

Klasifikasi Status Indeks Desa Membangun Jawa Barat Menggunakan Algoritma XGBoost

Shafira Agnia Latfalia, Reny Rian Marliana*

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

ARTICLE INFO

Article history :

Received : 27/09/2024
Revised : 27/12/2024
Published : 29/12/2024



Creative Commons Attribution-ShareAlike 4.0 International License.

Volume : 4
No. : 2
Halaman : 75 - 82
Terbitan : **Desember 2024**

Terakreditasi Sinta [Peringkat 4](#)
berdasarkan Ristekdikti
No. 177/E/KPT/2024

ABSTRAK

Data Badan Pusat Statistik tahun 2020 menunjukkan bahwa rata-rata wilayah perdesaan di Jawa Barat memiliki persentase kemiskinan sebesar 10,64% yang lebih tinggi dari wilayah perkotaan sebesar 7,79%. Dalam upaya melaksanakan pembangunan desa yang terukur dan berkelanjutan, Kementerian Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi Republik Indonesia menyusun suatu indeks komposit untuk menentukan status kemajuan suatu desa yang disebut Indeks Desa Membangun (IDM). Untuk mengatasi tantangan waktu dalam pengklasifikasian status IDM di Jawa Barat dengan jumlah desa yang sangat besar, diperlukan algoritma lain yang lebih efektif. XGBoost dipilih karena mampu mengatasi kompleksitas dataset besar dengan kelas data yang tidak seimbang dan dapat mencegah *overfitting*. Selain itu, dilakukan *hyperparameter tuning* untuk meningkatkan performa model. Tujuan dari artikel ini yaitu melihat seberapa akurat algoritma XGBoost dalam mengklasifikasikan status IDM. Berdasarkan analisis yang dilakukan pada data status IDM di Provinsi Jawa Barat tahun 2020 menggunakan 52 variabel dengan data *training* sebanyak 4.249 dan data *testing* sebanyak 1.063, telah dibentuk model XGBoost. Hasil analisis menunjukkan bahwa nilai *accuracy*, *precision*, *recall*, dan *f1-score* masing-masing sebesar 89%, 84%, 72%, dan 76%. Nilai *accuracy* yang diperoleh sudah tinggi sehingga model XGBoost yang dibangun sudah dapat mengklasifikasikan status IDM dengan baik dan dapat diimplementasikan.

Kata Kunci : *Indeks Desa Membangun, Klasifikasi, XGBoost*

ABSTRACT

Based on data from Statistics Indonesia 2020 shows that rural areas in West Java have an average poverty rate of 10,64%, which is higher than urban areas at 7,79%. To establish a measurable and sustainable village, the Ministry of Villages, Development of Disadvantaged Regions, and Transmigration of the Republic of Indonesia developed a composite index to determine the progress of a village, called the Village Development Index. To overcome the time challenge in classifying Village Development Index in West Java with a very large number of villages, an effective algorithm is needed. XGBoost was chosen because it can handle the complexity of large datasets with imbalanced data classes and can prevent overfitting. In addition, hyperparameter tuning is conducted to improve the model's performance. The aim of this research is to determine how accurately the XGBoost algorithm can classifying the Village Development Index. Based on the analysis conducted on Village Development Index data in West Java 2020 using 52 variables with 4.249 training data and 1.063 testing data, the XGBoost model has been formed. The analysis results show that the accuracy, precision, recall and f1-score are 89%, 84%, 72% and 76% respectively. The high accuracy obtained indicates that the XGBoost model that is built can classify Village Development Index well and can be implemented.

Keywords : *Classification, Developing Village Index, XGBoost*

Copyright© 2024 The Author(s)..

A. Pendahuluan

Indonesia merupakan negara kepulauan yang terdiri atas desa-desa yang tersebar luas. Dinas Pemberdayaan Masyarakat dan Desa tahun 2020 menyatakan bahwa jumlah desa di Provinsi Jawa Barat ada sebanyak 5.312 [1]. Berdasarkan data Badan Pusat Statistik tahun 2020, wilayah perdesaan di Jawa Barat memiliki persentase kemiskinan sebesar 10,64% yang lebih tinggi dari wilayah perkotaan sebesar 7,79%. Sehingga pembangunan desa diperlukan untuk mengurangi kesenjangan antara wilayah perdesaan dan perkotaan akibat adanya pembangunan ekonomi yang lebih berfokus di perkotaan [2].

Dalam upaya melaksanakan pembangunan desa yang terencana dan *sustainable*, Kementerian Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi Republik Indonesia kemudian menyusun Indeks Desa Membangun (IDM). IDM merupakan suatu indeks komposit yang digunakan untuk menentukan status kemajuan dan kemandirian suatu desa serta mendukung upaya pemerintah dalam mengatasi permasalahan desa tertinggal dan meningkatkan desa mandiri [3]. Berdasarkan Pasal 3 Peraturan Menteri Nomor 2 Tahun 2016 bahwa IDM didasarkan pada tiga indeks, yaitu: 1) Indeks Ketahanan Sosial (IKS); 2) Indeks Ketahanan Ekonomi (IKE); dan 3) Indeks Ketahanan Ekologi/Lingkungan (IKL) [4].

Sesuai dengan Pasal 5 Peraturan Menteri Nomor 2 Tahun 2016, status kemajuan dan kemandirian desa yang ditetapkan berdasar IDM dapat diklasifikasikan ke dalam 5 status desa, yaitu: 1) Mandiri; 2) Maju; 3) Berkembang; 4) Tertinggal; dan 5) Sangat Tertinggal. Klasifikasi desa dibuat untuk menunjukkan variasi karakteristik yang dimiliki setiap desa, meningkatkan status kemajuan desa, dan memberikan rekomendasi intervensi kebijakan yang diperlukan [5].

Pengklasifikasian status IDM dilakukan setiap satu tahun sekali dan membutuhkan waktu kurang lebih 8-9 bulan. Untuk mengatasi tantangan waktu dalam pengklasifikasian status IDM dengan jumlah desa yang sangat besar, diperlukan algoritma lain yang lebih efektif [6]. Data mining dengan teknik klasifikasi dapat dilakukan untuk membantu dalam mengidentifikasi status IDM [7]. Dalam hal ini, penerapan algoritma *ensemble* dapat menjadi solusi yang efektif untuk mengatasi hambatan waktu pengklasifikasian yang lama [8].

Algoritma *ensemble* merupakan pendekatan yang menggabungkan beberapa model *machine learning* untuk meningkatkan kinerja dan keakuratan hasil klasifikasi dibandingkan dengan menggunakan model tunggal, contohnya algoritma XGBoost (*Extreme Gradient Boosting*) [9]. XGBoost dipilih karena mampu mengatasi kompleksitas dataset besar dengan kelas data yang tidak seimbang dan dapat mencegah *overfitting* [10]. Artikel ini menggunakan 52 variabel prediktor untuk mengklasifikasikan status IDM di Provinsi Jawa Barat [11] [12]. Variabel-variabel tersebut mencakup tiga aspek, yaitu sosial, ekonomi, dan ekologi/lingkungan. Setiap variabel berkontribusi dalam menentukan status kemajuan dan kemandirian desa, sehingga penggunaan algoritma XGBoost diharapkan dapat menghasilkan model yang baik dalam mengklasifikasikan status IDM.

Tujuan dari artikel ini adalah mengetahui seberapa akurat algoritma XGBoost dalam mengklasifikasikan status IDM di Provinsi Jawa Barat tahun 2020. Dengan model XGBoost yang akurat, artikel ini memiliki potensi untuk menciptakan dampak positif yang langsung dirasakan oleh masyarakat desa serta pihak-pihak yang terlibat dalam proses pembangunan di daerah. Dengan demikian, banyaknya jumlah desa tidak lagi menjadi hambatan utama, melainkan menjadi tantangan yang dapat diatasi melalui pemanfaatan teknologi dan algoritma yang modern. Dengan pendekatan ini, waktu pengklasifikasian dapat dioptimalkan sehingga memungkinkan proses pengambilan keputusan terkait pembangunan desa dapat dilakukan lebih cepat dan efisien.

B. Metode Penelitian

Artikel ini menggunakan algoritma XGBoost dengan data sekunder yang bersumber dari website open data Jabar yang dihasilkan oleh Dinas Pemberdayaan Masyarakat dan Desa, yaitu data status Indeks Desa Membangun Provinsi Jawa Barat tahun 2020. Unit penelitian yang digunakan adalah 5.312 desa dan variabel yang digunakan terdiri atas 1 variabel respon (Y) dengan 4 kelas/label, yaitu Mandiri, Maju, Berkembang, dan Tertinggal serta 52 variabel prediktor (X).

C. Hasil dan Pembahasan

Pengecekan *Missing Value*

```
#cek missing value
print(df.isnull().values.any())

False
```

Gambar 1. Hasil Pengecekan *Missing Value*

Berdasarkan Gambar 1 diketahui bahwa tidak terdapat *missing value* dalam data sehingga tidak diperlukan imputasi atau pengelolaan khusus terhadap nilai yang hilang.

Pengecekan Data Duplikat

```
#cek data duplikat
print(df.duplicated().sum())

0
```

Gambar 2. Hasil Pengecekan Data Duplikat

Berdasarkan Gambar 2 diketahui bahwa tidak terdapat data duplikat sehingga dapat disimpulkan bahwa setiap baris dalam *data frame* adalah unik.

Mengubah Data Kategorik Menjadi Ordinal *Encoder*

Artikel ini terdiri dari 52 variabel prediktor (X) meliputi 28 variabel numerik yang tidak diberikan perlakuan khusus dan 24 variabel kategorik yang akan dilakukan transformasi data kategorikal dalam bentuk string menjadi bentuk numerik dengan ordinal *encoder*, dengan hasil sebagai berikut:

Tabel 1. Data Hasil Ordinal *Encoder*

X ₁	...	X ₄₃	X ₄₄	X ₄₅	X ₄₆	...	Y
5	...	1	1	1	3	...	2
5	...	1	1	1	3	...	3
2	...	1	0	1	3	...	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	...	1	0	1	3	...	3
2	...	1	0	1	3	...	3
4	...	1	0	1	3	...	3

Eksplorasi Data

Dengan menggunakan data sebanyak 5.312 desa berdasarkan 4 kelas/label, yaitu tertinggal, berkembang, maju, dan mandiri diperoleh frekuensi untuk masing-masing kelas/label adalah sebagai berikut:

Tabel 2. Frekuensi Kelas Status IDM

Kelas/Label	Jumlah
0 (Tertinggal)	120
1 (Berkembang)	3.291
2 (Maju)	1.631
3 (Mandiri)	270

Berdasarkan Tabel 4 diketahui bahwa sebagian besar desa di Provinsi Jawa Barat berada dalam kelas desa berkembang. Meskipun kelas desa tertinggal relatif sedikit, namun masih terlihat adanya perbedaan tingkat pembangunan antara desa-desa di Provinsi Jawa Barat.

Untuk mengetahui gambaran dan informasi mengenai data dari salah satu variabel prediktor yang berupa numerik, dilakukan analisis deskriptif dengan hasil sebagai berikut:

Tabel 3. Analisis Deskriptif Variabel Waktu Tempuh ke Sarana Kesehatan Terdekat (X_1)

Status IDM	Waktu Tempuh ke Sarana Kesehatan Terdekat (menit)			
	<i>count</i>	<i>mean</i>	<i>min</i>	<i>max</i>
0 (Tertinggal)	120	14,37	1	180
1 (Berkembang)	3.291	8,99	1	3.000
2 (Maju)	1.631	6,28	1	500
3 (Mandiri)	270	4,69	1	20

Berdasarkan Tabel 5 diketahui bahwa rata-rata waktu tempuh ke sarana kesehatan terdekat untuk status desa tertinggal adalah 14,37 menit, status desa berkembang adalah 8,99 menit, status desa maju adalah 6,28 menit, dan status desa mandiri adalah 4,69 menit. Ini berarti semakin tinggi status IDM, semakin singkat waktu tempuh ke sarana kesehatan terdekat. Artinya, desa dengan status IDM yang lebih maju atau mandiri cenderung memiliki akses yang lebih baik atau lebih dekat ke sarana kesehatan dibandingkan dengan desa yang tertinggal atau berkembang. Oleh karena itu, perlunya perhatian lebih terhadap akses terhadap sarana kesehatan untuk meningkatkan pelayanan kesehatan masyarakat.

Selain itu, untuk mengetahui gambaran dan informasi mengenai data dari salah satu variabel prediktor yang berupa kategorik, dilakukan analisis deskriptif dengan hasil sebagai berikut:

Tabel 4. Analisis Deskriptif Variabel Ketersediaan Pusat Kegiatan Belajar Masyarakat (X_{12})

Status IDM	Ketersediaan Pusat Kegiatan Belajar Masyarakat (PKBM)	
	0 (Tidak Ada)	1 (Ada)
0 (Tertinggal)	107	13
1 (Berkembang)	2.469	822
2 (Maju)	1.015	616
3 (Mandiri)	111	159

Berdasarkan Tabel 6 untuk diketahui bahwa dari 120 desa dengan status tertinggal terdapat 107 desa yang tidak memiliki PKBM dan 13 desa yang memiliki PKBM. Dari 3.291 desa dengan status berkembang terdapat 2.469 desa yang tidak memiliki PKBM dan 822 desa yang memiliki PKBM. Dari 1.631 desa dengan status maju terdapat 1.015 desa yang tidak memiliki PKBM dan 616 desa yang memiliki PKBM. Dari 270 desa dengan status mandiri terdapat 111 desa yang tidak memiliki PKBM dan 159 desa yang memiliki PKBM. Hal ini menunjukkan bahwa desa dengan status tertinggal memiliki akses yang sangat terbatas terhadap pendidikan non-formal PKBM. Sedangkan untuk desa dengan status berkembang, maju, dan mandiri menunjukkan adanya peningkatan akses terhadap pendidikan non-formal di desa meskipun masih terdapat banyak desa yang belum memiliki PKBM.

Pembagian Data

Sebelum melakukan pemodelan klasifikasi menggunakan algoritma XGBoost, data status Indeks Desa Membangun yang berjumlah 5.312 akan dibagi dengan proporsi 80% untuk data *training* dan 20% untuk data *testing* secara acak, sehingga diperoleh hasil sebagai berikut:

Tabel 5. Hasil Pembagian Data

Keterangan	Data Training	Data Testing	Total
Proporsi	80%	20%	100%
Jumlah	4.249	1.063	5.312

Berdasarkan Tabel 7 diketahui bahwa dari 4.249 data *training*, terdapat 107 data yang termasuk dalam kelas 0 (tertinggal), 2.632 data dalam kelas 1 (berkembang), 1.297 data dalam kelas 2 (maju), dan 213 data dalam kelas 3 (mandiri) sedangkan dari 1.063 data *testing*, terdapat 13 data yang termasuk dalam kelas 0 (tertinggal), 659 data dalam kelas 1 (berkembang), 334 data dalam kelas 2 (maju), dan 57 data dalam kelas 3 (mandiri).

Hasil Pemodelan XGBoost

Dalam pembuatan model dengan algoritma XGBoost, pemodelan dilakukan menggunakan data *training* dan dimulai dengan mendefinisikan parameter model XGBoost yang ditentukan, yaitu sebagai berikut:

Tabel 6. Parameter Model XGBoost

Parameter Model XGBoost	
<i>objective</i>	multi:softmax
<i>eval_metric</i>	mlogloss, merror
<i>num_class</i>	4
<i>tree_method</i>	exact

Berdasarkan Tabel 8 diketahui bahwa artikel ini akan menggunakan *objective function* dengan fungsi *softmax*. Evaluasi metrik menggunakan *mlogloss* karena *loss function* yang digunakan adalah *loss logistic* dan *merror* akan digunakan untuk pengecekan *overfitting*. Selain itu, *tree method* menggunakan *exact* yang berarti tidak ada pendekatan pembobotan dalam pembagian *node* pada pohon.

Setelah itu, dilakukan *hyperparameter tuning* menggunakan *grid search* agar mendapatkan nilai terbaik dari masing-masing parameter yang digunakan, dengan hasil sebagai berikut:

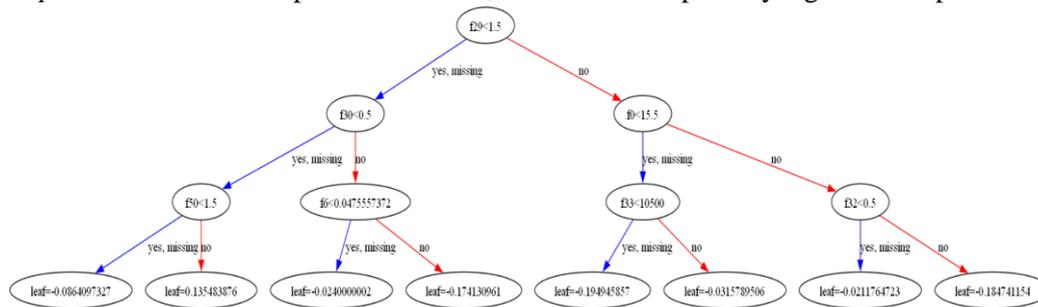
Tabel 7. Hasil *Hyperparameter Tuning Grid Search*

Parameter	Grid Search Values	Best Parameter
<i>n_estimators</i>	[25, 50, 100]	100
<i>eta (learning_rate)</i>	[0,05, 0,1, 0,3]	0,3
<i>max_depth</i>	[3, 4, 6]	3
<i>min_child_weight</i>	[0,5, 1, 2]	1
<i>gamma</i>	[0,1, 1, 2]	0,1
<i>reg_lambda</i>	[0,5, 1, 2]	0,5
<i>subsample</i>	[0,5, 0,8, 1]	0,8
<i>colsample_bylevel</i>	[0,5, 0,8, 1]	0,8

Tabel 9 merupakan hasil *hyperparameter tuning* yang diperoleh melalui proses “gridsearchCV” dengan melakukan pencarian secara menyeluruh pada parameter yang diuji. Artikel ini menggunakan *5-fold cross validation* untuk mengevaluasi kinerja model sebanyak lima kali perulangan dalam proses *grid search* pada setiap parameter. Nilai parameter terbaik yang diperoleh dari proses *grid search* kemudian digunakan dalam pembuatan model klasifikasi.

Berdasarkan hasil *hyperparameter tuning* pada Tabel 9, model klasifikasi XGBoost akan dibentuk 100 pohon dengan 3 percabangan pada setiap pohonnya. Model ini menggunakan pendekatan *One-vs-All (OvA)* untuk klasifikasi multikelas, yang berarti terdapat 4 model biner (satu untuk setiap kelas), masing-masing dengan 100 pohon sehingga terdapat $4 \times 100 = 400$ pohon yang dibentuk untuk membedakan satu kelas dengan kelas lainnya. Jumlah minimum berat yang dibutuhkan pada *child node* bernilai 1, *learning rate* sebesar 0,3, regularisasi pada pembobot (λ) sebesar 0,5, dan minimum *loss* yang dibutuhkan untuk membuat partisi *node*

pada pohon sebesar 0,1. Banyaknya bagian sampel yang dipilih dalam membangun pohon sebesar 0,8 dan rasio *subsample* kolom untuk setiap level sebesar 0,8. Berikut hasil pohon yang dibentuk pada iterasi pertama:



Gambar 3. Hasil Pohon Pertama dari Iterasi Pertama Pemodelan XGBoost

Evaluasi Model

Evaluasi model akan dijelaskan berdasarkan *confusion matrix* yang diperoleh dari hasil klasifikasi menggunakan data *testing*, sebagai berikut:

Tabel 8. *Confusion Matrix* dari Data *Testing* dengan Pemodelan XGBoost

		Nilai Prediksi			
		Tertinggal	Berkembang	Maju	Mandiri
Nilai Aktual	Tertinggal	5	8	0	0
	Berkembang	2	635	22	0
	Maju	0	66	261	7
	Mandiri	0	0	15	42

Dari Tabel 10 dilakukan perhitungan untuk mengetahui nilai *accuracy*, *precision*, *recall*, dan *f1-score*, dengan hasil sebagai berikut:

Tabel 9. Hasil Evaluasi Model

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
89%	84%	72%	76%

Berdasarkan Tabel 11 diketahui bahwa model klasifikasi XGBoost menunjukkan performa yang sangat baik dan dapat diimplementasikan. Beberapa indikator evaluasi yang mendukung kesimpulan ini adalah diperoleh *accuracy* sebesar 89% yang menggambarkan model sudah baik dalam mengklasifikasikan data. Nilai *precision* sebesar 84% menggambarkan tingkat akurasi antara data yang diminta dengan nilai prediksi model. Nilai *recall* sebesar 72% menggambarkan kemampuan sistem dalam mengambil informasi. Nilai *f1-score* sebesar 76% menggambarkan kemampuan model atau sistem memiliki keseimbangan yang baik antara *precision* dan *recall*.

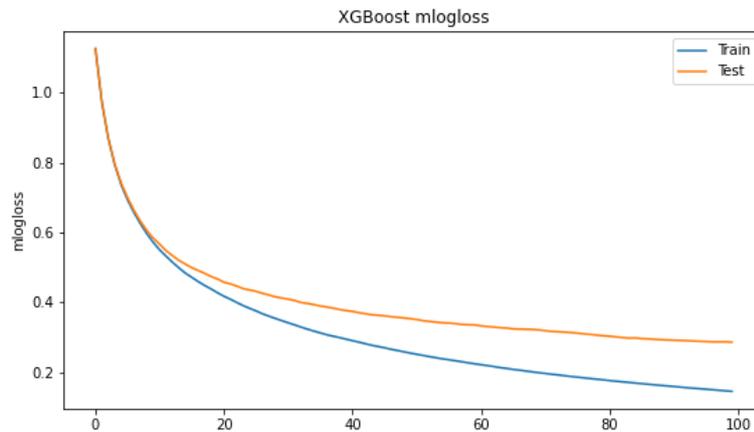
Evaluasi model juga dapat dilakukan dengan melihat apakah model yang telah dibuat terindikasi *overfitting* atau tidak. Pengecekan *overfitting* dilakukan untuk memastikan bahwa model tidak terlalu "terlatih" pada data *training* dan dapat menggeneralisasi pada data baru atau data *testing*. Berikut hasil *error* pada iterasi ke-100 dari data *training* dan data *testing*.

Tabel 10. Perbandingan *Error* pada Data *Training* dan Data *Testing*

<i>Error Data Training</i>	<i>Error Data Testing</i>
0,027536	0,112888

Berdasarkan Tabel 12 diketahui bahwa selisih *error* antara data *training* dan data *testing* sangat kecil, sekitar 0,085352 atau dapat dikatakan bahwa model yang dibuat tidak *overfitting* maupun *underfitting*.

Pengecekan *overfitting* atau *underfitting* juga dapat dilakukan menggunakan evaluasi metrik dengan *multilogloss* sebagai berikut:



Gambar 4. Grafik Evaluasi Metrik *Multilogloss*

Berdasarkan Gambar 4 diketahui bahwa dari 100 iterasi, hubungan antara nilai *mlogloss* dan jumlah iterasi menggambarkan hubungan linear negatif serta *mlogloss* pada data *training* dan data *testing* beriringan atau tidak berbeda secara signifikan sehingga tidak terjadi *overfitting* maupun *underfitting* pada model yang telah dibuat. Maka dapat dikatakan bahwa model memiliki keseimbangan yang baik antara kemampuan untuk menangkap pola yang kompleks dalam data *training* dan kemampuan untuk menggeneralisasi pada data baru.

D. Kesimpulan

Berdasarkan analisis yang dilakukan pada data status Indeks Desa Membangun di Provinsi Jawa Barat tahun 2020 menggunakan 52 variabel dengan data *training* sebanyak 4.249 dan data *testing* sebanyak 1.063, akan dibentuk model XGBoost dengan 100 pohon dan 3 percabangan pada setiap pohonnya. Jumlah minimum berat yang dibutuhkan pada *child node* bernilai 1, *learning rate* sebesar 0,3, regularisasi pada pembobot (λ) sebesar 0,5, dan *minimum loss* yang dibutuhkan untuk membuat partisi *node* pada pohon sebesar 0,1. Banyaknya bagian sampel yang dipilih dalam membangun pohon sebesar 0,8 dan rasio *subsample* kolom untuk setiap level sebesar 0,8. Dari model tersebut, diperoleh nilai *accuracy* sebesar 89%, *precision* sebesar 84%, *recall* sebesar 72%, dan *f1-score* sebesar 76%. Berdasarkan hasil tersebut, nilai *accuracy* yang diperoleh sudah tinggi sehingga dapat dikatakan bahwa model XGBoost yang dibangun sudah dapat mengklasifikasikan status Indeks Desa Membangun dengan baik dan dapat diimplementasikan.

Daftar Pustaka

- [1] E. Setyowati, "Tata Kelola Pemerintahan Desa pada Perbedaan Indeks Desa Membangun (IDM): Studi Tiga Desa di Kabupaten Malang," 2019.
- [2] A. S. Budiman and N. Hajarisman, "Analisis Mediasi Multipel Paralel Kausal Step pada Data Stunting menurut Kabupaten/Kota," *Jurnal Riset Statistika*, pp. 31–40, Jul. 2024, doi: 10.29313/jrs.v4i1.3860.
- [3] RPJMN, "Lampiran Peraturan Presiden Republik Indonesia Nomor 18 Tahun 2020 Tentang Rencana Pembangunan Jangka Menengah Nasional 2020-2024 Narasi Republik Indonesia," 2020.
- [4] Permendes, "Permendesa Nomor 2 Tahun 2016," 2016.
- [5] F. Xaverius, "Perkembangan Desa Berdasarkan Indeks Desa Membangun di Kabupaten Mempawah," vol. 7, no. 1, pp. 10–20, 2021.
- [6] T. S. Lestari and D. A. N. Sirodj, "Klasifikasi Penipuan Transaksi Kartu Kredit Menggunakan Metode Random Forest," *Jurnal Riset Statistika*, vol. 1, no. 2, pp. 160–167, Feb. 2022, doi: 10.29313/jrs.v1i2.525.
- [7] S. Bahri, "Implementasi Data Mining untuk Memprediksi Keterlambatan Jam Masuk Kerja Menggunakan Algoritma Klasifikasi," vol. 1, no. 1, pp. 11–20, 2020.
- [8] Desi Permatasari and Suliadi, "Algoritma Pemilihan Variabel untuk Klasifikasi dan Penerapannya pada Klasifikasi Desa-Kelurahan," *Jurnal Riset Statistika*, pp. 49–56, Jul. 2024, doi: 10.29313/jrs.v4i1.3876.

- [9] F. R. Noor and D. A. N. Sirodj, “Penerapan Support Vector Machine untuk Klasifikasi Opini Masyarakat Terhadap Isu Bullying,” *Jurnal Riset Statistika*, pp. 57–66, Jul. 2024, doi: 10.29313/jrs.v4i1.3877.
- [10] J. M. A. S. Dachi, “Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit,” *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam (JURRIMIPA)*, vol. 2, no. 2, 2023, doi: 10.55606/jurrimipa.v2i2.1336.
- [11] Tasya Azzahra, “Pemodelan Regresi Hurdle Negative Binomial pada Jumlah Kasus Difteri Jawa Barat 2020,” *Jurnal Riset Statistika*, pp. 125–130, Dec. 2023, doi: 10.29313/jrs.v3i2.3014.
- [12] P. D. Inas Azizah, “Penerapan Probabilistic Neural Network pada Klasifikasi Berat Bayi Baru Lahir,” *Jurnal Riset Statistika*, vol. 1, no. 2, pp. 152–159, Feb. 2022, doi: 10.29313/jrs.v1i2.524.