

OPTIMALISASI KLASIFIKASI INDEKS DESA MEMBANGUN MENGGUNAKAN METODE ENSEMBLE DAN ALGORITMA RANDOM FOREST

Ahmad Rifai¹, Sandy Eka Permana², Ryan Hamonangan³

^{1,3} Teknik Informatika, STMIK IKMI Cirebon

² Manajemen Informatika, STMIK IKMI Cirebon

Jalan Perjuangan 10 B Majasem Kota Cirebon, Indonesia

a.rifaai1408@gmail.com

ABSTRAK

Indeks Desa Membangun (IDM) adalah sistem klasifikasi yang digunakan untuk menentukan status pertumbuhan desa di Indonesia. Terdapat lima kategori status desa, yaitu desa mampu mandiri, desa telah berkembang, desa sedang dalam proses perkembangan, desa tertinggal, dan desa sangat tertinggal. IDM bertujuan membantu pembangunan desa dan memberikan saran kebijakan sesuai dengan status desa. Desa dengan status tertentu memerlukan intervensi kebijakan yang berbeda. Pertumbuhan desa terjadi jika status IDM meningkat dari tahun ke tahun, sedangkan jika status IDM menurun, maka desa tersebut mengalami penurunan. Penelitian ini menggunakan teknik *ensemble*, khususnya membangun banyak pohon keputusan pada setiap *bootstrap* data yang diambil dari data latih, mirip dengan metode *Random Forest*. Prediksi yang dihasilkan kemudian digabungkan untuk menghasilkan hasil akhir yang lebih akurat dan konsisten dalam klasifikasi desa berdasarkan IDM. Model *ensemble* ini dihasilkan secara *iteratif* dengan menyesuaikan bobot data yang salah diklasifikasikan oleh model sebelumnya sehingga akurasi klasifikasi desa pada setiap iterasi model menjadi lebih akurat dan dapat diandalkan. Hasil analisis data klasifikasi IDM menerapkan metode *ensemble optimize parameters* menunjukkan bahwa terdapat desa mandiri sebanyak 0,19%. Selain itu, empat desa memiliki indeks ketahanan sosial (IKS) tertinggi, enam desa memiliki indeks ketahanan ekonomi (IKE) tertinggi, dan tiga puluh desa memiliki indeks ketahanan lingkungan (IKL) tertinggi.

Kata kunci : Klasifikasi, Indeks Desa Membangun (IDM), *Random Forest*, *Ensemble Optimize Parameters*

1. PENDAHULUAN

Pertumbuhan desa bisa diukur menggunakan Indeks Desa Mandiri (IDM) dengan tujuan untuk mengevaluasi pengaruh jangka panjang pada kesejahteraan desa. Indeks Desa Membangun (IDM) digunakan sebagai alat untuk mengklasifikasikan status desa menjadi lima tingkatan, yaitu desa yang mandiri, desa yang maju, desa yang berkembang, desa yang tertinggal, dan desa yang sangat tertinggal [1]. Desa memiliki status dengan situasi serta keadaan tiap desa memerlukan pendekatan serta intervensi kebijakan yang berbeda. Desa tumbuh mempunyai karakteristik berbentuk pembangunan dan pengembangan segala infrastruktur perekonomian mulai dari penciptaan, retribusi, serta pemasaran dan berupaya buat bisa jadi lumbung pangan serta usaha perekonomian warga desa [2]. Pentingnya pembangunan di desa sangatlah besar mengingat mayoritas penduduk Indonesia tinggal di daerah pedesaan. Berdasarkan data Kementerian Desa, sebanyak 74.955 desa pada tahun 2022, desa mandiri mencapai 8,23% atau 6.238 desa, desa maju mencapai 27,01% atau 20.249 desa, desa berkembang mencapai 45,23% atau 33.902 desa, desa tertinggal mencapai 12,79% atau 9.584 desa, dan desa sangat tertinggal mencapai 6,65% atau 4.982 desa [3].

Undang-undang No. 6 Tahun 2014 telah disahkan oleh pemerintah dengan tujuan untuk meningkatkan kesejahteraan dan kualitas hidup penduduk desa, serta mengatasi masalah kemiskinan melalui pemenuhan kebutuhan dasar, pembangunan infrastruktur dan fasilitas, pengembangan ekonomi

lokal, serta pengelolaan sumber daya manusia dan lingkungan secara berkelanjutan [4]. Pemerintah mengeluarkan peraturan perundangan tersebut untuk mengendalikan desa, dalam pandangan baru mengenai pembangunan desa, peran utama diberikan kepada desa dalam upaya mencapai kesejahteraan warga desa melalui pemenuhan kebutuhan sosial dasar, pengembangan potensi ekonomi lokal, dan penggunaan daerah secara berkelanjutan. Dalam konteks ini, pembangunan desa menjadi suatu usaha untuk kemajuan desa itu sendiri [5].

Berdasarkan data IDM tahun 2022 di Provinsi Maluku Utara, sekitar 56% dari total 1.033 desa, yaitu sebanyak 602 desa, masih termasuk dalam kategori desa tertinggal dan sangat tertinggal. Sementara itu, hanya sekitar 8% atau 87 desa yang berada dalam kategori lainnya. Desa yang berhasil mencapai pertumbuhan sementara sisanya, yaitu sebanyak 976 desa, tidak berhasil mencapai pertumbuhan atau bahkan beberapa desa justru mengalami penurunan. Desa dikatakan mengalami pertumbuhan jika status IDM meningkat dari tahun ke tahun, sedangkan jika status IDM-nya tidak berubah maka desa tersebut dikatakan tidak mengalami pertumbuhan, dan sebaliknya jika status IDM-nya menurun maka desa tersebut dikatakan mengalami penurunan.

Desa memiliki faktor-faktor yang menghambat pertumbuhan tersebut antara lain (1) ketidakbergantian *mindset* atau mentalitas dari birokrasi dalam hal pelayanan; (2) keterbatasan sumber daya manusia; (3) rendahnya tingkat ekonomi warga; (4) kurangnya partisipasi warga; (5) keterbatasan sumber daya alam;

(6) posisi geografis wilayah; (7) keberlanjutan kebudayaan di wilayah tersebut. Perencanaan pembangunan di era otonomi saat ini yang menghalangi proses pertumbuhan desa adalah faktor-faktor tersebut di atas [6]. Sebuah desa tidak berhasil mencapai pertumbuhan disebabkan belum adanya sistem yang dapat mengidentifikasi *ketidaksiapan* desa sejak dini, sehingga pemerintah desa maupun pemerintah pusat sulit dalam melakukan upaya-upaya percepatan pemenuhan kebutuhan sarana dan prasarana yang diperlukan dalam mewujudkan pertumbuhan desa.

Peneliti mengembangkan sebuah model yang dapat memprediksi dengan teknik klasifikasi status IDM sebuah desa di tahun berikutnya berdasarkan data indikator-indikator kesiapan pada tahun berjalan. Pemerintah desa maupun pemerintah pusat dapat melakukan proses prediksi klasifikasi untuk mengetahui IDM pada tahun berikutnya. Jika diketahui hasil prediksi klasifikasi menunjukkan belum adanya peningkatan status IDM, maka pemerintah dapat melakukan upaya-upaya yang lebih optimal dengan harapan hasil prediksi klasifikasi tidak adanya pertumbuhan di tahun berikutnya dapat dicegah. Pemilihan model *random forest* pada *ensemble learning* dapat meningkatkan kinerja prediksi, mengurangi *overfitting*, meningkatkan stabilitas model, dan membantu dalam pemilihan fitur yang penting dalam data. Hal ini dikarenakan algoritma *random forest* membantu dalam memilih fitur yang paling penting dalam data. Pemilihan model ini dilakukan dengan menghitung seberapa sering fitur muncul dalam pemisahan pohon keputusan. Selain itu, *random forest* lebih stabil dibandingkan dengan model tunggal karena terdiri dari beberapa pohon keputusan (*decision trees*). Pemilihan model ini membuatnya lebih tahan terhadap gangguan atau variasi pada data. Teknik *ensemble* ini merupakan model yang dihasilkan secara iteratif dan setiap model ditambahkan dengan menyesuaikan bobot data yang salah diklasifikasikan oleh model sebelumnya agar meningkatkan akurasi klasifikasi desa pada setiap iterasi model dengan hasil klasifikasi desa berdasarkan IDM menjadi lebih akurat dan reliable. Dengan demikian, hal ini dapat memberikan pemahaman yang lebih komprehensif mengenai perkembangan desa.

Penelitian ini dapat membantu pemerintah dalam merencanakan kebijakan pembangunan desa yang lebih tepat sasaran dan berkelanjutan untuk meningkatkan kesejahteraan penduduk desa dan mengurangi kemiskinan di Indonesia. Penelitian ini menggunakan *dataset* indeks desa membangun yang diperoleh dari [7], [8], dan dengan latar belakang masalah tersebut di atas akan dijadikan sebagai landasan dalam pemilihan judul tesis dengan tema “Klasifikasi Indeks Desa Membangun dengan Menerapkan Metode *Ensemble Optimize Parameters* Menggunakan Algoritma *Random Forest*” studi kasus di desa membangun di Maluku Utara.

2. TINJAUAN PUSTAKA

2.1 Indeks Desa Membangun (IDM)

Indeks Desa Membangun (IDM) merupakan Indeks Komposit yang dibangun bersumber pada tiga indeks, yaitu Indeks Ketahanan Sosial, Indeks Ketahanan Ekonomi dan Indeks Ketahanan Ekologi/Lingkungan. Indeks Desa Membangun memotret pertumbuhan desa berlandaskan implementasi undang-undang desa dengan dukungan dana desa dan juga pendamping desa. Indeks Desa Membangun memusatkan ketepatan intervensi dalam kebijakan dengan korelasi intervensi pembangunan dari Pemerintah sesuai dengan partisipasi warga yang berkorelasi dengan karakteristik wilayah Desa yaitu tipologi serta modal sosial [5].

2.2 Data Mining

Data mining yaitu proses yang memakai metode statistik, matematika, kecerdasan buatan, serta *machine learning* untuk mengekstraksi dan mengidentifikasi data yang bernilai serta pengetahuan yang terikat dari berbagai *database* besar, sehingga *data mining* menggambarkan pengetahuan yang tersembunyi di dalam *database* yang di proses untuk mendapatkan pola serta metode statistik matematika, kecerdasan buatan, serta *machine learning* untuk mengekstraksi serta mengidentifikasi data pengetahuan dari *database* tersebut klasifikasi [10].

2.3 Tahapan Data Mining (CRISP-DM)

CRISP-DM telah lama diakui secara luas menjadi sebuah proses terstruktur atau kerangka kerja yang digunakan dalam memandu proyek data mining dan *machine learning* metode CRISP-DM bertujuan untuk menemukan pola yang menarik dan memiliki makna pada data yang digunakan [11]. CRISP-DM memiliki tahapan dan kerangka kerja yang terstruktur sehingga pengguna metode ini akan lebih terarah dan mengetahui langkah yang harus dikerjakan dalam penelitian [12].

2.4 Metode Ensemble

Metode *Ensemble* merupakan algoritma dalam *machine learning* dimana algoritma ini selaku pencarian pemecahan prediksi terbaik dibanding dengan algoritma yang lain sebab tata cara *ensemble* ini memakai sebagian algoritma *machine learning* buat pencapaian pemecahan prediksi yang lebih baik daripada algoritma yang dapat diperoleh dari salah satu algoritma *machine learning*. *Ensemble machine learning* cuma terdiri dari seperangkat model alternatif yang bersifat terbatas, tetapi umumnya membolehkan buat jadi lebih banyak lagi struktur fleksibel yang terdapat di antara alternatif model itu sendiri. Evaluasi prediksi dari *ensemble* biasanya memerlukan banyak komputasi daripada evaluasi prediksi model tunggal (*single model*), jadi *ensemble* ini memungkinkan untuk mengimbangi *poor learning algorithms* oleh performansi lebih dari komputasi itu [10].

2.5 Random Forest

Random Forest merupakan sebuah metode *ensemble*. Metode *ensemble* merupakan cara untuk

meningkatkan akurasi metode klasifikasi dengan cara mengombinasikan metode klasifikasi. *Random Forest* diawali dengan teknik dasar *data mining* yaitu *decision tree*. Pada *decision tree input* dimasukan pada bagian atas (*root*) kemudian turun kebagian bawah (*leaf*) untuk menentukan data tersebut termasuk kelas apa. Dengan kata lain *Random Forest* terdiri dari sekumpulan *decision tree* (pohon keputusan), di mana kumpulan *decision tree* tersebut digunakan untuk mengklasifikasi data ke suatu kelas [10].

2.6 Penelitian Terdahulu

Penelitian Aram Palilu dkk. mengenai analisis indeks desa membangun di kampung Klayili distrik Klaliyi Kabupaten Sorong pada tahun 2021 memiliki tujuan untuk menunjukkan bahwa indeks jaminan sosial di Desa Klayili masing-masing sebesar 0,485 untuk ketiga variabel tersebut; 0,543 untuk indeks ketahanan lingkungan dan 0,490 untuk indeks ketahanan ekonomi. Hasilnya, hasil indeks IDM Desa Klayili untuk ketiga variabel tersebut rata-rata mencapai 0,506. Jika dikonsultasikan, hasil akhir ini memiliki rentang >0,491 hingga 0,599, yang menunjukkan bahwa Desa Klayili tergolong desa tertinggal [13]. Penelitian lainnya dari Nasiya Alifah Utami dkk. pada tahun 2022 memanfaatkan data yang dikumpulkan dari situs web Kemendesa dan publikasi IPD oleh BPS tahun 2018. Penelitian tersebut pendekatan *naïve bayes* digunakan untuk mengklasifikasikan status indeks pembangunan desa. Berdasarkan hasil penelitian, metode pembelajaran terbaik adalah *MST (Maximum Spanning Tree)* dengan diskritisasi data menggunakan algoritma K-means, yang mencapai akurasi sebesar 90,69%. [14].

Penelitian Muhammadiyah mengenai indeks desa membangun dan kesejahteraan petani di Desa Kalimas Kecamatan Sungai Kakap Kabupaten Kubu Raya tahun 2021 dengan mengadopsi pendekatan deskriptif eksploratif dan menghasilkan temuan bahwa Desa Kalimas dapat diklasifikasikan sebagai desa maju dengan skor IDM sebesar 0,7493. Lebih dari 44% masyarakat Desa Kalimas percaya bahwa desa mereka berada dalam status desa maju. Skor IDM dan status desa ini dipengaruhi oleh tiga indeks ketahanan, yaitu Indeks Ketahanan Sosial, Indeks Ketahanan Ekonomi, dan Indeks Ketahanan Lingkungan [15]. Berdasarkan penelitian Mutiara Cindy Nur Fitria dkk. pada tahun 2021 dengan menggunakan 156 data dengan akurasi 82,35 persen dan kualitas model yang baik untuk mengklasifikasikan status desa menggunakan algoritma C5.0 [16]. Serta penelitian yang dilakukan Devi Triady Bachruddin, dkk. mengenai analisis pembangunan desa berdasarkan capaian indeks desa membangun di Kabupaten Serang pada tahun 2020. Penelitian tersebut dilakukan dalam bentuk deskriptif berdasarkan data IDM yang diterbitkan oleh Kementerian Desa pada tahun 2015 dan 2018. Temuan penelitian menunjukkan bahwa pada tahun 2018, tidak ada desa mandiri di Kabupaten Serang berdasarkan nilai IDM, sedangkan pada tahun 2015 terdapat tiga desa mandiri berdasarkan data IDM. Jumlah desa maju

mengalami penurunan dari 20 pada tahun 2015 menjadi 11 pada tahun 2018. Jumlah desa di Kabupaten Serang meningkat dari 98 menjadi 138, menunjukkan peningkatan status desa secara keseluruhan. Pada tahun 2018, terdapat 166 desa tertinggal dan 11 desa sangat tertinggal, mengalami penurunan dari 183 pada tahun 2017. Skor IKL dan IKE Kabupaten Serang lebih tinggi daripada rata-rata Provinsi Banten, tetapi skor IKL mengalami penurunan dari 0,6288 (2015) menjadi 0,6163 (2018). Skor IDM Kabupaten Serang lebih rendah dibandingkan dengan skor Provinsi Banten secara keseluruhan. Meskipun memiliki skor IDM sebesar 0,5930, Kabupaten Serang masih diklasifikasikan sebagai desa tertinggal. Untuk meningkatkan tekanan pada sektor yang memiliki potensi peningkatan, seperti IKL dan IKE, pemerintah desa perlu melaksanakan prakarsa perencanaan pembangunan. Oleh karena itu, Pemerintah Kabupaten Serang perlu melakukan intervensi kebijakan [17].

3. METODE PENELITIAN

3.1 Jenis Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder diperoleh dari sumber-sumber yang sudah ada dan dikumpulkan oleh pihak lain. Dalam konteks penelitian ini, data sekunder mencakup informasi tentang Indeks Desa Membangun (IDM) yang diambil dari berbagai sumber resmi, seperti *website* Kementerian Desa.

3.2 Sumber Data

Sumber data dalam penelitian ini adalah data yang diperoleh dari dua sumber utama:

- a. *Website* Kementerian Desa: Informasi terkait status desa, seperti Indeks Desa Membangun (IDM) dan berbagai indikator terkait, diambil dari publikasi yang tersedia di <https://idm.kemendesa.go.id/view/detil/3/publikasi>.
- b. Sistem Informasi Desa (SID): Data tambahan mengenai dana desa dan bantuan langsung tunai (BLT) diperoleh dari <https://sid.kemendesa.go.id/village-fund>.

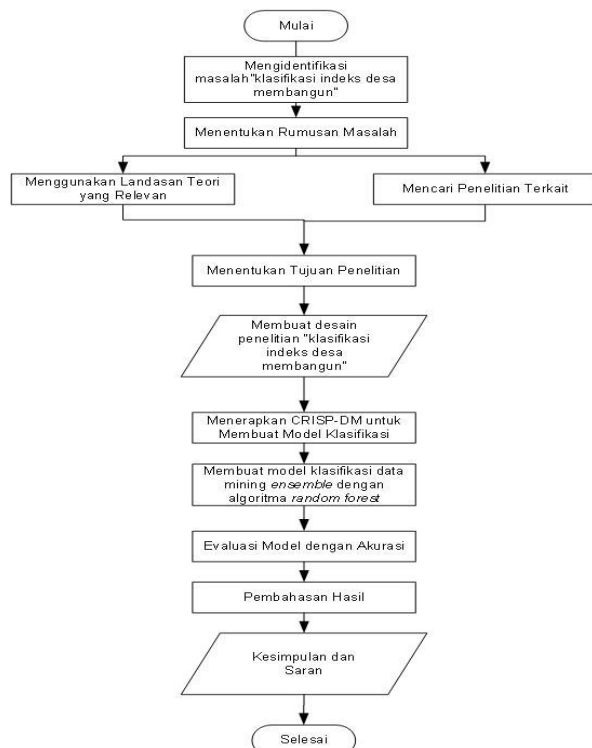
3.3 Teknik Pengambilan Data

Teknik pengambilan data dalam penelitian ini melibatkan pengunduhan dan ekstraksi data dari sumber-sumber resmi yang telah disebutkan. Proses pengambilan data dilakukan dengan langkah-langkah berikut:

- a. Identifikasi Sumber Data: Menentukan dan mengakses situs web resmi Kementerian Desa dan Sistem Informasi Desa.
- b. Pengunduhan Data: Mengunduh *dataset* yang diperlukan, termasuk data IDM, dana desa, dan BLT.
- c. Pembersihan Data: Melakukan pembersihan data untuk memastikan bahwa data yang digunakan dalam analisis bebas dari kesalahan atau data yang tidak relevan.

3.4 Metode Penelitian

Penelitian ini menggunakan metode *ensemble* dengan algoritma *Random Forest* untuk klasifikasi status desa. Berikut adalah tahapan metode penelitian yang digunakan:



Gambar 1. Alur Penelitian

Penelitian ini dimulai dengan mengidentifikasi masalah utama, yaitu "klasifikasi indeks desa membangun." Langkah pertama ini melibatkan pengenalan masalah yang spesifik dan menentukan rumusan masalah yang akan menjadi fokus utama penelitian. Setelah masalah diidentifikasi, peneliti kemudian menggunakan landasan teori yang relevan untuk membangun dasar teoritis yang kuat dan mencari penelitian terkait. Penelitian terkait ini membantu dalam menemukan celah atau gap yang ada serta memperkuat basis teoritis penelitian.

Selanjutnya, peneliti menentukan tujuan penelitian yang ingin dicapai. Tujuan ini harus spesifik dan terukur, mencakup aspek-aspek yang akan dijelaskan dan dibuktikan dalam penelitian. Dengan tujuan yang jelas, peneliti kemudian membuat desain penelitian yang rinci, termasuk metodologi yang akan digunakan, jenis data yang dibutuhkan, dan teknik analisis data yang akan diterapkan.

Tahap berikutnya adalah menerapkan metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) untuk membimbing proses data mining. Langkah-langkah dalam CRISP-DM meliputi pemahaman tujuan bisnis dan konteks penelitian, pengumpulan dan pemahaman data yang ada, penyiapan data untuk analisis, pembangunan model klasifikasi menggunakan teknik data mining, evaluasi model yang telah dibangun, dan implementasi model yang telah divalidasi.

Dalam pembangunan model klasifikasi, peneliti menggunakan algoritma *Random Forest* dengan pendekatan *ensemble*. Algoritma ini dipilih karena kemampuannya dalam meningkatkan akurasi prediksi dan mengurangi *overfitting*. Setelah model dibangun, langkah selanjutnya adalah mengevaluasi model untuk menentukan akurasi. Evaluasi ini mencakup analisis *metrics* seperti *confusion matrix*, sensitivitas, spesifisitas, dan AUC (*Area Under Curve*).

Hasil evaluasi kemudian dibahas secara mendalam, menganalisis kekuatan dan kelemahan model yang telah dibangun, serta membandingkannya dengan penelitian sebelumnya. Diskusi ini membantu dalam memahami efektivitas model dan bagaimana model tersebut dapat diterapkan dalam konteks yang lebih luas.

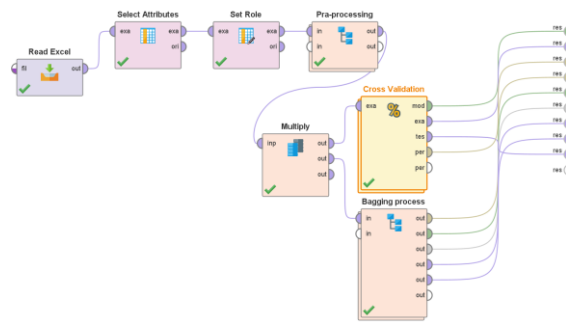
Pada akhir penelitian, peneliti menyusun kesimpulan dari penelitian yang telah dilakukan dan memberikan saran untuk penelitian selanjutnya. Kesimpulan ini mencakup temuan utama dari penelitian, serta implikasi praktis dan teoritis dari hasil penelitian. Dengan menyelesaikan seluruh langkah ini, penelitian dapat diselesaikan dengan baik, memastikan bahwa setiap langkah dalam proses penelitian dilakukan secara sistematis dan terstruktur, sehingga hasil yang diperoleh valid dan dapat dipertanggungjawabkan.

4. HASIL DAN PEMBAHASAN

Bantuan langsung tunai (BLT) dari hasil temuan *dataset* tersebut dapat dikemukakan bahwa jumlah desa yang memiliki BLT yang terendah sebanyak 39 desa, sedangkan jumlah desa yang memiliki BLT tertinggi sebanyak 3 desa. Berdasarkan data statistik penyaluran dana yang tercantum pada tabel 4.11 diatas dapat diperoleh informasi bahwa atribut sebaran penyaluran dana itu adalah rendah/kecil atau tidak adanya kesenjangan yang cukup besar dari penyaluran dana, hal ini dikarenakan nilai deviasi lebih kecil dari nilai rata-rata (*mean*) sehingga, atribut penyaluran dana ini dapat dikatakan Baik. Penyaluran dana berdasarkan *dataset* dari desa wilayah maluku utara tersebut diperoleh informasi bahwa jumlah desa yang memiliki penyaluran dana rata-rata sebanyak 355 desa.

4.1 Implementasi Model

Penelitian tentang klasifikasi indeks desa membangun dengan menerapkan metode *ensemble optimize parameters* menggunakan algoritma *random forest* ini menggunakan PC/Laptop Lenovo dengan RAM 8GB dan *hardisk dirve* SSD 512 GB dengan *system* operasi Windows 11 dan menggunakan *tools* aplikasi Rapidminer Studio versi 9.10. Adapun model proses dari penelitian ini menggunakan *tools* Rapidminer dapat dilihat pada gambar 2 :

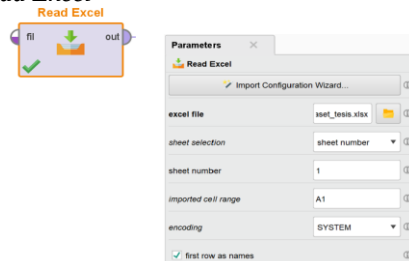


Gambar 2 Model Klasifikasi Desa Membangun

Dengan Menerapkan Metode *Ensemble Optimize Parameters* Menggunakan Algoritma *Random Forest*

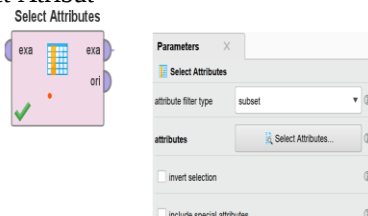
Implementasi dari metode ini dimulai dengan tahap membaca data menggunakan operator *Read excel*, memilih atribut dengan menggunakan operator *Select Atribut*, menetapkan atribut target dengan operator *Set Role*, melakukan *pra-processing*/pembersihan data, membagi proses menjadi 2 menggunakan operator *Multiply*, melakukan *K-Fold Cross Validation*, menerapkan proses *bagging* dan menerapkan algoritma *Random Forest* untuk melakukan evaluasi klasifikasi desa membangun dengan menerapkan metode *ensemble optimize parameters*. Proses klasifikasi desa membangun dengan menerapkan metode *ensemble optimize parameters* memiliki beberapa tahapan pada penelitian ini yang dapat diuraikan sebagai berikut :

a. *Read Excel*

Gambar 3. *Read Excel*

Read excel merupakan operator untuk membaca *dataset* yang tersimpan pada *file excel*, dengan *setting* parameter *excel file* = d:\dataset_tesis.xlsx, *sheet selection* = *sheet number*, *sheet number* = 1.

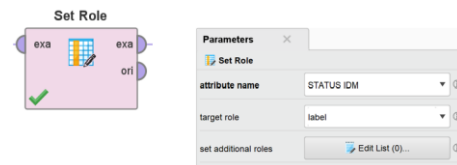
b. *Select Atribut*

Gambar 4 *Select Atribut*

Select Atribut merupakan operator untuk memilih atribut yang akan dijadikan *variable/atribut predictor*, dengan mengatur parameter *attribute filter type*=

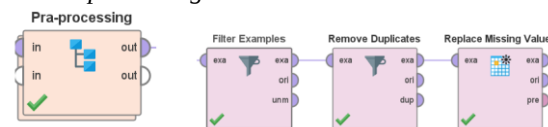
subset dan atribut yang dipilih terdiri atas *IKE, IKS, IKL, NAMA DESA* dan *STATUS IDM*.

c. *Set Role*

Gambar 4 *Set Role*

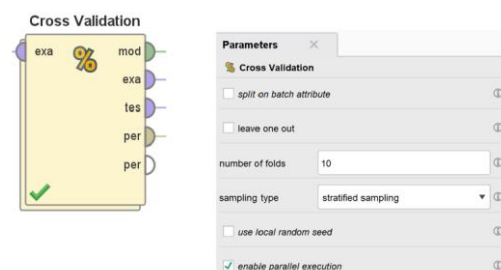
Set Role merupakan operator untuk merubah atribut menjadi *variable target*, dengan mengatur parameter *attribute name*= *STATUS IDM* dan *target role*=*label*.

d. *Pra-processing*

Gambar 5 *Pre-processing*

Pra-processing merupakan operator untuk sub-proses yang memuat operator lain, seperti operator *Filter Example*, operator *Remove Duplicate* dan Operator *Replace Missing Value*. Operator *Filter Example* untuk memilih *dataset* berdasarkan filter class, *Remove duplicate* untuk menghilangkan data yang sama atau duplikat agar tidak terjadi *redundancy*, *replace missing value* untuk mengisi data yang missing dengan nilai tertentu dalam hal ini *default* adalah *average* (nilai rata-rata).

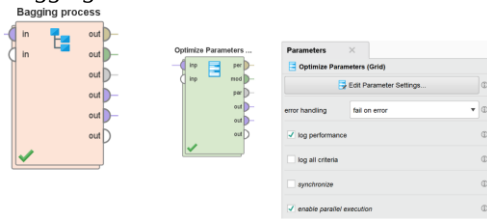
e. *Cross Validation*

Gambar 6 *Cross Validation*

Cross Validation adalah operator yang digunakan untuk membagi data menjadi data latihan (*training*) dan data uji (*testing*) dengan tujuan menghasilkan akurasi yang baik. Dalam *Cross Validation*, data latihan ditentukan sebesar 90% dan data uji sebesar 10%. *K-Fold Cross Validation* dilakukan dengan mengulang proses sebanyak *k*-kali untuk membagi himpunan sampel menjadi *k*-subset yang saling lepas. Parameter yang digunakan adalah *number of folds*=10 dan jenis pengambilan sampel menggunakan *stratified sampling*. *Stratified* sampling merupakan metode pengambilan sampel secara acak dengan mempertahankan proporsi kelas yang sama di setiap subset. Kelebihan dari *Stratified K-Fold Cross Validation* adalah mencegah terjadinya *overfitting* pada data latihan. *Overfitting* adalah kondisi di mana

model memiliki prediksi yang sangat baik pada data latihan, namun kinerjanya buruk pada data uji.

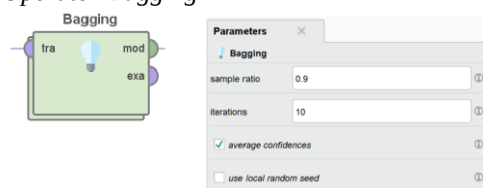
f. Bagging Process



Gambar 7 Ensemble dengan Teknik Bagging

Operator *Bagging process* merupakan sub operator sebagai ensemble yang memuat operator *optimize parameter* dengan mengatur parameter *random forest criterion* yang terdiri atas *gain ratio*, *information gain* dan *gini index*.

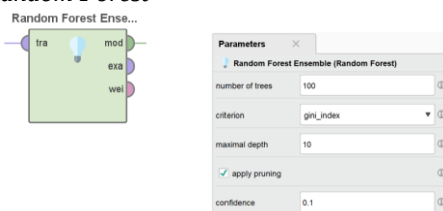
g. Operator Bagging



Gambar 8 Operator Bagging

Bootstrap aggregating (bagging) adalah *meta-algoritma ensemble* pembelajaran mesin untuk meningkatkan model klasifikasi dalam hal stabilitas dan akurasi klasifikasi untuk mengurangi variasi dan membantu menghindari *overfitting*. Operator *bagging* biasanya diterapkan pada model pohon keputusan, namun dapat digunakan dengan semua jenis model, mengatur parameter yang diterapkan pada penelitian ini secara default *sample ratio* = 0.9 dan *iterations* = 10.

h. Random Forest



Gambar 9 Operator Random Forest

Operator random forest merupakan operator dari algoritma random forest dengan parameter *number of trees*=100, *criterion*=*gini_index*, *maximal_dept*=10, *confidence*=0.1.

4.2 Evaluasi Tingkat Akurasi

Evaluasi pengujian dilakukan dengan penerapan 2 (dua) algoritma yaitu algoritma *Random Forest* dan algoritma *Random Forest* menggunakan *ensemble* dengan teknik *bagging* menggunakan *Cross Validation* dengan *nFold* sebesar 10 serta dengan teknik *stratified sampling*, yaitu sampling bertingkat membangun himpunan bagian acak. Evaluasi pengujian untuk memastikan bahwa distribusi kelas (didefinisikan oleh atribut label atau target) di subset *dataset* sama dengan di seluruh *dataset*. Klasifikasi

binominal salah satu contoh kasusnya, *stratified sampling* membangun subset acak sehingga setiap subset berisi kira-kira proporsi yang sama dari dua nilai label tersebut. Eksperimen yang dilakukan dalam penelitian ini diperoleh *confusion matrix* untuk penerapan algoritma *Random Forest* diperoleh nilai akurasi sebesar 92,67% seperti terlihat pada gambar 4.16 :

accuracy: 92.67% +/- 3.20% (micro average: 92.66%)

	true BERKEMBA...	true TERTINGGAL	true SANGAT TE...	true MAJU	true MANDIRI	class precision
pred. BERKEMB...	373	23	0	9	0	92.10%
pred. TERTINGG...	21	497	18	0	0	92.72%
pred. SANGAT T...	0	4	60	0	0	93.75%
pred. MAJU	1	0	0	55	2	94.83%
pred. MANDIRI	0	0	0	0	0	0.00%
class recall	94.43%	94.85%	76.92%	85.94%	0.00%	

Gambar 7 Confusion Matrix dari Algoritma *Random Forest*

Eksperimen yang dilakukan dalam penelitian ini menggunakan algoritma *Random Forest* menggunakan *ensemble optimize parameter* dengan teknik *bagging* diperoleh *confusion matrix* dengan nilai akurasi sebesar 93,88% seperti terlihat pada gambar 7 :

accuracy: 93.88% +/- 1.74% (micro average: 93.89%)

	true BERKEMBA...	true TERTINGGAL	true SANGAT TE...	true MAJU	true MANDIRI	class precision
pred. BERKEMB...	380	20	1	8	0	92.91%
pred. TERTINGG...	14	501	16	0	0	94.35%
pred. SANGAT T...	0	3	61	0	0	95.31%
pred. MAJU	1	0	0	56	2	94.92%
pred. MANDIRI	0	0	0	0	0	0.00%
class recall	95.20%	95.61%	78.21%	87.50%	0.00%	

Gambar 7 Confusion Matrix dari Algoritma *Random Forest* Dengan Metode *Ensemble* Dengan Teknik *Bagging*

Eksperimen pada penelitian ini menggunakan algoritma *Random Forest* menggunakan *ensemble optimize parameter* dengan teknik *bagging* tersebut dihasilkan akurasi berdasarkan 3 (tiga) kriteria dalam *optimize parameter* yang diterapkan seperti terlihat pada tabel 1 :

Tabel 1 Ensemble Optimize Parameter RF Criteria

No	RF Crireria	Akurasi
1	<i>gain rasio</i>	82,60%
2	<i>information gain</i>	93,00%
3	<i>gini index</i>	93,90%

Hasil eksperimen tersebut dapat dikemukakan bahwa *RF Criteria* yang menghasilkan nilai akurasi tertinggi 93,9% pada penerapan *optimize parameter* yaitu menggunakan kriteria *gini index*, sedangkan penerapan kriteria *gain ratio* menghasilkan nilai akurasi terendah 82,60%. Visualisasi nilai akurasi dari hasil eksperimen *optimize parameter* dalam penelitian ini seperti terlihat pada tabel 2

Dalam penelitian ini juga dilakukan perhitungan nilai *AUC* berdasarkan *confusion matrix* yang diperoleh dari penerapan algoritma *Random Forest* dan perhitungan nilai *AUC* berdasarkan *confusion matrix* yang diperoleh dari penerapan algoritma *Random Forest* menggunakan *ensemble optimize*

parameter dengan teknik *bagging*. Temuan hasil penelitian seperti ditunjukkan pada tabel 2 dan gambar 7 menunjukkan bahwa nilai AUC dari penerapan algoritma *Random Forest* menggunakan *ensemble optimize parameter* dengan teknik *bagging* (86,11%) hasilnya lebih baik dari pada nilai AUC dari penerapan algoritma *Random Forest* saja (83,44%). Temuan ini menunjukkan bahwa penerapan menggunakan *ensemble optimize parameter* dengan teknik *bagging* dalam model tersebut berdasarkan teori AUC pada sub bab menghasilkan klasifikasi indeks desa membangun dengan menerapkan *ensemble optimize parameters* sebagai baik.

Tabel 2 Tabel Perbandingan Akurasi Dengan AUC

No	Model	Performance accuracy	AUC
1	Random Forest+Ensemble	93,88%	86,11%
2	Random Forest	92,67%	83,44%

4.3 Pembahasan

Penelitian yang dilakukan ini dibandingkan dengan penelitian terkait yang dilakukan oleh [1], [2], [3], [4] dan [5] merupakan penelitian yang membahas penerapan algoritma *Random Forest* dalam bidang *Landslide susceptibility*, dalam bidang Kesehatan, dalam bidang iklim, topografi, vegetasi, dan sosial ekonomi terhadap kejadian kebakaran hutan, dalam bidang budaya keselamatan, dalam bidang *Groundwater Potential*, namun dalam penelitian ini dilakukan penerapan algoritma *Random Forest* dalam bidang indeks desa membangun di wilayah Maluku Utara.

Hasil penelitian menunjukkan bahwa informasi tentang prediksi desa berkembang dan desa itu memang benar desa berkembang yaitu sebesar 380 desa, sedangkan prediksi desa berkembang dan ternyata desa itu bukan desa berkembang melainkan masuk ke dalam desa tertinggal sebanyak 20 desa. Pengukuran *performance* yang telah dilakukan berikut di bawah ini adalah tabel 3 perbandingan *performance* yang dihasilkan dari penerapan 2 buah metode yang berbeda terlihat bahwa penerapan *ensemble learning* pada metode klasifikasi menggunakan *Random Forest* menghasilkan akurasi yang lebih baik,.

Tabel 3 Pengukuran *Performance*

Nilai Pengukuran Performance		Random Forest	Random Forest + Ensemble	Peningkatan akurasi
Recall	Berkembang	94,43	96,20	1,77
	Tertinggal	94,85	95,61	0,76
	Sangat Tertinggal	76,92	78,21	1,29
	Maju	85,94	87,5	1,56
	Mandiri	0	0	0

Nilai Pengukuran Performance		Random Forest	Random Forest + Ensemble	Peningkatan akurasi
Precision	Berkembang	92,10	92,91	0,81
	Tertinggal	92,72	94,35	1,63
	Sangat Tertinggal	93,75	95,31	1,56
	Maju	94,83	94,92	0,09
	Mandiri	0	0	0

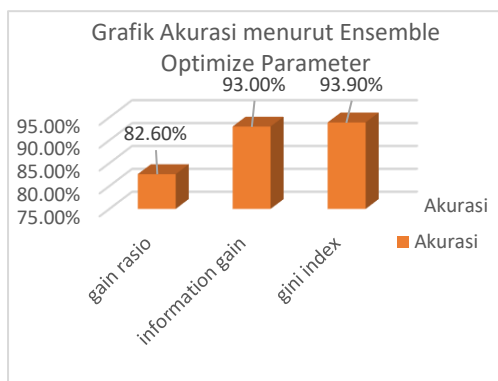
Pengukuran *performance* ini ditandai dengan adanya beberapa prediksi yang kurang tepat namun karena tingkat errornya cukup kecil 5% maka dapat disimpulkan model ini memiliki perform cukup baik sehingga dapat digunakan dalam memprediksi klasifikasi status IDM sebuah desa, walaupun demikian karena rata rata error yang relative kecil maka dapat dikatakan bahwa model klasifikasi yang dihasilkan masih dapat diterima. Dibandingkan dengan *performance* model klasifikasi yang hanya menerapkan algoritma *Random Forest* yaitu menghasilkan rata rata akurasi di nilai 92,67% maka terdapat peningkatan akurasi yang signifikan yaitu sebesar 1,21% . maka model klasifikasi menggunakan penerapan *ensemble learning* memberikan kontribusi peningkatan *performance* dan dapat dijadikan sebagai model klasifikasi dalam menentukan nilai IDM Desa.

Prediksi desa tertinggal dan desa itu memang benar tertinggal masih tergolong sangat tinggi yaitu sebesar 501 desa, sedangkan prediksi desa tertinggal dan ternyata desa itu bukan desa tertinggal melainkan masuk ke dalam desa berkembang sebanyak 14 desa dan masuk desa sangat tertinggal sebanyak 16 desa. Prediksi desa sangat tertinggal dan desa itu memang benar desa sangat tertinggal sebanyak 61 desa, sedangkan prediksi desa sangat tertinggal dan ternyata desa itu bukan desa sangat tertinggal melainkan masuk ke dalam desa tertinggal sebanyak 3 desa. Prediksi desa maju dan desa itu memang benar desa maju sebanyak 56 desa, sedangkan prediksi desa maju dan ternyata desa itu bukan desa maju melainkan masuk ke dalam desa berkembang sebanyak 1 desa dan masuk desa Mandiri sebanyak 2 desa.

Penelitian ini dilakukan klasifikasi indeks desa membangun dengan menerapkan *ensemble optimize parameters* dengan teknik *bagging*. Dimana penerapan teknik *bagging* ini dapat meningkatkan hasil akurasi yang semula 92,67% menjadi 93,88% sehingga meningkat sebesar 1,21%, dikarenakan teknik *bagging* mengambil sampel acak dengan penggantian dari dataset training untuk membangun kumpulan model yang berbeda. Setiap model dalam kumpulan model ini dibangun dari dataset yang berbeda, sehingga memiliki varian yang berbeda dalam kesalahan prediksinya. Prediksi pada data yang tidak terlihat selama training, kumpulan model akan menggabungkan hasil prediksi dari setiap model

individu dan menghasilkan prediksi rata-rata atau mayoritas.

Kumpulan model ini memiliki kecenderungan untuk mengurangi varian dalam kesalahan prediksi, yang dapat menyebabkan *overfitting* pada model individu, sehingga meningkatkan nilai akurasi. Teknik *bagging* ini digunakan dalam *ensemble learning*, di mana beberapa model berbeda digabungkan untuk meningkatkan kinerja keseluruhan. Dengan menggunakan teknik *bagging*, maka dapat memperkuat kinerja model dengan meningkatkan akurasi dan mengurangi varian dalam kesalahan prediksi, dengan menerapkan *ensemble optimize parameters* yang dibuat dalam penelitian ini.



Gambar 8 Grafik Akurasi Penerapan Ensemble Optimize Parameter

Dengan penerapan *ensemble optimize parameters* pada algoritma *random forest* dengan teknik *bagging* ini menjadi poin terpenting yang membedakan penelitian ini dengan lima penelitian terkait sebelumnya. Penerapan *ensemble optimize parameters* pada algoritma *random forest* dengan teknik *bagging* ini mampu meningkatkan nilai akurasi dan nilai AUC, sehingga berkontribusi dalam meningkatkan performance dari model tersebut.

5. KESIMPULAN DAN SARAN

Penggunaan teknik *bagging* dalam *ensemble learning* membantu mengurangi *overfitting* dan meningkatkan stabilitas serta akurasi model. Model *Random Forest* dengan *ensemble optimize parameters* mencapai akurasi sebesar 93.88%, lebih tinggi dibandingkan model *Random Forest* tanpa optimisasi yang mencapai akurasi sebesar 92.67%. Penerapan kriteria *Gini Index* pada *optimize parameters* menghasilkan nilai akurasi tertinggi sebesar 93.90%. Di sisi lain, kriteria *Gain Ratio* menghasilkan nilai akurasi terendah sebesar 82.60%. Teknik *bagging* yang digunakan dalam metode *ensemble* terbukti efektif dalam meningkatkan akurasi dan mengurangi varians kesalahan prediksi. Dengan menggabungkan hasil prediksi dari beberapa model individu, teknik ini menghasilkan prediksi yang lebih stabil dan akurat. Hasil penelitian ini dapat membantu pemerintah dalam merencanakan kebijakan pembangunan desa yang lebih tepat sasaran. Model klasifikasi yang dihasilkan

dapat digunakan untuk memprediksi status IDM desa pada tahun berikutnya, sehingga upaya percepatan pembangunan dapat dilakukan dengan lebih optimal. Metode *ensemble optimize parameters* dengan algoritma *Random Forest* merupakan pendekatan yang efektif dan dapat diandalkan untuk klasifikasi status IDM. Penelitian ini memberikan kontribusi penting dalam bidang klasifikasi data dan analisis pembangunan desa.

DAFTAR PUSTAKA

- [1] M. Cindy Nur Fitria, N. Nessyana Debataraja, S. Wira Rizki, P. Studi Statistika FMIPA Universitas Tanjungpura, dan J. H. Hadari Nawawi, "Classification of Village Status in Landak Regency Using C5.0 Algorithm," vol. 3, no. 1, 2022.
- [2] J. B. Akcaya, F. Xaverius, B. K. Mempawah, dan J. Chandramidi, "Perkembangan Desa Berdasarkan Indeks Desa Membangun Di Kabupaten Mempawah Village Progress By Village Developing Index In Mempawah Regency," vol. 7, no. 1, hlm. 10–20, 2021.
- [3] P. -Kabupaten -Kecamatan -Desa, "Peringkat Nilai Rata-Rata Indeks Desa Membangun Tahun 2022."
- [4] Republic Indonesia, "2-Dengan Persetujuan Bersama Dewan Perwakilan Rakyat Republik Indonesia Dan Presiden Republik Indonesia," 2014.
- [5] J. I. Sosial, D. Humaniora, E. Pembangunan, D. Berdasarkan, A. Nyoman Astika, dan N. S. Subawa, "Evaluasi Pembangunan Desa Berdasarkan Indeks Desa Membangun," Nyoman Sri Subawa, vol. 5, 2021, [Daring]. Tersedia pada: <http://jurnal.um-tapsel.ac.id/index.php/muquoddimah>
- [6] M. Warouw, R. Leonardus, R. Paulus, dan A. Pangemanan, "Kajian Faktor-Faktor Dalam Proses Pembangunan Desa Di Era Otonomi Daerah Di Kecamatan Sinonsayang Study of Factors in the process of Rural Development in the Era of Regional Autonomy in District Sinonsayang," 2015.
- [7] Kementerian Desa Pembangunan Daerah Tertinggal dan Transmigrasi, "Standar Operasional Prosedur Pengukuran Indeks Desa Membangun Tahun 2019," 2019.
- [8] G. Shmueli, P. C. Bruce, I. Yahav, N. R. Patel, dan Jr. Kenneth C. Lichtendahl, Data Mining For Business Analytics Concepts, Techniques, and Applications in R. John Wiley & Sons, Inc., 2018.
- [9] D. P. Utomo dan M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," Jurnal Media Informatika Budidarma, vol. 4, no. 2, hlm. 437, Apr 2020, doi: 10.30865/mib.v4i2.2080.

- [10] C. Q. X. Poh, C. U. Ubeynarayana, dan Y. M. Goh, "Safety leading indicators for construction sites: A machine learning approach," *Autom Constr*, vol. 93, hlm. 375–386, Sep 2018, doi: 10.1016/j.autcon.2018.03.022.
- [11] F. N. Dhewayani dkk., "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *Jurnal Teknologi dan Informasi*, doi: 10.34010/jati.v12i1.
- [12] S. Huber, H. Wiemer, D. Schneider, dan S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model," *Procedia CIRP*, vol. 79, hlm. 403–408, 2019, doi: 10.1016/j.procir.2019.02.106.
- [13] A. Palilu, J. E. Lopulalan, dan R. R. Pakpahan, "Analisis Indeks Desa Membangun Di Kampung Klayili Distrik Klayili Kabupaten Sorong," *Jurnal Jendela Ilmu*, vol. 2, no. 1, hlm. 13–18, 2021.
- [14] N. A. Utami dan A. W. Wijayanto, "Classification Of Village Development Index At Regency / Municipality Level Using Bayesian Network Approach With K-Means Discretization," hlm. 95–106.
- [15] M. Muhammadiyah, "Indeks Desa Membangun Dan Kesejahteraan Petani Di Desa Kalimas Kecamatan Sungai Kakap Kabupaten Kubu Raya," *Jurnal Pembangunan dan Pemerataan*, vol. 09, no. 02, hlm. 2–6, 2021.
- [16] S. W. Fitria, M. C. N., Debataraja, N. N., & Rizki, "Classification of Village Status in Landak Regency Using C5.0 Algorithm," *Tensor: Pure and Applied Mathematics Journal*, vol. 3, no. 1, hlm. 33–42, 2022.
- [17] D. T. Bachruddin dan B. A. Darma, "Analisis Pembangunan Desa Berdasarkan Capaian Indeks Desa Membangun di Kabupaten Serang," *Jurnal Analisis Kebijakan*, vol. 4, no. 2, hlm. 26–39, 2020.