JURNAL FASILKOM
Volume 16 No. 1 | April 2026: 1-9

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Perbandingan Algoritma Regresi dalam Memprediksi Penjualan Berdasarkan Indikator Sosial Ekonomi Kabupaten Cirebon (2010-2023)

Muthia Rahmah[1], Kanaya Ramadanti[2], Imelda Fransiska Aulia[3]
[1,2,3] Pendidikan Teknik Informatika dan Komputer, Pendidikan, Institut Prima Bangsa
[1]12323031@ipbcirebon.ac.id*, [2]12323030@ipbcirebon.ac.id, [3]12323033@ipbcirebon.ac.id

Abstract

A comparative study of four regression algorithms, namely Support Vector Regression (SVR), Gradient Boosting Regressor (GBR), Random Forest Regressor (RFR), and Extreme Gradient Boosting (XGBoost), was conducted to predict annual aggregate sales based on socioeconomic indicators in Cirebon Regency from 2010 to 2023. The study utilized secondary data obtained from the Central Bureau of Statistics (Badan Pusat Statistik) of Cirebon Regency. Five predictor variables were employed, including life expectancy, expected years of schooling, mean years of schooling, per capita expenditure, and the Human Development Index (HDI). Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R-squared). The experimental results indicate that the GBR model achieved the best predictive performance, with the lowest error values (MAE = 127.98 and RMSE = 185.63) and the highest R^2 value (0.94), outperforming RFR, XGBoost, and SVR after parameter tuning. Feature importance analysis consistently identified life expectancy as the most influential variable across models. These findings demonstrate that ensemble-based regression methods, particularly boosting algorithms, are effective for modeling complex socioeconomic patterns and can support data-driven economic forecasting and regional policy planning.

Keywords: ***regression, SVR, GBR, RFR, XGBoost***

Abstrak

Penelitian ini membandingkan kinerja empat algoritma regresi, yaitu *Support Vector Regression* (SVR), *Gradient Boosting Regressor* (GBR), *Random Forest Regressor* (RFR), dan *Extreme Gradient Boosting* (XGBoost), dalam memprediksi nilai penjualan agregat tahunan berdasarkan indikator sosial ekonomi Kabupaten Cirebon periode 2010–2023. Data yang digunakan merupakan data sekunder yang bersumber dari *Badan Pusat Statistik* Kabupaten Cirebon. Lima variabel prediktor digunakan dalam pemodelan, yaitu umur harapan hidup, harapan lama sekolah, rata-rata lama sekolah, pengeluaran per kapita, dan Indeks Pembangunan Manusia (IPM). Evaluasi performa model dilakukan menggunakan metrik *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan koefisien determinasi (*R-squared*). Hasil pengujian menunjukkan bahwa model GBR memberikan performa terbaik dengan nilai kesalahan terendah (MAE = 127,98 dan RMSE = 185,63) serta nilai R^2 tertinggi sebesar 0,94, dibandingkan dengan RFR, XGBoost, dan SVR setelah dilakukan *tuning* parameter. Analisis *feature importance* mengungkapkan bahwa variabel umur harapan hidup merupakan faktor paling berpengaruh dalam memprediksi penjualan. Temuan ini menunjukkan bahwa model regresi berbasis *ensemble*, khususnya metode *boosting*, efektif dalam memodelkan hubungan kompleks indikator sosial ekonomi dan dapat dimanfaatkan dalam perencanaan ekonomi daerah berbasis data.

Kata kunci: regresi, SVR, GBR, RFR, XGBoost

1. Pendahuluan

Prediksi penjualan sangat penting dalam perencanaan strategis dan pengambilan keputusan operasional, terutama untuk mengantisipasi permintaan pasar [1]. Manfaat utama dari prediksi penjualan adalah pengelolaan stok bahan baku yang lebih efisien [2]. Secara tradisional, prediksi penjualan sering bergantung pada data historis internal perusahaan. Namun, dalam skala wilayah, nilai penjualan agregat juga dipengaruhi oleh kondisi sosial ekonomi masyarakat.

Dalam hal ini, penjualan merujuk pada total nilai penjualan agregat tahunan di Kabupaten Cirebon, yang mencerminkan tingkat konsumsi masyarakat secara umum. Faktor-faktor sosial ekonomi seperti Indeks Pembangunan Manusia (IPM), pengeluaran per kapita, harapan lama sekolah, rata-rata lama sekolah, serta umur harapan hidup tidak secara langsung mencatat transaksi penjualan. Namun, indikator-indikator tersebut dapat digunakan sebagai proksi untuk mengukur daya beli serta potensi konsumsi masyarakat [3]. Misalnya, IPM mencerminkan aspek pendidikan, kesehatan, dan standar hidup yang berdampak pada perilaku konsumsi, sementara pengeluaran per kapita merefleksikan kapasitas konsumsi masyarakat pada periode tertentu [4].

Dengan demikian, indikator sosial ekonomi layak digunakan sebagai variabel prediktor dalam model estimasi penjualan. Model ini dapat membantu memahami pola-pola makro yang memengaruhi dinamika pasar lokal, terutama di wilayah dengan karakteristik pembangunan yang beragam seperti Kabupaten Cirebon.

Seiring dengan kemajuan teknologi, penerapan *data mining* menggunakan algoritma pembelajaran mesin *(machine learning)* telah menunjukkan potensi besar

dalam berbagai bidang prediksi, termasuk prediksi penjualan. Beberapa algoritma regresi yang banyak digunakan dan terbukti unggul dalam berbagai studi adalah *Support Vector Regression* yang dikenal dengan singkatan SVR, model Gradient *Boosting* Regression yang biasa disebut GBR, model *Random Forest Regression* atau RFR, serta XGBoost.

Penerapan algoritma-algoritma tersebut mencakup berbagai domain. Random Forest, sebagai metode *ensemble* berbasis *bagging*, terbukti stabil dan andal dalam berbagai skenario. Penelitian menunjukkan bahwa RFR lebih stabil dibandingkan XGBoost dalam memprediksi penjualan kopi [2]. RFR juga berhasil diterapkan untuk prediksi penjualan di supermarket dengan tingkat akurasi yang tinggi [5]. Hasil komparatif menunjukkan bahwa SVR dapat lebih unggul dari RFR dalam prediksi risiko kredit [6], sementara dalam prediksi harga emas, SVR menunjukkan akurasi lebih tinggi dan RFR menghasilkan tingkat galat yang lebih rendah [7].

SVR secara konsisten memberikan hasil yang baik, terutama ketika parameter dan fungsi kernel dioptimalkan. SVR telah berhasil diterapkan dalam peramalan penjualan roti memperoleh Root Mean Square Error (RMSE) yang sangat rendah [1]. Studi lainnya menunjukkan bahwa SVR mengungguli RFR dalam prediksi laju penguapan harian [8]. Pengadopsi kernel *Radial Basis Function* (RBF) juga terbukti meningkatkan akurasi dalam prediksi risiko kredit [6], dan hasil serupa diperoleh dalam studi peramalan curah hujan [9].

Metode *boosting* seperti Gradient *Boosting* dan XGBoost juga dilaporkan memberikan performa prediksi yang sangat baik dalam berbagai studi. *Gradient Boosting* menunjukkan hasil yang lebih baik dibandingkan *Random Forest* dalam klasifikasi IPM [3], serta unggul dalam prediksi depresi mahasiswa [10], emisi karbon [11], dan estimasi performa kompresor [12]. Sementara itu, XGBoost menunjukkan performa terbaik dalam prediksi kehilangan pelanggan (*customer churn*) [13], serta dalam prediksi cuaca, mengalahkan SVR dan *Random Forest* [14]. Studi komprehensif mengenai prediksi magnitudo gempa juga menempatkan algoritma berbasis *boosting* seperti LightGBM dan XGBoost sebagai model terbaik [15]. Dalam studi klasifikasi lainnya, Random Forest dan XGBoost sering bersaing dalam hal akurasi [16].

Meskipun banyak studi telah membandingkan algoritma regresi dalam berbagai domain seperti prediksi kualitas udara [17], analisis serbuk sari [18], hingga risiko kredit [6], belum ditemukan penelitian yang secara khusus membandingkan kinerja SVR, RFR, GBR, dan XGBoost untuk memprediksi volume penjualan berdasarkan indikator sosial ekonomi regional dalam jangka panjang, khususnya di Kabupaten Cirebon. Kesenjangan kajian ini menunjukkan adanya celah pengetahuan yang signifikan dalam penelitian ini, yaitu bagaimana efektivitas keempat algoritma tersebut dalam memodelkan penjualan berdasarkan data sosial ekonomi lokal.

Dengan demikian, kajian ini bertujuan agar kinerja empat algoritma regresi tersebut dibandingkan dalam memprediksi penjualan berdasarkan data indikator sosial ekonomi Kabupaten Cirebon dari tahun 2010 hingga 2023. Temuan kajian ini diharapkan bermanfaat secara praktis bagi pelaku usaha maupun perencana kebijakan ekonomi daerah dalam merumuskan strategi berbasis data.

Sebagai usulan pengembangan, penelitian selanjutnya dapat mengeksplorasi teknik optimisasi *hyperparameter* lanjutan, seperti *Particle Swarm Optimization* (PSO), yang telah terbukti efektif dalam meningkatkan performa SVR pada berbagai studi [9].

2. Metode Penelitian

Setelah mengembangkan permasalahan dan tujuan penelitian, tahapan selanjutnya adalah menyusun metode yang sesuai untuk menguji hipotesis dan menjawab pertanyaan penelitian. Dalam penelitian ini, diterapkan metode analisis untuk mempelajari interaksi serta keterhubungan antar variabel sosial ekonomi dan nilai penjualan agregat melalui teknik proyektif berbasis algoritma regresi. Pemilihan metode eksperimen berbantuan *machine learning* bertujuan untuk memperoleh pola terbaik yang dapat merepresentasikan pola historis secara kuantitatif. Dengan demikian, metode ini tidak hanya menjelaskan hubungan antarvariabel, tetapi juga menguji performa prediksi dari masing-masing algoritma yang diterapkan pada data multivariat dalam rentang waktu yang telah ditentukan.

Penentuan empat algoritma regresi, seperti didasarkan pada keunikan dan keistimewaan masing- *Support Vector Regression* (SVR), *Gradient Boosting Regressor* (GBR), *Random Forest Regressor* (RFR), dan *Extreme Gradient Boosting* (XGBoost) masing yang telah terbukti efektif dalam penelitian sebelumnya. SVR unggul dalam mengatasi data berdimensi tinggi dan memberikan prediksi yang stabil, terutama pada dataset berukuran kecil hingga menengah, dengan menentukan *hyperplane* yang optimal dalam ruang fitur berdimensi tinggi. Di sisi lain, RFR merupakan model *ensemble* dari pohon keputusan yang dikenal tangguh terhadap *overfitting* dan mampu mengatasi data dengan *noise* tinggi, berkat kemampuannya menyatukan hasil dari banyak pohon keputusan untuk meningkatkan akurasi prediksi. Sementara itu, GBR memiliki kemampuan untuk mengoreksi kesalahan prediksi secara bertahap, menjadikannya sangat cocok untuk memodelkan hubungan non-linier antarvariabel, karena membangun model secara iteratif dengan fokus pada kesalahan model sebelumnya. XGBoost, sebagai pengembangan

dari GBR, menawarkan kinerja komputasi yang tinggi dan teknik regularisasi yang kuat, sehingga sangat

P-ISSN : 2089-3353
E-ISSN : 2808-9162

populer dalam berbagai kompetisi *data science* dan aplikasi prediksi berskala besar. Dengan menggabungkan keempat model ini, diharapkan analisis dapat memberikan tinjauan lengkap mengenai performa algoritma regresi dalam konteks prediksi penjualan yang berbasis pada indikator sosial ekonomi, sehingga menunjukkan prediksi yang menunjukkan prediksi yang lebih meyakinkan.

2.1 Data dan Variabel

Analisis ini menerapkan metode statistik pengujian yang berbasis pada teknik *data mining* untuk membandingkan hasil empat algoritma regresi dalam mengantisipasi total penjualan tahunan di Kabupaten Cirebon. Rentang waktu yang digunakan mencakup tahun 2010 hingga 2023, dengan total empat belas observasi tahunan yang berasal dari sumber terverifikasi dan resmi, yaitu Badan Pusat Statistik (BPS) Kabupaten Cirebon, melalui laman: https://cirebonkab.bps.go.id/. Variabel yang menjadi target dalam penelitian ini adalah nilai penjualan agregat tahunan, yang mencerminkan tingkat konsumsi dan daya beli masyarakat secara umum di wilayah tersebut. Dalam penelitian ini, nilai penjualan agregat diproksikan menggunakan variabel pengeluaran per kapita sebagai representasi daya beli masyarakat.

Sebagai variabel predikator, digunakan lima faktor sosial ekonomi yang dianggap mencerminkan aspek kesejahteraan masyarakat, yaitu umur harapan hidup (UHH), harapan lama sekolah (HLS), rata-rata lama sekolah (RLS), pengeluaran per kapita, dan indeks pembangunan manusia (IPM). Penentuan kelima faktor ini didasarkan pada penelitian sebelumnya yang menunjukkan hubungannya dengan pola konsumsi dan aktivitas ekonomi masyarakat. Dengan demikian, faktor sosioekonomi dianggap memiliki dampak tidak langsung tetapi signifikan terhadap nilai penjualan tahunan karena terkait dengan kemampuan daya beli dan perilaku konsumsi penduduk.

Untuk memperjelas struktur data yang digunakan, Tabel 1 menyajikan cuplikan dataset yang menggambarkan format dan variabel penelitian. Cuplikan ini ditampilkan sebagai contoh representatif dari data tahunan yang digunakan dalam proses analisis dan pemodelan.

Tabel 1. Cuplikan Dataset Indikator Sosial Ekonomi Kabupaten Cirebon

Tahun	Umur Harapan Hidup (tahun)	Harapan Lama Sekolah (tahun)	Rata-rata Lama Sekolah (tahun)	Pengeluaran per Kapita (ribu rupiah)	IPM
2010	71,09	10,66	5,92	8.866,25	63,64
2015	71,38	11,79	6,32	9.261,30	66,07
2020	71,99	12,25	6,92	10.342,00	68,75
2023	72,76	12,41	7,64	11.128,00	70,95

2.2. Teknik Analisis dan Proses Eksperimen

Eksperimen ini dilaksanakan dengan menggunakan bahasa pemrograman Python dan platform *Google Colaboratory* (Colab) sebagai lingkungan pemrosesan data. Beberapa *library* penting diterapkan dalam proses ini, dengan menggunakan pandas dan numpy sebagai pengolahan data, matplotlib dan seaborn yang digunakan untuk visualisasi data, serta scikit-learn untuk pemodelan dan xgboost sebagai *library* utama untuk pembangunan dan pelatihan model regresi.

Tahapan awal dalam proses analisis dimulai dengan pengecekan kelengkapan data, termasuk identifikasi dan penanganan terhadap nilai-nilai yang hilang atau tidak konsisten. Setelah memastikan data dalam kondisi bersih, dilakukan normalisasi menggunakan metode StandardScaler untuk menyeragamkan skala antar fitur, terutama karena model seperti *Support Vector Regression* (SVR) sangat sensitif terhadap perbedaan skala. Selanjutnya, data dibagi menjadi dua subset dengan proporsi 70 persen untuk pelatihan dan 30 persen untuk pengujian menggunakan fungsi train_test_split.

Empat algoritma regresi digunakan dalam penelitian ini, yaitu *Support Vector Regression* yang dikenal dengan singkatan SVR, Gradient *Boosting* Regression yang biasa disebut GBR, Random Forest Regression atau RFR, serta XGBoost. Radial Basis Function (RBF) digunakan sebagai kernel dalam model SVR dengan parameter C dan epsilon yang dioptimalkan menggunakan teknik GridSearchCV untuk memperoleh model yang optimal. *Random Forest Regressor* diimplementasikan sebagai model *ensemble* berbasis *bagging* yang dikenal tangguh terhadap overfitting dan mampu menangani data kompleks. Sementara itu, *Gradient Boosting Regressor* membangun pohon keputusan secara sekuensial untuk meminimalkan kesalahan prediksi sebelumnya. XGBoost sebagai pengembangan dari GBR menawarkan efisiensi komputasi yang lebih tinggi dan mendukung regularisasi tambahan; model ini di *tuning* berdasarkan kombinasi parameter learning_rate, max_depth, n_estimators, serta subsample.

Sebagai contoh penerapan, berikut adalah potongan kode yang digunakan untuk melakukan *tuning* parameter SVR menggunakan GridSearchCV:

Program Analisis Regresi

```
from sklearn.model_selection import
GridSearchCV
from sklearn.svm import SVR

param_grid = {'C': [1, 10, 100], 'epsilon':
[0.01, 0.1, 1]}
svr = SVR(kernel='rbf')
grid_svr = GridSearchCV(svr, param_grid,
cv=5)
grid_svr.fit(X_train_scaled, y_train)
```

Seluruh proses eksperimen, termasuk *preprocessing*, pelatihan model, *tuning*, dan visualisasi, diarsipkan kode program tersedia secara online dalam bentuk

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Google Colab Notebook yang dapat diakses melalui tautan ini:

https://colab.research.google.com/drive/

2.3. Evaluasi dan Visualisasi Hasil

Penilaian terhadap performa model dilakukan dengan menggunakan tiga indikator evaluasi regresi yang umum digunakan, yaitu koefisien determinasi yang umum dikenal dengan istilah R-squared atau ditulis sebagai R^2, kemudian *Mean Absolute Error* yang sering disingkat MAE, serta *Root Mean Square Error* dengan singkatan RMSE. Koefisien dterminasi digunakan dalam menentukan seberapa besar variabilitas target dapat diprediksi oleh proporsi variansi dari perubahan nilai variabel dependen yang dapat ditangkap oleh model. Nilai R2 yang mendekati satu menunjukkan bahwa model memiliki performa prediktif yang unggul. Di sisi lain, MAE menilai akurasi model dengan menghitung rata-rata jarak absolut antara nilai perkiraan dan nilai aktual, tanpa mempertimbangkan arah kesalahan, sedangkan RMSE cenderung menekankan kesalahan prediksi yang besar karena menggunakan kuadrat selisih antara nilai aktual dengan nilai prediksi. Secara matematis, rumus dari MAE dan RMSE dituliskan pada rumus 1:

$$MAE = (1/n)\,\Sigma\,|yi - \hat{y}i|, RMSE = \sqrt{[(1/n)\,\Sigma\,(yi - \hat{y}i)^2]} \quad (1)$$

Model yang paling baik adalah model yang dimana hasil nilai R^2 tertinggi serta nilai MAE dan RMSE yang paling rendah. Sebagai pelengkap dari evaluasi numerik, penelitian ini juga menyajikan visualisasi hasil model dalam bentuk grafik hubungan antara nilai prediksi dan nilai aktual untuk masing-masing model. Grafik ini memungkinkan pembaca untuk melihat secara langsung apakah hasil prediksi dari model sesuai dengan data sebenarnya atau justru menyimpang jauh. Semakin mendekati garis lurus diagonal, maka semakin baik kualitas prediksi yang dihasilkan model.

Selain itu, diperlihatkan pula struktur pohon keputusan untuk model *Gradient Boosting Regressor* dan XGBoost sebagai bentuk visualisasi proses pengambilan keputusan dari masing-masing model. Visualisasi ini dilakukan menggunakan fungsi plot tree() yang tersedia dalam pustaka xgboost dan sklearn, yang menggambarkan bagaimana fitur-fitur dibagi di setiap node hingga membentuk keputusan akhir. Untuk model *Random Forest Regressor*, karena terdiri atas ratusan pohon dalam satu *ensemble*, hanya ditampilkan satu pohon sebagai contoh kasus untuk memperlihatkan prinsip kerja model tersebut. Seluruh proses evaluasi dan visualisasi dilakukan dalam satu notebook interaktif yang tersimpan di Google Colab, sehingga keseluruhan eksperimen dapat direplikasi oleh peneliti lain secara mudah dan sistematis.

3. Hasil dan Pembahasan

Penelitian ini membandingkan empat algoritma regresi, yaitu *Support Vector Regression* yang biasa disebut SVR, model Gradient *Boosting* Regression yang dikenal dengan GBR, model Random Forest Regression yang disingkat RFR, serta *Extreme Gradient Boosting* atau XGBoost, dalam memprediksi nilai penjualan berdasarkan indikator sosial ekonomi Kabupaten Cirebon dari tahun 2010 - 2023. Dataset historis dengan variabel prediktor digunakan untuk pelatihan dan pengujian model seperti umur harapan hidup, harapan lama sekolah, rata-rata lama sekolah, pengeluaran per kapita, dan Indeks Pembangunan Manusia (IPM), dengan pengeluaran per kapita yang dijadikan sebagai proksi nilai penjualan.

3.1. Evaluasi Kinerja Model

Evaluasi performa diimplementasikan melalui tiga metrik regresi utama: *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan koefisien determinasi (R^2). Kajian ini membuktikan bahwa ketiga model memiliki tingkat akurasi prediksi cukup baik, namun XGBoost secara konsisten menunjukkan performa terbaik dengan nilai MAE dan MSE paling rendah serta R^2 paling tinggi. Temuan ini selaras dengan penelitian sebelumnya yang menunjukkan keunggulan XGBoost dalam mendeteksi interaksi kompleks dan non-linier antar variabel [14].

Sebaliknya, SVR dengan kernel RBF mencapai performa terbaik di antara kernel lainnya, dengan prediksi yang stabil terutama setelah *tuning* parameter menggunakan Grid Search. Meskipun tidak seakurat RFR pada data training, SVR RBF menunjukkan generalisasi yang lebih baik pada data testing [6]. Random Forest menunjukkan stabilitas yang baik dan resistensi terhadap *overfitting* pada dataset kecil atau bernoise, sedangkan XGBoost menunjukkan performa yang sebanding tetapi lebih rentan terhadap *overfitting* jika tidak dikonfigurasi dengan tepat [2]. Model *Random Forest Regression* menunjukkan kemampuan unggul dalam menangani variabel kompleks dan interdependen, serta terbukti efektif dalam mencegah *overfitting*, dengan hasil evaluasi OOB Score dan R^2 yang sangat tinggi, yaitu 0.9999 [5]. Yang menunjukkan efektivitas RFR dalam konteks prediksi penjualan produk ritel.

Untuk memperkuat pembahasan tersebut, dilakukan evaluasi kuantitatif terhadap semua model menggunakan tiga metrik regresi: MAE, RMSE, dan R^2. Rincian hasil ditampilkan pada Tabel 2.

Tabel 2. Perbandingan Performa Model Regresi dalam Prediksi Penjualan

Model	MAE	RMSE	R^2
SVR (sebelum *tuning*)	845.85	940.31	-0.45
SVR (setelah *tuning*)	191.63	309.00	0.84
Random Forest Regressor	138.93	223.60	0.91
Gradient Boosting Regressor	127.98	185.63	0.94
XGBoost (sebelum *tuning*)	169.24	265.21	0.89
XGBoost (setelah *tuning*)	187.74	278.36	0.88

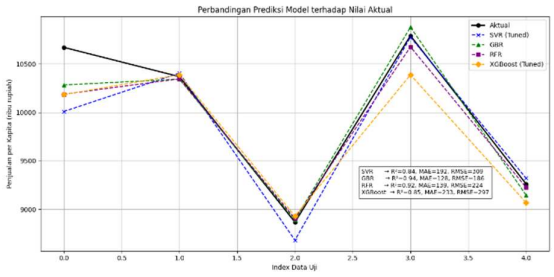
P-ISSN : 2089-3353
E-ISSN : 2808-9162

Hasil evaluasi kuantitatif performa model regresi disajikan dalam Tabel 2, menggunakan tiga metrik evaluasi, yaitu MAE, RMSE, dan R^2. Berdasarkan temuan tersebut, dapat dilihat bahwa model *Support Vector Regression* (SVR) menunjukkan peningkatan performa yang signifikan setelah dilakukan optimasi parameter, dengan nilai R^2 meningkat dari -0.45 menjadi 0.84 dan penurunan RMSE dari 940.31 menjadi 309.00.

Model *Gradient Boosting Regressor* (GBR) mencapai performa unggul secara keseluruhan dengan galat minimum (MAE dan RMSE) dan koefisien determinasi tertinggi, membuktikan kemampuannya dalam memodelkan hubungan non-linier antarvariabel. Disisi lain, *Random Forest Regressor* (RFR) juga menunjukkan performa yang sangat baik dengan prediksi error yang relatif kecil dan R^2 sebesar 0.91. Model XGBoost menunjukkan hasil yang kompetitif, meskipun sedikit penurunan performa terjadi setelah proses *tuning*. Hal ini menunjukkan bahwa konfigurasi awal model sudah cukup optimal, atau proses *tuning* belum berhasil menemukan parameter yang lebih efektif.

3.2. Visualisasi Prediksi

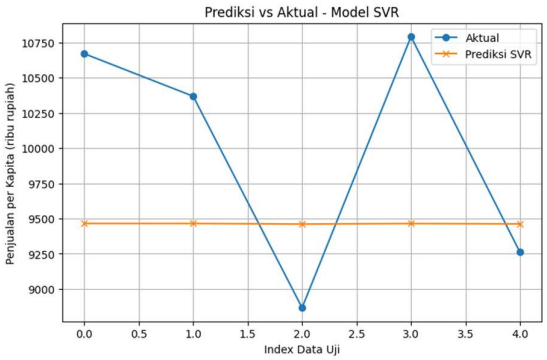
Hasil prediksi yang divisualisasikan menunjukkan bahwa XGBoost memiliki kurva yang paling dekat dengan nilai aktual, sementara SVR sedikit lebih menyimpang, tetapi masih dalam batas yang dapat diterima. RFR dan GBR menunjukkan fluktuasi prediksi yang lebih signifikan terhadap data uji, khususnya pada periode dengan perubahan drastis dalam indikator ekonomi. Grafik perbandingan antara hasil prediksi dan nilai aktual untuk memperjelas performa setiap model ditampilkan pada Gambar 1.



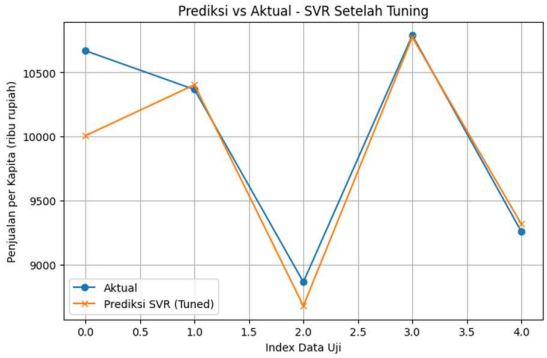
Gambar 1. Grafik perbandingan hasil prediksi empat model regresi terhadap nilai actual

Berdasarkan Gambar 1, dapat dilihat bahwa model *Gradient Boosting* dan XGBoost menunjukkan tren prediksi yang paling sesuai dengan nilai aktual, sedangkan SVR menunjukkan penyimpangan pada beberapa titik, terutama sebelum proses *tuning*. RFR menunjukkan kinerja yang stabil, namun kurang responsif terhadap fluktuasi data ekonomi. Grafik ini memperkuat hasil evaluasi bahwa model berbasis *boosting* cenderung unggul dalam memodelkan hubungan non-linear pada data penjualan.

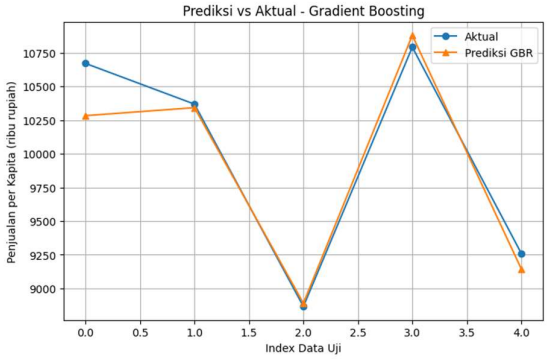
Selain visualisasi komprehensif, Gambar 2 hingga Gambar 7 menampilkan hasil prediksi setiap model secara terpisah. Visualisasi ini memungkinkan analisis lebih mendalam tentang pola kesalahan prediksi dan kesesuaian model dengan nilai aktual secara rinci.



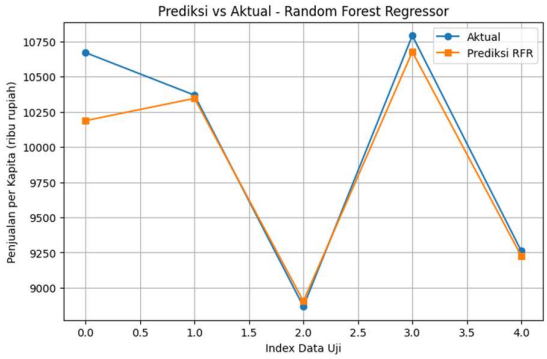
Gambar 2. Grafik prediksi vs aktual model SVR sebelum *tuning*



Gambar 3. Grafik prediksi vs aktual model SVR setelah *tuning*

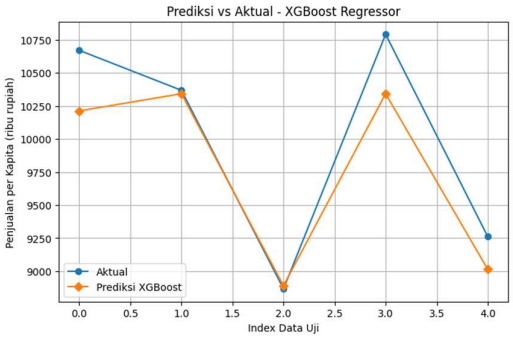


Gambar 4. Grafik prediksi vs aktual model *Gradient Boosting Regressor*

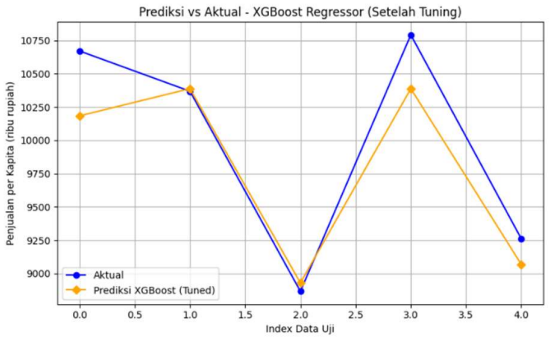


P-ISSN : 2089-3353
E-ISSN : 2808-9162

Gambar 5. Grafik prediksi vs aktual model *Random Forest Regressor*



Gambar 6. Grafik prediksi vs aktual model XGBoost sebelum *tuning*



Gambar 7. Grafik prediksi vs aktual model XGBoost setelah *tuning*

Berdasarkan visualisasi yang ditampilkan pada Gambar 2 hingga Gambar 7, terlihat perbedaan mencolok dalam pola prediksi setiap model regresi. Pada Gambar 2, model *Support Vector Regression* (SVR) sebelum dilakukan *tuning* menunjukkan penyimpangan yang cukup besar terhadap nilai aktual. Prediksi model tampak tidak mengikuti tren data sebenarnya, utamanya pada periode tertentu dengan lonjakan atau penurunan nilai yang signifikan, sehingga hasilnya mengindikasikan bahwa parameter awal SVR belum mampu menangkap hubungan kompleks antar variabel prediktor secara optimal.

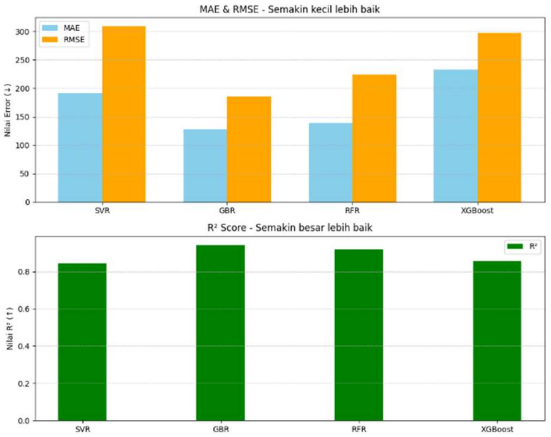
Setelah dilakukan *tuning* (Gambar 3), performa SVR meningkat secara visual. Pola prediksi menjadi lebih selaras dengan nilai aktual, meskipun terdapat beberapa titik penyimpangan. Hal ini menunjukkan bahwa *tuning* parameter seperti C dan epsilon melalui GridSearchCV berhasil meningkatkan kemampuan generalisasi model terhadap data uji, meskipun SVR tetap memiliki keterbatasan dalam menyesuaikan diri terhadap fluktuasi tajam.

Gambar 4 memperlihatkan hasil prediksi dari model *Gradient Boosting Regressor* (GBR). Dari visualisasi, terlihat bahwa GBR sangat efektif dalam memodelkan pola tren dan mengikuti kurva aktual secara halus. Model ini terlihat sangat adaptif terhadap perubahan nilai yang cepat, seperti lonjakan atau penurunan drastis pada tahun tertentu. Hal ini selaras dengan karakteristik algoritma *boosting* yang secara bertahap memperbaiki kesalahan prediksi sebelumnya.

Pada Gambar 5, model *Random Forest Regressor* (RFR) menampilkan prediksi cukup stabil, namun dengan kecenderungan untuk meratakan nilai prediksi. Meskipun akurat pada nilai tengah, RFR tampak kurang responsif terhadap lonjakan maupun penurunan tajam pada data. Hal ini mencerminkan kecenderungan *ensemble bagging* untuk menghasilkan prediksi yang lebih konservatif dan merata.

Model XGBoost sebelum *tuning* (Gambar 6) sudah menunjukkan hasil prediksi yang baik dan cukup mendekati nilai aktual. Pola prediksi relatif stabil dan mengikuti arah tren data. Namun, setelah dilakukan *tuning* (Gambar 7), peningkatannya tidak terlalu signifikan secara visual. Justru terlihat bahwa pada beberapa titik, model menunjukkan gejala *overfitting* atau kurang akurat. Ini menunjukkan bahwa *tuning* harus dilakukan secara hati-hati karena tidak selalu membawa perbaikan signifikan, terutama jika konfigurasi awal model sudah cukup baik.

Selain visualisasi hasil prediksi terhadap nilai aktual, penting pula untuk melihat performa setiap model secara kuantitatif berdasarkan metrik evaluasi. Untuk memberikan gambaran yang lebih intuitif, dilakukan visualisasi nilai koefisien determinasi (R^2) dalam bentuk bar chart yang ditampilkan pada Gambar 8.



Gambar 8. Bar Chart Evaluasi Model

Gambar 8 memperlihatkan perbandingan nilai MAE, RMSE, dan R^2 dari empat model regresi: SVR, GBR, RFR, dan XGBoost. Terlihat bahwa GBR mencatatkan performa terbaik dengan nilai R^2 tertinggi dan MAE serta RMSE terendah, diikuti oleh RFR dan XGBoost. *Support Vector Regression* (SVR), meskipun performanya meningkat setelah *tuning*, masih mencatatkan akurasi yang lebih rendah dibandingkan model berbasis *ensemble*.

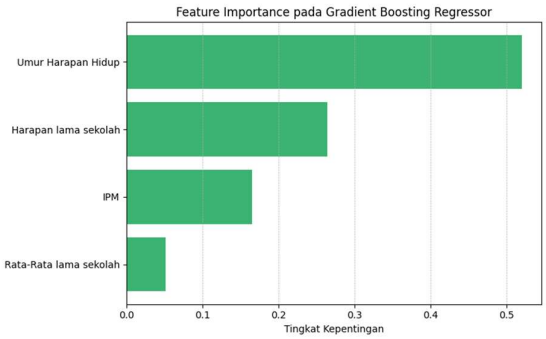
Berdasarkan nilai R^2, model *boosting* seperti GBR dan XGBoost menunjukkan performa yang unggul, menguatkan temuan bahwa model ini lebih sensitif terhadap kompleksitas antar fitur dalam data penjualan. Visualisasi ini memperkuat hasil evaluasi numerik yang sebelumnya disajikan dalam tabel, dan memberikan representasi intuitif terhadap efektivitas masing-masing

P-ISSN : 2089-3353
E-ISSN : 2808-9162

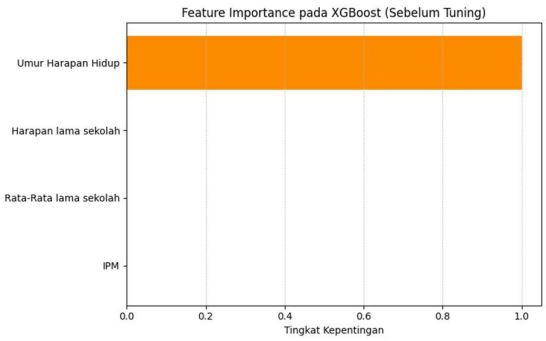
model dalam prediksi penjualan. Setelah dilakukan *tuning* (Gambar 3), performa SVR meningkat secara visual. Pola prediksi menjadi lebih selaras dengan nilai aktual, meskipun masih terdapat beberapa penyimpangan. Hal ini menunjukkan bahwa *tuning* parameter seperti C dan epsilon melalui GridSearchCV berhasil meningkatkan kemampuan generalisasi model terhadap data uji, meskipun SVR tetap memiliki keterbatasan dalam menyesuaikan diri terhadap fluktuasi tajam.

3.3. Analisis Pentingnya Fitur Prediktor

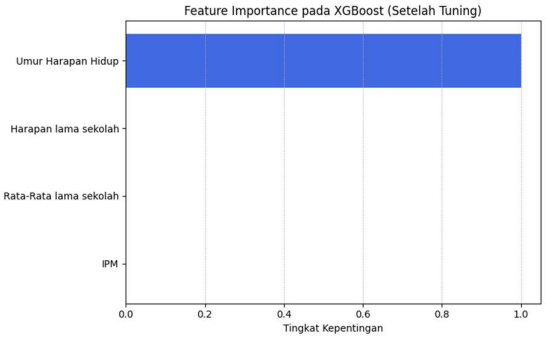
Setelah dilakukan evaluasi performa model dari sisi akurasi prediksi, analisis dilanjutkan dengan meninjau kontribusi setiap fitur input terhadap hasil prediksi. *Feature importance* digunakan untuk mengidentifikasi fitur mana yang paling berpengaruh terhadap model, sehingga memberikan wawasan lebih dalam tentang hubungan antara indikator sosial ekonomi dengan penjualan yang diprediksi. Visualisasi pada Gambar 9 hingga Gambar 11 menampilkan urutan kepentingan fitur berdasarkan dua model unggulan, yaitu *Gradient Boosting Regressor* dan XGBoost (sebelum dan sesudah *tuning*).



Gambar 9. Feature importance pada model *Gradient Boosting Regressor*



Gambar 10. Feature importance pada model XGBoost sebelum *tuning*



Gambar 11. Feature importance pada model XGBoost setelah *tuning*

Berdasarkan Gambar 9 hingga Gambar 11, fitur Umur Harapan Hidup secara konsisten muncul sebagai variabel paling berpengaruh dalam ketiga model. Pada model *Gradient Boosting Regressor* (Gambar 9), Umur Harapan Hidup memiliki tingkat kepentingan tertinggi, diikuti oleh Harapan Lama Sekolah, IPM, serta Rata-rata Lama Sekolah. Hasil tersebut menegaskan bahwa kontribusi terbesar terhadap prediksi penjualan berasal dari indikator yang menggambarkan kondisi kesehatan masyarakat.

Di sisi lain, pada model XGBoost sebelum *tuning* (Gambar 10) dan sesudah *tuning* (Gambar 11) menunjukkan bahwa hanya fitur Umur Harapan Hidup yang digunakan dalam pembentukan pohon keputusan, sebagaimana ditunjukkan oleh nilai F-score yang sangat tinggi dan dominan. F-score dalam XGBoost mengindikasikan seberapa sering sebuah fitur digunakan untuk membagi node pada pohon yang dibentuk selama pelatihan. Nilai F-score yang sangat tinggi untuk satu fitur menandakan bahwa fitur tersebut dianggap jauh lebih informatif dibandingkan fitur lainnya. Hal ini menyebabkan fitur lain seperti IPM atau Pengeluaran Per Kapita tidak muncul dalam hasil visualisasi, karena kontribusinya dianggap kecil atau bahkan nol oleh algoritma selama proses pelatihan.

Temuan ini mengindikasikan bahwa variabel Umur Harapan Hidup tidak hanya berkorelasi tinggi dengan data target (penjualan), tetapi juga secara statistik dianggap sebagai satu-satunya fitur yang membawa informasi penting oleh model XGBoost. Meskipun hal ini memperkuat kemampuan prediksi model, namun berisiko mengabaikan peran variabel lain yang mungkin memiliki dampak kontekstual atau jangka panjang. Oleh karena itu, interpretasi model harus dilakukan dengan hati-hati, terutama ketika fitur yang dianggap "tidak penting" oleh algoritma memiliki nilai strategis dalam pengambilan keputusan.

Hasil ini menekankan perlunya memahami bukan hanya kemampuan prediksi model, tetapi juga bagaimana model menginterpretasikan struktur data, khususnya dalam konteks indikator sosial ekonomi yang saling terkait.

3.4. Implikasi Temuan

Temuan dari penelitian ini memberikan sejumlah implikasi penting dalam konteks pengembangan model prediksi penjualan berbasis data sosial ekonomi regional. Model *Gradient Boosting Regressor* (GBR) secara konsisten menunjukkan performa paling unggul dibandingkan tiga model lainnya (SVR, RFR, dan XGBoost), baik dilihat dari hasil MAE, RMSE, maupun R^2. Hal ini menunjukkan bahwa GBR memiliki keunggulan dalam mengidentifikasi pola hubungan non-linier dan menangani kompleksitas antar fitur prediktor secara lebih akurat. Temuan ini selaras dengan karakteristik pendekatan *boosting* yang mengoptimalkan hasil dengan memfokuskan pada prediksi yang keliru dari iterasi sebelumnya, sehingga menghasilkan output yang lebih presisi.

Implikasinya, GBR dapat direkomendasikan sebagai pendekatan yang efektif dalam peramalan penjualan berbasis indikator sosial ekonomi, khususnya di wilayah yang memiliki keterbatasan jumlah data historis namun dengan variabel yang kompleks seperti Kabupaten Cirebon. Model ini dapat digunakan oleh pemerintah daerah atau pelaku usaha sebagai alat bantu dalam perencanaan strategi ekonomi, misalnya untuk merumuskan estimasi permintaan barang di masa depan, mengatur distribusi logistik, atau menentukan prioritas program intervensi sosial yang dapat mendorong pertumbuhan konsumsi masyarakat.

Lebih lanjut, analisis *feature importance* dalam penelitian ini memperlihatkan bahwa variabel Umur Harapan Hidup (UHH) memiliki pengaruh yang signifikan dalam membentuk prediksi penjualan. Temuan ini mempertegas keterkaitan antara indikator kesehatan dengan perilaku konsumsi. Dalam konteks Kabupaten Cirebon, meningkatnya usia harapan hidup dapat diasosiasikan dengan peningkatan kualitas hidup, stabilitas ekonomi rumah tangga, serta peningkatan kapasitas konsumsi. Dengan demikian, perbaikan pada sektor kesehatan masyarakat secara tidak langsung berpotensi meningkatan aktivitas ekonomi, yang tercermin dalam nilai penjualan agregat tahunan yang signifikan.

Model lain seperti *Random Forest Regressor* (RFR) juga menunjukkan performa prediksi yang tinggi, meskipun cenderung menghasilkan prediksi yang lebih konservatif dan kurang adaptif terhadap lonjakan tajam dalam data. Namun demikian, kestabilan dan ketahanannya terhadap overfitting membuat RFR tetap layak dipertimbangkan sebagai baseline model dalam konteks yang serupa. Sebaliknya, model *Support Vector Regression* (SVR) awalnya menunjukkan performa yang buruk, namun kemudian mengalami peningkatan drastis setelah dilakukan *tuning* parameter. Hal ini menunjukkan bahwa model yang awalnya berkinerja rendah tidak boleh langsung diabaikan, tetapi perlu dilakukan optimisasi parameter untuk meningkatkan kinerjanya.

Hal menarik juga ditemukan pada XGBoost, di mana proses *tuning* justru menurunkan performa model. Hal ini memberikan pelajaran metodologis penting, bahwa *tuning* parameter tidak selalu memberikan hasil yang lebih baik. Bahkan, jika tidak dilakukan dengan cermat atau tanpa mempertimbangkan karakteristik data, *tuning* dapat memperburuk performa model akibat overfitting, underfitting, atau hilangnya keunggulan konfigurasi awal. Oleh karena itu, *tuning* perlu dilakukan secara sistematis, dengan pendekatan eksperimental berbasis validasi silang dan interpretasi performa multi-metrik.

Secara keseluruhan, temuan penelitian ini memberikan kontribusi yang bermakna dalam pengembangan model prediksi penjualan berbasis indikator sosial ekonomi lokal. Hasil evaluasi dan visualisasi mendukung kesimpulan bahwa model *ensemble boosting* seperti GBR cocok digunakan untuk prediksi berbasis data tahunan, sementara faktor-faktor seperti kualitas hidup dan pendidikan memiliki peran yang sentral dalam menentukan proyeksi penjualan masa depan. Dengan menerapkan model dan fitur yang relevan, pemangku kebijakan dapat menyusun kebijakan pembangunan ekonomi yang lebih presisi, terukur, dan berbasis bukti (*data-driven policy*).

3.5. Keterbatasan dan Rekomendasi

Keterbatasan pada jumlah observasi tahunan (n=14) menjadi faktor yang perlu dipertimbangkan dalam menggeneralisasi hasil penelitian ini. Selain itu, data hanya mencakup wilayah Kabupaten Cirebon, sehingga penerapan model pada konteks geografis lain memerlukan penyesuaian. Sebagai rekomendasi, penelitian selanjutnya dapat menerapkan teknik optimisasi lanjutan seperti Particle Swarm Optimization (PSO) untuk *tuning* parameter SVR sebagaimana direkomendasikan oleh studi sebelumnya [9], serta memperluas cakupan data dan variabel prediktor untuk validasi model lintas wilayah

4. Kesimpulan

Penelitian ini bertujuan untuk membandingkan kinerja empat algoritma regresi, yaitu *Support Vector Regression* yang dikenal dengan singkatan SVR, Gradient *Boosting* Regression yang biasa disebut GBR, Random Forest Regression yang disingkat RFR, serta XGBoost dalam memprediksi nilai penjualan agregat tahunan berdasarkan indikator sosial ekonomi di Kabupaten Cirebon periode 2010–2023. Hasil evaluasi menunjukkan bahwa *Gradient Boosting Regressor* (GBR) merupakan model dengan performa unggul, ditunjukkan oleh nilai kesalahan prediksi terendah dan koefisien determinasi tertinggi. Model *Random Forest Regressor* (RFR) menempati posisi kedua dengan performa stabil dan akurasi tinggi, sementara SVR mengalami peningkatan signifikan setelah *tuning*. Sebaliknya, *tuning* pada XGBoost tidak menghasilkan peningkatan performa yang berarti temuan ini menunjukkan bahwa model berbasis *ensemble*, khususnya yang menerapkan metode *boosting*, sangat cocok untuk digunakan dalam memodelkan hubungan

kompleks antar variabel sosial ekonomi dan penjualan. Model tersebut berpotensi diterapkan dalam sistem pendukung keputusan untuk perencanaan logistik, strategi distribusi, dan perumusan kebijakan ekonomi berbasis data di tingkat daerah. Selain itu, variabel Umur Harapan Hidup terbukti memiliki kontribusi prediktif tertinggi, yang mengindikasikan bahwa peningkatan kualitas hidup masyarakat dapat berdampak positif terhadap aktivitas ekonomi.

Penelitian selanjutnya disarankan untuk memperluas jumlah observasi dengan memanfaatkan data yang lebih luas dan bersifat representatif, baik dari segi kuantitas sampel maupun keberagaman wilayah geografis, guna menguji kinerja model pada berbagai konteks dan memperkuat kemampuan generalisasinya. Jumlah observasi yang ideal seharusnya mampu merepresentasikan variasi karakteristik data secara memadai agar pola hubungan antar variabel dapat dipelajari secara konsisten dan stabil. Jika perluasan jumlah observasi tidak dilakukan, model berisiko memiliki daya generalisasi yang terbatas, rentan terhadap overfitting, serta mengalami penurunan tingkat akurasi saat diterapkan pada data atau kondisi wilayah yang berbeda. Di samping itu, penerapan metode optimisasi parameter yang lebih adaptif, seperti Particle Swarm Optimization dan Bayesian Optimization, dapat dipertimbangkan untuk meningkatkan performa model, terutama pada algoritma yang peka terhadap pengaturan parameter awal.

Daftar Rujukan

[1] N. D. Maulana, B. D. Setiawan, and C. Dewi, "Implementasi Metode *Support Vector Regression* (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus: Harum Bakery)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2986–2995, 2019, [Online]. Available: http://j-ptiik.ub.ac.id

[2] Y. M. Nurak, S. Wahyu Irianandа, & F. Marisa, "Prediksi penjualan Warung Kopi OI menggunakan metode *Random Forest* dan *XGBoost*," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 5796–5802, Aug. 2025, doi: 10.36040/jati.v9i4.13922.

[3] Z. Aini, "Implementasi Random Forest Dan Gradient *Boosting* Pada Klasifikasi Indeks Pembangunan Manusia (IPM)," *Skripsi*, p. 90, 2023, [Online]. Available: https://repository.uinjkt.ac.id/dspace/handle/123456789/74286

[4] A. Dahlan and R. Anggara, "Pengaruh Faktor Pendapatan, Konsumsi/Pengeluaran Rumah Tangga, Dan Tabungan Terhadap Tingkat Kesejahteraan Buruh Pada Industri Batu Alam Di Desa Bobos Dukupuntang Cirebon," *Syirkatuna J. Ekon. Islam*, vol. 5, no. 1, pp. 1–14, 2017, [Online]. Available: https://ejournal.steialishlah.ac.id/index.php/syirkatuna/article/view/21

[5] R. Hidayat *et al.*, "Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket," *Simkom*, vol. 10, no. 1, pp. 101–109, 2025, doi: 10.51717/simkom.v10i1.703.

[6] N. A. A. Z. Tualeka AC, R. M. Atok, and A. U. Alfajriyah, "Perbandingan Metode Random Forest Regression (RFR) dan *Support Vector Regression* (SVR) dalam Memprediksi Risiko Kredit pada Bank XYZ," *J. Sains dan Seni ITS*, vol. 13, no. 6, 2025, doi: 10.12962/j23373520.v13i6.150012.

[7] S. V. Hutagalung, Y. Yennimar, E. R. Rumapea, M. J. G. Hia, T. Sembiring, and D. R. Manday, "Comparison of *Support Vector Regression* and Random Forest Regression Algorithms on Gold Price Predictions," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 7, no. 1, pp. 255–262, 2023, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v7i1.4125.

[8] F. E. Penalun, A. Hermawan, and D. Avianto, "Perbandingan Random Forest Regression dan *Support Vector Regression* Pada Prediksi Laju Penguapan," *J. Fasilkom*, vol. 13, no. 02, pp. 104–111, 2023, doi: 10.37859/jf.v13i02.4976.

[9] F. Yulianto, W. F. Mahmudy, and A. A. Soebroto, "Comparison of Regression, *Support Vector Regression* (SVR), and SVR-Particle Swarm Optimization (PSO) for Rainfall Forecasting," *J. Inf. Technol. Comput. Sci.*, vol. 5, no. 3, pp. 235–247, 2020, doi: 10.25126/jitecs.20205374.

[10] A. N. M. Pudjianto and E. Y. Hidayat, "Perbandingan Prediksi Depresi Mahasiswa dengan Linear Regression, Random Forest, dan Gradient *Boosting*," *SINTECH (Science Inf. Technol. J.*, vol. 7, no. 3, pp. 180–189, 2024, doi: 10.31598/sintechjournal.v7i3.1729.

[11] I. Palupi, B. ari Wahyudi, N. AL Mamuda, and A. Shabrina, "Predicting Forest Fire Hotspots with Carbon Emission Insights Using Random Forest and Gradient *Boosting* Regression," *Int. J. Inf. Commun. Technol.*, vol. 9, no. 2, pp. 137–149, 2023, doi: 10.21108/ijoict.v9i2.865.

[12] K. C. Liao, H. Y. Wu, H. T. Wen, J. T. Sung, M. Hidayat, and W. W. J. Wang, "Compressor performance prediction: gradient *boosting* regression model and sensitivity analysis," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 37, no. 2, pp. 1201–1208, 2025, doi: 10.11591/ijeecs.v37.i2.pp1201-1208.

[13] A. N. Rachmi, "Implementasi Metode Random Forest Dan Xgboost Pada Klasifikasi Customer Churn," *Univ. Islam Indones.*, vol., no., pp. 1–101, 2020.

[14] A. Syahreza, N. K. Ningrum, and M. A. Syahrazy, "Perbandingan Kinerja Model Prediksi Cuaca: Random Forest, *Support Vector Regression*, dan XGBoost," *Edumatic J. Pendidik. Inform.*, vol. 8, no. 2, pp. 526–534, 2024, doi: 10.29408/edumatic.v8i2.27640.

[15] I. Maulita and A. M. Wahid, "Prediksi Magnitudo Gempa Menggunakan Random Forest, *Support Vector Regression*, XGBoost, LightGBM, dan Multi-Layer Perceptron Berdasarkan Data Kedalaman dan Geolokasi," *J. Pendidik. dan Teknol. Indones.*, vol. 4, no. 5, pp. 221–232, 2024, doi: 10.52436/1.jpti.470.

[16] R. Hidayat, D. Mahdiana, and A. Fergina, "Comparative Analysis of Logistic Regression, SVM, Xgboost, and Random Forest Algorithms for Diabetes Classification," *J. Teknol. Sist. Inf. dan Apl.*, vol. 7, no. 1, pp. 281–291, 2024, doi: 10.32493/jtsi.v7i1.38258.

[17] E. Banjarnahor, R. Belferik, W. Cendana, Y. Adi, and S. Abraham, "Analisis Implementasi Support Vector Machine dan Random Forest untuk Prediksi Kategori Indeks Kualitas Udara Jakarta under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)," vol. 10, no. 1, pp. 2541–1179, 2025, doi: 10.24252/instek.v10i1.56477.

[18] S. Papadogiannaki, S. Kontos, D. Parliari, and D. Melas, "Machine Learning Regression to Predict Pollen Concentrations of Oleaceae and Quercus Taxa in Thessaloniki, Greece," p. 2, 2023, doi: 10.3390/environsciproc2023026002.