “Modelling Covid-19 Infections in Zambia Using Data Mining Techniques”. *Results in Engineering*, 13, 2022.
- [9] D. Lion, “Artificial Inteligence: A Beginner Guide”, 2021.
- [10] M. Kantardzic, “Data Mining: Concepts, Models, Methods, and Algorithms” Third Edition. New York: Wiley. 2020
- [11] C. Lopez, “Data Mining: CRISP-DM Methodology, The Clem Language, and IBM SPSS Modeler”, 2021.
- [12] D. Berrar, “Bayes’ Theorem and Naive Bayes Classifier”. *Reference Module in Life Sciences*, 1, 2018.
- [13] P. Jaramillo, and D. Juan, “A Reviews of Machine Learning Algorithms Used for Modelling Travel Mode Choice” *DYNA*, 86(211), pp. 32-41, 2019. <https://doi.org/10.15446/dyna.v86n211.79743>
- [14] R. Dash, R. Paramguru, and R. Dash, “Comparative Analysis of Supervised and Unsupervised Discretization Techniques”. *International Journal of Advances in Science and Technology*, 2(3), 2011.
- [15] K. Kharisma and U. S. Aesy, “Analisis Tingkat Kebermanfaatan MyPertamina Menggunakan K-Means Clustering”, *JOISM*, 4(2), pp. 91 - 96, 2023. <https://doi.org/10.24076/joism.2023v4i2.982>