



Identification of Determinants of Inclusive Economic Growth Using the Metaheuristic Whale Optimization Algorithm Approach

Firza Septian¹, Nina Dwi Putriani²

^{1,2}Prodi Informatika, Fakultas Ilmu Komputer, Universitas SereLO Lahat

¹firzaseptian09@gmail.com, ²ninadwiputriani0@gmail.com

ABSTRACT

Inclusive economic growth demands the identification of key factors that drive equitable improvements in regional welfare. However, the complex interrelationships among social, economic, and demographic variables make traditional approaches insufficient for handling High-dimensional data. This study introduces an innovative approach by combining the Whale Optimization Algorithm (WOA) for feature selection with a Random Forest Regressor model to predict Gross Regional Domestic Product (GRDP) per capita as the main indicator of regional prosperity. The dataset consists of 210 regional observations and 18 independent variables. Feature selection using WOA was guided by minimizing the mean squared error (MSE), resulting in the identification of the 8 most relevant features. The retrained Random Forest model on the selected features achieved a High prediction performance, with an R^2 value of 0.9938 and a Low RMSE. Furthermore, GRDP values were categorized into three regional welfare classes (Low, Medium, High), and the classification yielded 97.92% accuracy with High precision, recall, and F1-scores across all classes. These findings demonstrate that combining metaheuristic optimization and Machine Learning enables efficient and accurate identification of the key determinants of inclusive economic growth. The results provide valuable insights for formulating more targeted regional development policies. However, this approach still has limitations in terms of result interpretability, particularly in explicitly explaining the causal relationships among the selected variables. Although the implementation of WOA-RF is technically accurate, it lacks contextual representation of inclusivity dimensions, such as social engagement, service accessibility, and the participation of vulnerable communities in economic growth. This presents an interesting avenue for future research.

Keywords : Inclusive Economic Growth, GRDP, Whale Optimization Algorithm, Feature Selection, Random Forest, Machine Learning

ABSTRAK

Pertumbuhan ekonomi inklusif menuntut identifikasi faktor-faktor utama yang mendorong peningkatan kesejahteraan secara merata di berbagai wilayah. Namun, kompleksitas hubungan antar variabel sosial, ekonomi, dan demografis membuat pendekatan tradisional kurang efektif dalam menangani data berdimensi tinggi. Penelitian ini menawarkan pendekatan inovatif dengan menggabungkan algoritma *Whale Optimization Algorithm* (WOA) sebagai metode seleksi fitur dan model *Random Forest Regressor* untuk prediksi Produk Domestik Regional Bruto (PDRB) per kapita sebagai indikator utama kesejahteraan daerah. Dataset yang digunakan terdiri dari 210 observasi regional dan 18 variabel independen. Seleksi fitur dilakukan menggunakan WOA dengan fungsi objektif berupa *Mean squared error* (MSE), di mana 8 fitur terbaik dipilih berdasarkan minimasi error prediksi. Model *Random Forest* yang dilatih dengan fitur terpilih menghasilkan nilai R^2 sebesar 0,9938 dan RMSE yang rendah, menunjukkan performa prediksi yang sangat akurat. Selanjutnya, nilai PDRB diklasifikasikan ke dalam tiga kategori kesejahteraan (*Low, Medium, High*), dan hasil klasifikasi menunjukkan akurasi sebesar 97,92% dengan nilai presisi, recall, dan *F1-score* yang tinggi di seluruh kelas. Temuan ini menunjukkan bahwa pendekatan berbasis metaheuristik dan pembelajaran mesin mampu mengidentifikasi determinan utama pertumbuhan ekonomi inklusif secara efisien dan akurat. Hasil penelitian ini diharapkan dapat menjadi dasar dalam perumusan kebijakan pembangunan regional yang lebih tepat sasaran. Namun demikian, pendekatan ini masih memiliki keterbatasan pada interpretabilitas hasil, terutama dalam menjelaskan hubungan kausal antar variabel yang terpilih secara eksplisit. Meskipun implementasi WOA-RF akurat secara teknis, pendekatan ini masih kurang menunjukkan dimensi inklusivitas secara kontekstual, seperti keterlibatan sosial, aksesibilitas layanan, dan partisipasi masyarakat rentan dalam pertumbuhan ekonomi. Hal ini menarik diangkat sebagai penelitian selanjutnya.

Kata kunci : Pertumbuhan Ekonomi Inklusif, PDRB, *Whale Optimization Algorithm*, *Feature Selection*, *Random Forest*, *Machine Learning*.

1. PENDAHULUAN

Pertumbuhan ekonomi yang inklusif telah menjadi agenda strategis dalam pembangunan global, karena tidak hanya menekankan pada peningkatan Produk Domestik Bruto (PDB) [1], tetapi juga pada pemerataan hasil pembangunan dan keterlibatan seluruh kelompok masyarakat dalam proses ekonomi. Di tengah ketimpangan sosial dan ekonomi yang masih tinggi di berbagai wilayah, identifikasi determinan utama dari pertumbuhan ekonomi yang inklusif menjadi urgensi kebijakan yang signifikan [2]. Namun, kompleksitas hubungan antar variabel sosial, demografis, dan ekonomi menjadikan proses identifikasi tersebut menantang, terutama ketika pendekatan tradisional mengalami keterbatasan dalam mengelola data berdimensi tinggi dan nonlinieritas hubungan antar variabel [3], [4], [5].

Berbagai studi sebelumnya telah mengkaji faktor-faktor yang memengaruhi pertumbuhan ekonomi secara parsial, seperti indeks pembangunan manusia (HDI) [6], ketimpangan pendapatan (Gini Ratio) [7], [8], tingkat kemiskinan [9], [10], dan investasi [11], [12]. Namun demikian, pendekatan konvensional seperti regresi linier sering kali tidak cukup sensitif dalam menyaring kombinasi variabel terbaik dari dataset yang kompleks dan besar. Beberapa penelitian mutakhir mulai menerapkan metode pembelajaran mesin (*Machine Learning*), tetapi sebagian besar masih menggunakan pendekatan brute-force atau metode optimasi yang rentan terhadap solusi lokal dan membutuhkan tuning manual yang intensif.

Dalam konteks ini, penelitian ini menawarkan pendekatan baru dengan menerapkan algoritma metaheuristik *Whale Optimization Algorithm* (WOA) untuk seleksi fitur dalam rangka mengidentifikasi determinan utama pertumbuhan ekonomi inklusif. WOA, yang meniru perilaku sosial paus dalam strategi berburu mangsa, terbukti efisien dalam pencarian solusi global pada ruang solusi multidimensi yang kompleks [13], [14]. Pendekatan ini diintegrasikan dengan model *Random Forest* (RF) untuk mengevaluasi performa prediksi, menjadikannya kombinasi yang kuat antara seleksi variabel otomatis dan akurasi model non-linear berbasis ensemble. RF dipilih karena kemampuannya yang tinggi dalam menangani data dengan dimensi besar dan interaksi antar variabel yang kompleks [15], serta ketahanannya terhadap *overfitting*. Dengan menggabungkan WOA untuk optimalisasi subset fitur dan RF untuk klasifikasi atau regresi, pendekatan WOA-RF ini diharapkan dapat meningkatkan akurasi prediksi sekaligus mengurangi kompleksitas model, sehingga memperkuat validitas hasil dalam konteks analisis ekonomi makro. Pendekatan WOA-RF ini sudah terbukti cocok digunakan untuk seleksi banyak variabel seperti yang telah dilakukan oleh Liu, 2023 pada tulisannya [16].

Tujuan dari penelitian ini adalah untuk menjawab pertanyaan kunci: faktor-faktor apa saja yang secara signifikan memengaruhi pertumbuhan ekonomi inklusif, dan bagaimana algoritma optimasi berbasis WOA dapat meningkatkan ketepatan identifikasi faktor-faktor tersebut? Penelitian ini menutup celah (gap) dalam literatur yang masih terbatas dalam menggabungkan metode metaheuristik modern dengan model prediksi dalam konteks ekonomi inklusif [17], [18], [19], serta menawarkan *novelty* dalam bentuk penerapan WOA yang belum banyak digunakan dalam studi kebijakan pembangunan ekonomi di tingkat regional. Meski metode ini menunjukkan akurasi teknis yang tinggi, masih terdapat kekurangan dalam hal kemudahan interpretasi hasil, khususnya dalam mengungkap secara jelas hubungan kausal antara variabel-variabel yang dipilih. Selain itu, pendekatan WOA-RF ini belum sepenuhnya merepresentasikan aspek-aspek inklusivitas secara kontekstual, seperti peran serta sosial, kemudahan akses terhadap layanan, serta keterlibatan kelompok rentan dalam proses pertumbuhan ekonomi. Oleh karena itu, aspek-aspek tersebut menjadi peluang menarik untuk dikaji lebih lanjut dalam penelitian mendatang.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksploratif dengan tujuan untuk mengidentifikasi determinan utama pertumbuhan ekonomi inklusif di Indonesia berdasarkan variabel-variabel sosial, ekonomi, dan investasi. Data yang digunakan berasal dari dataset sekunder yang telah dibersihkan dan disusun dalam format lembar Excel, mencakup keseluruhan 24 variabel. Demi kesesuaian dengan penelitian penulis maka diseleksi secara manual sebanyak 18 variabel independen dan 1 variabel dependen. Variabel independen sebanyak 18 buah itu terdiri dari: *hdi*, *gini*, *poverty*, *gov_spending*, *gov_capex*, *unemployed_percent*, *formal_sector*, *health_complaint*, *decent_sanitation*, *clean_water*, *hs_graduation*, *life_expectancy*, *private_invest*, *foreign_invest*, *domestic_invest*, *agri*, *grdp_employed*, dan *population_thousand*. Kemudian variabel dependen yang digunakan ialah *grdp_capita* atau sering disebut sebagai Produk Domestik Regional Bruto (PDRB) per kapita. Terkait total jumlah observasi, terdapat 210 sampel regional. Kemudian dilakukan seleksi dengan menggunakan *Whale Optimization Algorithm* dan bertindak sebagai *feature selection*.

Untuk pemisahan data, dilakukan pembagian training set dan testing set dengan rasio 80:20 menggunakan metode `train_test_split` dari pustaka Scikit-learn. Selanjutnya, dilakukan proses seleksi fitur menggunakan algoritma *Whale Optimization Algorithm* (WOA), yaitu algoritma metaheuristik berbasis populasi yang meniru perilaku berburu mangsa paus bungkuk (humpback whale) melalui mekanisme spiral dan pengepungan [4]. Implementasi WOA dalam penelitian ini memanfaatkan pustaka `pyMetaheuristic`, dengan parameter sebagai berikut: jumlah populasi (`hunting_party`) sebanyak 10 agen, parameter spiral (`spiral_param`) diset sebesar 1, dimensi pencarian disesuaikan dengan jumlah fitur (18 dimensi), dan iterasi dijalankan selama 30 siklus pencarian.

Fungsi objektif (fitness function) yang digunakan dalam proses seleksi fitur adalah *mean squared error* (MSE) dari hasil prediksi model *Random Forest* Regressor yang dibangun dari subset variabel hasil seleksi. Semakin kecil nilai MSE, semakin baik kombinasi fitur yang dipilih. Hanya subset fitur yang menghasilkan error terkecil yang dipilih sebagai hasil akhir.

Setelah proses seleksi, model akhir dilatih ulang menggunakan *Random Forest* Regressor pada fitur terpilih, kemudian dievaluasi terhadap testing set menggunakan metrik performa regresi seperti R^2 (koefisien determinasi) dan *Root Mean squared error* (RMSE). Untuk mengevaluasi kemampuan klasifikasi dalam konteks kategorisasi kesejahteraan daerah, nilai PDRB per kapita dikonversi ke dalam tiga kelas: *Low*, *Medium*, dan *High*, dan dilakukan evaluasi klasifikasi dengan metrik akurasi, presisi, recall, dan *F1-score* serta ditampilkan dalam bentuk confusion matrix.

Seluruh proses pemodelan dan visualisasi dilakukan menggunakan Python (versi 3.x) dengan pustaka utama meliputi Pandas, NumPy, Scikit-learn, Seaborn, dan Matplotlib. Seluruh kode dan parameter model didesain agar dapat direplikasi secara langsung oleh peneliti lain dengan dataset yang serupa.

3. HASIL DAN PEMBAHASAN

3.1. Proses Optimasi dengan *Whale Optimization Algorithm* (WOA)

Proses optimasi dimulai dengan pemanggilan algoritma WOA selama 30 iterasi untuk menyeleksi kombinasi fitur terbaik yang meminimalkan nilai *Mean Squared Error* (MSE) dari model prediksi. Seperti ditunjukkan pada log iterasi, nilai fungsi objektif menurun secara bertahap dari awal sebesar 1.072.716 ke nilai akhir 962.529, yang mencerminkan peningkatan signifikan dalam performa model melalui seleksi fitur otomatis. WOA berhasil menyeleksi 8 variabel dari total 18 variabel awal yang ditunjukkan pada Tabel 1.

Tabel 1. Variabel Terpilih Hasil Seleksi *Whale Optimization Algorithm* (WOA)

No	Variabel	Definisi
1	gini	Indeks Gini – ukuran ketimpangan distribusi pendapatan; nilai antara 0 (pemerataan sempurna) hingga 1 (ketimpangan sempurna).
2	clean_water	Persentase rumah tangga yang memiliki akses terhadap air bersih yang layak.
3	hs_graduation	Persentase penduduk usia sekolah yang berhasil menyelesaikan pendidikan tingkat menengah atas (SMA/ sederajat).
4	foreign_invest	Total nilai investasi asing langsung (FDI) yang masuk ke suatu wilayah dalam satu tahun (dalam miliar rupiah).
5	domestic_invest	Total nilai investasi domestik dari sektor swasta atau BUMN di suatu wilayah dalam satu tahun (dalam miliar rupiah).
6	agri	Kontribusi sektor pertanian terhadap PDRB wilayah (dalam persen terhadap total PDRB).
7	grdp_employed	Jumlah penduduk yang bekerja (employed) dalam kaitannya dengan kontribusi terhadap PDRB wilayah (dalam satuan orang).
8	population_thousand	Jumlah total penduduk wilayah dalam ribuan orang.

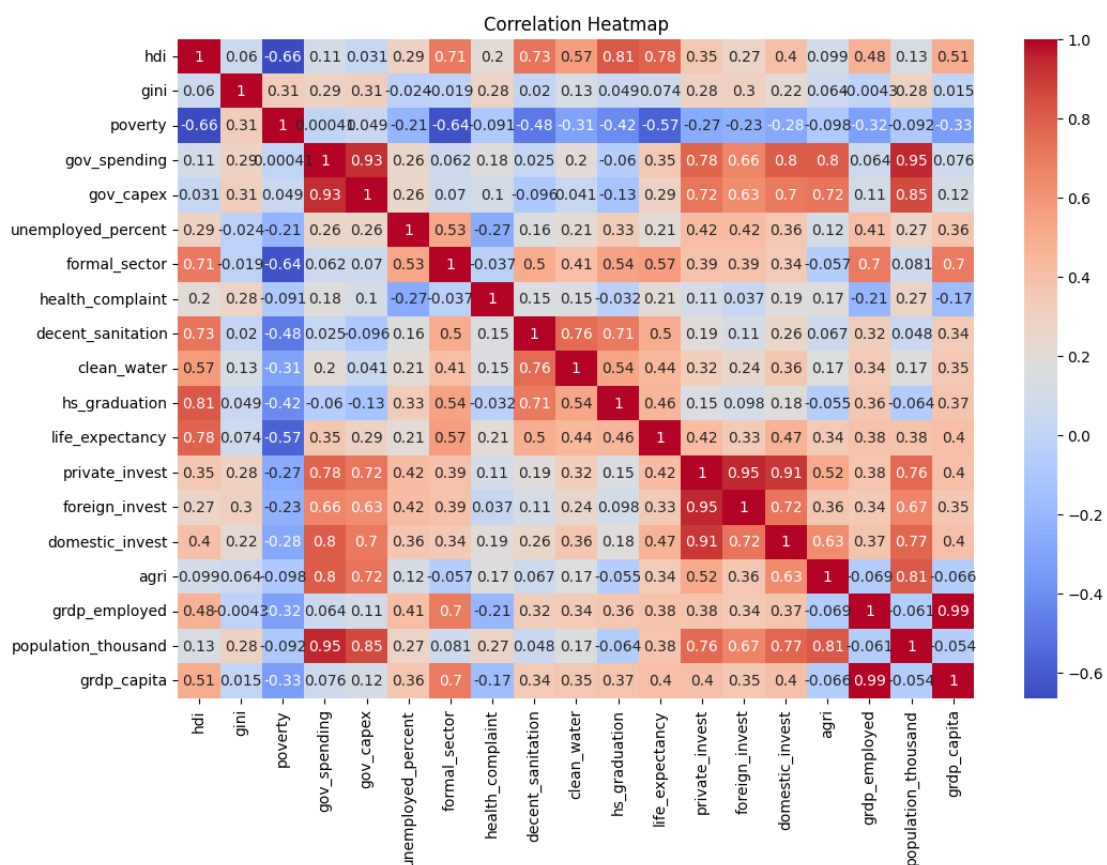
3.2. Evaluasi Model Regresi

Model *Random Forest* Regressor yang dilatih menggunakan fitur hasil seleksi menunjukkan performa prediksi yang sangat tinggi. Nilai koefisien determinasi (R^2) mencapai 0.9938, yang berarti bahwa hampir 99,4% variasi dalam PDRB per kapita dapat dijelaskan oleh kombinasi fitur terpilih. Sementara itu, nilai *Root Mean squared error* (RMSE) yang diperoleh adalah 2467.54, mengindikasikan bahwa kesalahan prediksi relatif kecil terhadap rentang nilai target.

Disajikan pada Gambar 1, heatmap korelasi antar variabel ekonomi dan sosial yang digunakan dalam penelitian ini, termasuk hubungan antara masing-masing variabel terhadap target utama, yaitu Produk Domestik Regional Bruto (PDRB) per kapita. Warna merah menunjukkan korelasi positif, sedangkan biru menandakan korelasi negatif, dengan intensitas warna menggambarkan kekuatan korelasi. Korelasi ditampilkan dalam bentuk angka antara -1 hingga 1.

Dari visualisasi ini, terlihat bahwa variabel `grdp_employed` memiliki korelasi sangat tinggi terhadap `grdp_capita` (0.99), yang menunjukkan bahwa jumlah pekerja yang berkontribusi dalam kegiatan ekonomi sangat berkaitan dengan tingkat PDRB per kapita. Selain itu, variabel seperti `hs_graduation`, `clean_water`, `foreign_invest`, dan `domestic_invest` juga memperlihatkan korelasi positif yang cukup kuat terhadap PDRB per kapita, menunjukkan pentingnya pendidikan menengah, akses terhadap layanan dasar, dan masuknya investasi dalam mendorong pertumbuhan ekonomi inklusif.

Sebaliknya, variabel `poverty` dan `unemployed_percent` menunjukkan korelasi negatif terhadap PDRB per kapita. Korelasi negatif ini logis karena tingginya tingkat kemiskinan dan pengangguran umumnya berasosiasi dengan rendahnya output ekonomi per kapita di suatu wilayah. Variabel seperti `gini` (ketimpangan), walaupun berkorelasi positif terhadap `grdp_capita`, menunjukkan nilai korelasi yang relatif lemah (0.13), menandakan bahwa meskipun wilayah dengan PDRB tinggi bisa mengalami ketimpangan, hubungan liniernya tidak dominan.



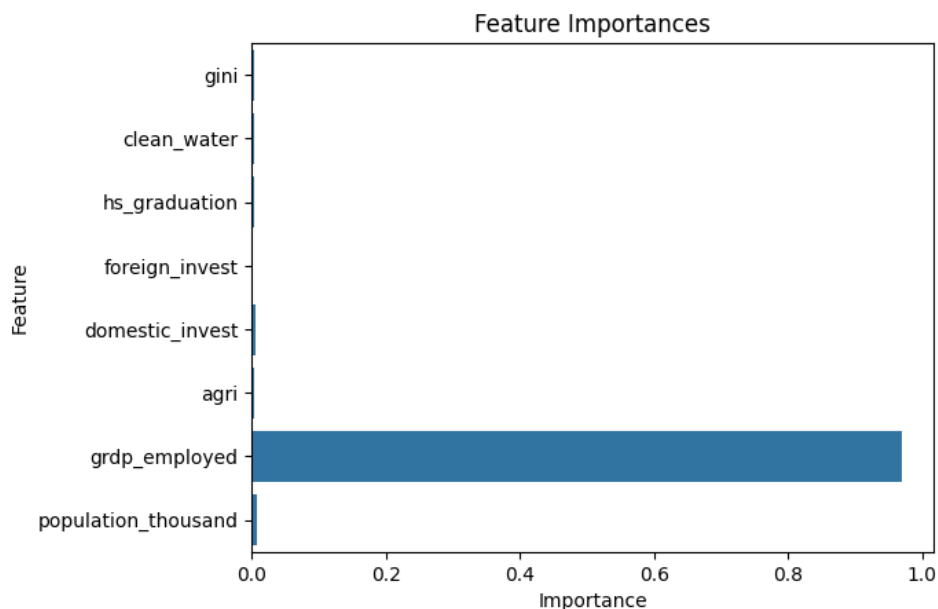
Gambar 1. Correlation Heatmap

Korelasi yang sangat tinggi antar beberapa variabel lainnya, seperti antara `gov_spending` dan `gov_capex` (0.93), serta `population_thousand` dan `gov_spending` (0.95), menunjukkan potensi multikolinearitas yang perlu diantisipasi dalam pemodelan, terutama dalam konteks seleksi fitur atau feature reduction. Oleh karena itu, heatmap ini menjadi alat awal yang penting dalam menilai keterkaitan antar variabel dan sebagai dasar awal dalam proses pemilihan fitur menggunakan algoritma optimasi seperti *Whale Optimization Algorithm* (WOA).

Visualisasi korelasi antar variabel yang ditunjukkan pada Gambar 1 menunjukkan adanya keterkaitan kuat antara variabel terpilih dengan target, yang memperkuat validitas hasil seleksi fitur. Selain itu, analisis feature importance yang ditunjukkan pada Gambar 2 dari model final mengonfirmasi bahwa variabel seperti `grdp_employed`, `population_thousand`, dan `foreign_invest` memiliki kontribusi dominan dalam model.

Kemudian Gambar 2 menampilkan grafik Feature Importances yang menggambarkan tingkat kontribusi masing-masing variabel input dalam mempengaruhi hasil prediksi PDRB per kapita oleh model. Nilai importance diperoleh dari proses pelatihan model regresi berbasis algoritma decision tree ensemble, yang secara umum menilai seberapa sering dan signifikan suatu fitur digunakan dalam membagi data selama proses pelatihan.

Hasil visualisasi menunjukkan bahwa variabel `grdp_employed` memiliki pengaruh dominan yang sangat signifikan dibandingkan fitur lainnya. Dengan tingkat kepentingan mendekati 1, `grdp_employed` menjadi penentu utama dalam prediksi nilai PDRB per kapita. Hal ini selaras dengan temuan pada Gambar 1 (*heatmap* korelasi) yang menunjukkan korelasi sangat tinggi (0.99) antara `grdp_employed` dan `grdp_capita`. Artinya, semakin banyak tenaga kerja yang terlibat dalam kegiatan ekonomi, semakin tinggi pula produktivitas wilayah secara keseluruhan.



Gambar 2. Tingkat Kepentingan (Importance) Variabel Terpilih

Sebaliknya, variabel lain seperti `gini`, `clean_water`, `hs_graduation`, `foreign_invest`, `domestic_invest`, `agri`, dan `population_thousand` memiliki kontribusi yang sangat kecil terhadap model. Meskipun secara teoritis variabel-variabel ini relevan terhadap perkembangan ekonomi, kontribusi mereka dalam model ini sangat rendah. Hal ini bisa disebabkan oleh adanya korelasi yang lebih tinggi antar variabel-variabel tersebut dengan `grdp_employed`, sehingga informasi yang mereka bawa telah terserap sepenuhnya oleh fitur dominan tersebut.

Temuan ini menggarisbawahi pentingnya proses seleksi fitur dalam pemodelan prediktif. Jika model hanya mengandalkan satu fitur utama, ada risiko *overfitting* atau ketergantungan yang terlalu besar terhadap satu indikator. Oleh karena itu, meskipun `grdp_employed` menjadi variabel paling informatif dalam kasus ini, evaluasi lanjutan terhadap interaksi variabel dan eksperimen dengan model regularisasi sangat disarankan untuk meningkatkan generalisasi model.

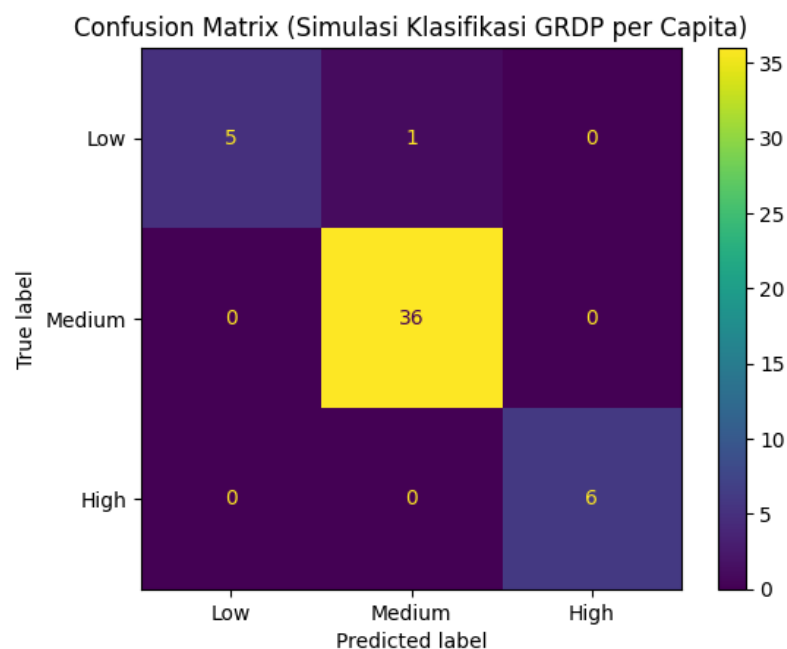
3.3. Evaluasi Klasifikasi Kesejahteraan Daerah

Untuk memberikan perspektif kebijakan yang lebih praktis, nilai PDRB per kapita diklasifikasikan ke dalam tiga kategori kesejahteraan: *Low*, *Medium*, dan *High*. Model menunjukkan performa klasifikasi yang sangat baik, dengan akurasi mencapai 97,92%, serta nilai precision, recall, dan *F1-score* yang konsisten tinggi di seluruh kelas, ditunjukkan pada Tabel 2 dan Gambar 3.

Tabel 2. Hasil klasifikasi PDRB per kapita

Kelas	Precision	Recall	F1-Score
<i>Low</i>	1.00	1.00	1.00
<i>Medium</i>	1.00	0.83	0.91
<i>High</i>	0.97	1.00	0.99

Hasil klasifikasi PDRB per kapita menunjukkan performa model yang sangat baik dalam membedakan tingkat kesejahteraan regional ke dalam tiga kelas, yaitu *Low*, *Medium*, dan *High*. Nilai precision dan recall yang sempurna (1.00) pada kelas *Low* menunjukkan bahwa model mampu mengenali seluruh wilayah dengan PDRB rendah tanpa kesalahan prediksi. Untuk kelas *Medium*, meskipun precision tetap tinggi (1.00), recall sedikit lebih rendah (0.83), yang mengindikasikan adanya beberapa wilayah kelas *Medium* yang salah diklasifikasikan ke kelas lain. Sementara itu, kelas *High* memperoleh recall sempurna (1.00) dan precision sebesar 0.97, menandakan bahwa hampir semua prediksi kelas tinggi akurat. Secara keseluruhan, kombinasi nilai *F1-score* yang tinggi di semua kelas mengonfirmasi bahwa model yang dibangun melalui pendekatan *Whale Optimization Algorithm* dan *Random Forest* memiliki kapabilitas klasifikasi yang andal dalam konteks analisis kesejahteraan ekonomi daerah.



Gambar 3. Confusion matrix Klasifikasi Tingkat Kesejahteraan Wilayah Berdasarkan PDRB

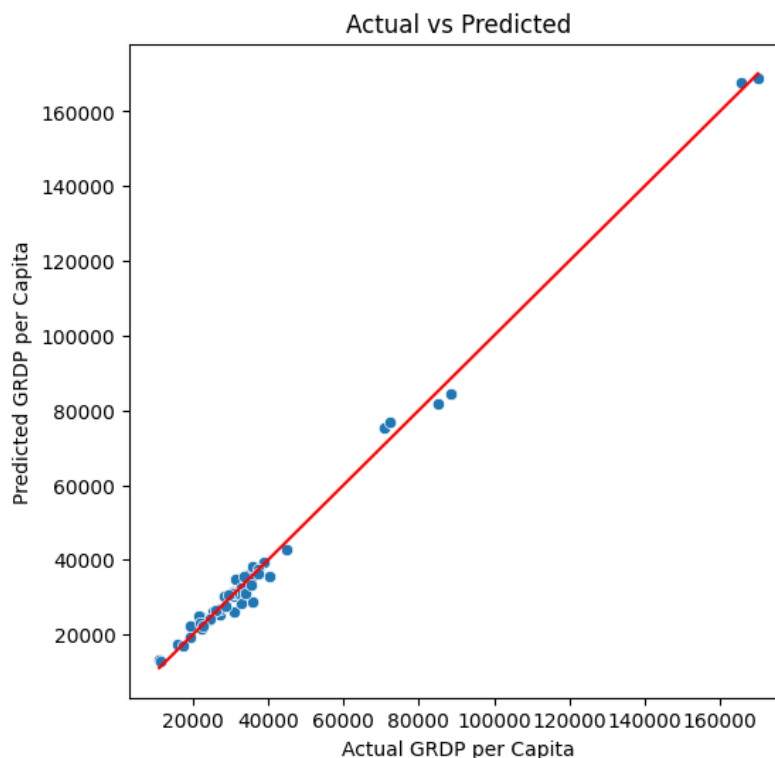
Lalu disajikan *Confusion matrix* dari hasil simulasi klasifikasi kategori Produk Domestik Regional Bruto (PDRB) per kapita ke dalam tiga kelas: *Low*, *Medium*, dan *High*. *Confusion matrix* ini digunakan untuk mengevaluasi performa model klasifikasi dalam memetakan prediksi terhadap nilai sebenarnya. Setiap sel pada matriks merepresentasikan jumlah observasi yang diklasifikasikan ke dalam masing-masing kategori.

Hasil klasifikasi menunjukkan bahwa model memiliki performa yang sangat baik, dengan sebagian besar prediksi berada pada diagonal utama (yang menandakan prediksi benar). Kategori *Medium* memiliki performa tertinggi dengan 36 data yang diprediksi benar dari total 36, menunjukkan akurasi sempurna untuk kelas tersebut. Kategori *High* juga menunjukkan performa sangat baik dengan 6 prediksi benar dari 6 data, tanpa adanya kesalahan klasifikasi. Untuk kategori *Low*, model memprediksi 5 dari 6 data dengan benar, dan 1 data salah diklasifikasikan ke dalam kategori *Medium*.

Tingkat kesalahan klasifikasi yang rendah ini diperkuat oleh nilai-nilai metrik evaluasi: akurasi mencapai 97.92%, dengan precision dan recall yang tinggi pada setiap kelas. Nilai *F1-score* juga sangat tinggi (rata-rata tertimbang sebesar 0.9784), menandakan keseimbangan antara precision dan recall. Kesalahan minor yang terjadi pada kategori *Low* dapat disebabkan oleh ambiguitas nilai ambang antar kelas, atau keterbatasan jumlah data pelatihan pada kelas tersebut.

Confusion matrix ini menunjukkan bahwa model klasifikasi yang digunakan sangat efektif dalam mengkategorikan wilayah berdasarkan PDRB per kapita. Namun, perlu dilakukan validasi silang tambahan serta eksplorasi terhadap teknik penyeimbangan kelas, terutama jika terjadi ketimpangan distribusi data pada masing-masing kategori.

Kemudian Gambar 4 memperlihatkan hubungan antara nilai aktual dan nilai prediksi yang hampir identik, menegaskan bahwa model memiliki presisi tinggi dan tidak mengalami *overfitting* signifikan.



Gambar 4. Plot Sebaran Nilai Aktual vs Prediksi PDRB per Kapita

Gambar 4 menunjukkan sebaran antara nilai aktual dan nilai prediksi PDRB per kapita yang divisualisasikan dalam bentuk diagram pencar (*scatter plot*). Setiap titik pada plot merepresentasikan satu observasi, di mana sumbu X menunjukkan nilai aktual PDRB per kapita dan sumbu Y menunjukkan nilai prediksi dari model. Garis merah diagonal yang membentang dari kiri bawah ke kanan atas merepresentasikan garis ideal di mana nilai aktual sama dengan nilai prediksi ($y = x$). Titik-titik yang berada tepat pada garis ini menunjukkan prediksi yang sangat akurat.

Sebagian besar titik terlihat berdekatan dengan garis merah, yang menandakan bahwa model prediksi memiliki performa yang baik dalam memperkirakan nilai PDRB per kapita. Kedekatan ini mencerminkan error yang rendah antara nilai aktual dan nilai prediksi, yang berarti bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang dianalisis. Titik-titik yang tersebar merata dan tidak membentuk pola menyimpang menunjukkan bahwa tidak ada bias sistematis dalam prediksi.

Namun, terdapat beberapa titik yang sedikit menjauh dari garis merah, khususnya pada nilai PDRB yang lebih tinggi. Hal ini menunjukkan bahwa pada wilayah dengan PDRB per kapita yang tinggi, prediksi model cenderung mengalami deviasi lebih besar dibandingkan dengan wilayah-wilayah yang memiliki nilai PDRB menengah atau rendah. Meskipun demikian, penyimpangan tersebut tidak terlalu signifikan dan masih berada dalam rentang toleransi yang wajar. Gambar 4 memberikan gambaran visual bahwa model regresi yang digunakan cukup efektif dalam memprediksi nilai PDRB per kapita. Kinerja model terlihat cukup konsisten di seluruh rentang data, dan tidak menunjukkan pola outlier yang ekstrem. Oleh karena itu, model ini dapat diandalkan untuk digunakan dalam analisis atau perencanaan kebijakan ekonomi yang memerlukan estimasi nilai PDRB per kapita.

3.4. Pembahasan Temuan

Hasil penelitian menunjukkan bahwa penerapan *Whale Optimization Algorithm* (WOA) dalam proses seleksi fitur berhasil meningkatkan efisiensi model prediksi. Penurunan nilai *mean squared error* (MSE) dari 1.072.716 menjadi 962.529 setelah 30 iterasi menunjukkan bahwa WOA efektif dalam menemukan kombinasi variabel yang relevan untuk meningkatkan akurasi prediksi PDRB per kapita. Seleksi terhadap delapan variabel terpilih mencerminkan tidak hanya kemampuan WOA dalam mengeliminasi variabel yang kurang informatif, tetapi juga dalam mempertahankan variabel-variabel yang secara ekonomis signifikan terhadap pertumbuhan wilayah.

Model regresi menggunakan *Random Forest Regressor* memberikan hasil yang sangat memuaskan, ditunjukkan oleh nilai R^2 sebesar 0.9938. Ini menandakan bahwa hampir seluruh variasi nilai PDRB per kapita dapat dijelaskan oleh variabel terpilih. Selain itu, nilai RMSE yang relatif kecil memperkuat kesimpulan bahwa model memiliki

tingkat kesalahan prediksi yang rendah. Hal ini menunjukkan bahwa pendekatan kombinasi antara seleksi fitur berbasis algoritma metaheuristik dan pemodelan regresi berbasis ensemble trees sangat efektif dalam konteks pemodelan ekonomi makro.

Dari analisis korelasi yang divisualisasikan dalam bentuk heatmap, ditemukan bahwa variabel *grdp_employed* memiliki korelasi yang sangat tinggi terhadap target variabel (*grdp_capita*), yang juga dikonfirmasi dalam analisis *feature importance* sebagai fitur dengan pengaruh paling dominan. Meskipun demikian, dominasi fitur tunggal ini menimbulkan risiko ketergantungan berlebih (*over-reliance*), yang bisa mengarah pada *overfitting* apabila digunakan pada data yang berbeda. Oleh karena itu, penting untuk mengkombinasikan pendekatan ini dengan teknik validasi lanjutan atau penggunaan model regularisasi agar ketahanan model tetap terjaga dalam konteks data baru.

Temuan lain yang menarik adalah bahwa beberapa variabel seperti *gini*, *hs_graduation*, dan *foreign_invest* meskipun secara teoritis relevan terhadap pertumbuhan ekonomi, memiliki pengaruh yang jauh lebih kecil terhadap model. Hal ini mungkin menunjukkan bahwa kontribusi variabel-variabel tersebut telah diwakili secara tidak langsung oleh variabel dominan seperti *grdp_employed* atau *population_thousand*, atau bisa juga karena keterbatasan dalam jumlah dan distribusi data. Fenomena ini menyoroti pentingnya tidak hanya seleksi fitur secara kuantitatif, tetapi juga pemahaman substansi terhadap relasi antar variabel dalam konteks kebijakan ekonomi dan pembangunan daerah.

Dalam konteks klasifikasi tingkat kesejahteraan daerah, model menunjukkan performa sangat baik, dengan akurasi mencapai hampir 98%. Kemampuan model dalam memetakan wilayah berdasarkan kategori PDRB per kapita memberikan nilai tambah yang besar dalam mendukung pengambilan kebijakan berbasis data. *Confusion matrix* memperlihatkan bahwa prediksi yang salah hanya terjadi dalam jumlah sangat terbatas, dan cenderung pada ambang batas antar kelas. Meski demikian, hasil ini menegaskan bahwa pendekatan klasifikasi dengan supervisi yang kuat serta seleksi fitur yang relevan dapat menjadi alat yang efektif dalam pengelompokan wilayah menurut tingkat kemakmuran secara objektif dan akurat.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penggunaan *Whale Optimization Algorithm* (WOA) sebagai metode seleksi fitur terbukti efektif dalam meningkatkan performa model prediksi PDRB per kapita. WOA berhasil menyaring 8 variabel paling relevan dari total 18 variabel awal, dengan menurunkan nilai MSE secara signifikan, yang mencerminkan perbaikan kualitas prediksi.

Model *Random Forest* Regressor yang dilatih menggunakan fitur hasil seleksi menunjukkan performa yang sangat tinggi, dengan nilai R^2 sebesar 0.9938 dan RMSE yang relatif rendah. Hal ini menunjukkan bahwa model mampu menjelaskan hampir seluruh variasi dalam data PDRB per kapita secara akurat. Visualisasi *feature importance* dan korelasi juga menunjukkan bahwa variabel *grdp_employed* menjadi faktor dominan dalam mempengaruhi prediksi, diikuti oleh variabel-variabel lain seperti *population_thousand* dan *foreign_invest*.

Selain itu, model klasifikasi kesejahteraan wilayah berdasarkan PDRB per kapita menghasilkan akurasi sebesar 97,92% dengan nilai *precision*, *recall*, dan *F1-score* yang tinggi di seluruh kategori. Hasil *Confusion matrix* menunjukkan bahwa model mampu mengklasifikasikan wilayah ke dalam kategori kesejahteraan (*Low*, *Medium*, *High*) dengan sangat baik.

Secara keseluruhan, pendekatan integratif antara seleksi fitur berbasis algoritma metaheuristik (WOA) dan model prediktif berbasis ensemble learning memberikan solusi yang akurat dan efisien untuk analisis dan klasifikasi kesejahteraan ekonomi wilayah. Temuan ini dapat dijadikan dasar dalam pengambilan keputusan kebijakan yang lebih tepat sasaran, terutama dalam upaya peningkatan kesejahteraan masyarakat secara regional. Didapatkan bahwa aspek-aspek inklusivitas secara kontekstual, seperti peran serta sosial, kemudahan akses terhadap layanan, serta keterlibatan kelompok rentan dalam proses pertumbuhan ekonomi menarik dikaji lebih lanjut dalam penelitian mendatang.

DAFTAR RUJUKAN

- [1] A. J. van Niekerk, "Inclusive economic sustainability: SDGs and global inequality," *Sustainability (Switzerland)*, vol. 12, no. 13, Jul. 2020, doi: 10.3390/su12135427.
- [2] S. D. Yeboah, J. G. Gatsi, M. O. Appiah, and M. P. Fumey, "Examining the drivers of inclusive growth: A study of economic performance, environmental sustainability, and life expectancy in BRICS economies," *Research in Globalization*, vol. 9, Dec. 2024, doi: 10.1016/j.resglo.2024.100267.
- [3] T. Kyriazos and M. Poga, "Application of *Machine Learning* Models in Social Sciences: Managing Nonlinear Relationships," *Encyclopedia*, vol. 4, no. 4, pp. 1790–1805, Nov. 2024, doi: 10.3390/encyclopedia4040118.
- [4] F. Pang, G. Miao, Y. Li, and Y. Shi, "Key factors influencing sustainable population growth: A DEMATEL-ANP combined approach," *Heliyon*, vol. 10, no. 21, Nov. 2024, doi: 10.1016/j.heliyon.2024.e39404.
- [5] H. Islam, "Nexus of economic, social, and environmental factors on sustainable development goals: The moderating role of technological advancement and green innovation," *Innovation and Green Development*, vol. 4, no. 1, Feb. 2025, doi: 10.1016/j.igd.2024.100183.
- [6] T. Sihol Nababan, R. Panjaitan, F. Panjaitan, and S. Napitupulu, "investigating the relationship between economic growth and human development index: are they simultaneously related?," 2024, [Online]. Available: <https://sultanist.ac.id/index.php/sultanist>
- [7] I. Mahdi and S. W. Sulistyono, "Determinants of gini ratio: a case study of Low gini ratio in Madura," *Entrepreneurship Bisnis Manajemen Akuntansi (E-BISMA)*, pp. 245–260, Dec. 2023, doi: 10.37631/ebisma.v4i2.1119.
- [8] M. Muhtar and D. L. W., "Analysis Inequality Income in Indonesia 2020," *Journal of International Conference Proceedings*, vol. 4, no. 3, Dec. 2021, doi: 10.32535/jicp.v4i3.1375.
- [9] I. Safitri and J. Hadi Susilo, "Analysis Factors Influencing Economic Growth, Unemployment and Poverty in Indonesia," *JEJAK: Jurnal Ekonomi dan Kebijakan*, vol. 16, no. 2, p. 16, 2023, doi: 10.15294/jejak.v16i2.42032.
- [10] M. Y. Messa, N. Bariyah, and E. P. Kurniasih, "Determining Factors of Economic Growth and Poverty in West Kalimantan," *Jurnal Ilmu Sosial dan Humaniora*, vol. 12, no. 3, pp. 530–546, Dec. 2023, doi: 10.23887/jish.v12i3.67197.
- [11] Suyanto, "Economics Development Analysis Journal Foreign Direct Investment: Does it Increase Economic Growth? Article Information," *Economics Development Analysis Journal*, vol. 12, no. 1, 2023, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/edaj>
- [12] L. Saly and A. Satrianto, "The Factors Influencing Economic Growth and Foreign Direct Investment: A Case Study in Cambodia," *Journal of Ecohumanism*, vol. 3, no. 8, Jan. 2025, doi: 10.62754/joe.v3i8.5685.
- [13] F. Septian, "Exploring the Performance of *Whale Optimization Algorithm* on Rosenbrock's Function," *Journal of Intelligent Systems and Information Technology*, vol. 1, no. 2, p. 2024, 2024, [Online]. Available: <https://jurnalapik.id/index.php/jisit>
- [14] I. Mediansyah, F. Septian, and A. Zikry, "Penerapan *Whale Optimization Algorithm* dalam Pengoptimalan Portofolio Investasi Menggunakan Model Prediktif Artificial Intelligence," vol. 2, no. 1, 2024.
- [15] R. Hornung and A. L. Boulesteix, "Interaction forests: Identifying and exploiting interpretable quantitative and qualitative interaction effects," *Comput Stat Data Anal*, vol. 171, Jul. 2022, doi: 10.1016/j.csda.2022.107460.
- [16] W. Liu, Z. Liu, Z. Liu, S. Xiong, and S. Zhang, "Random Forest and *Whale Optimization Algorithm* to Predict the Invalidation Risk of Backfilling Pipeline," *Mathematics*, vol. 11, no. 7, Apr. 2023, doi: 10.3390/math11071636.
- [17] A. M. Nassef, M. A. Abdelkareem, H. M. Maghrabie, and A. Baroutaji, "Review of Metaheuristic Optimization Algorithms for Power Systems Problems," Jun. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/su15129434.
- [18] Á. Bueno-Ferrer, J. De Pablo Valenciano, and J. De Burgos Jiménez, "Unveiling the Potential of Metaheuristics in Transportation: A Path Towards Efficiency, Optimization, and Intelligent Management," *Infrastructures (Basel)*, vol. 10, no. 1, Jan. 2025, doi: 10.3390/infrastructures10010004.
- [19] K. Lepenioti, A. Bousdekis, D. Apostolou, and G. Mentzas, "Prescriptive analytics: Literature review and research challenges," Feb. 01, 2020, *Elsevier Ltd*. doi: 10.1016/j.ijinfomgt.2019.04.003.