

PENERAPAN METODE REGRESI DALAM ANALISIS TINGKAT KONSUMSI IKAN DI JAWA TIMUR

Hikmal Akbar Fitrianyah¹, Mahathir Muhamad Ashari², Naufal Rahaditya Aryadi³, Willdan Aprizal Arifin⁴

^{1,2,3,4}Sistem Informasi Kelautan, Universitas Pendidikan Indonesia
Kampus Daerah Serang

¹hikmal2601@upi.edu, ²mahathirashari@upi.edu, ³naufalrahaditya@upi.edu, ⁴willdanarifin@upi.edu

ABSTRACT

The Indonesian marine ecosystem, rich in water resources, makes the fishing sector important for the economy, food, and jobs. East Java, with the highest catch production in Indonesia, has huge potential in the fishing sector. Despite its high potential, fish consumption in Indonesia is still relatively low due to lack of awareness, unoptimal distribution, and other factors. The study aims to analyze the level of fish consumption in Eastern Java and determine the best regression method to predict the rate of fish intake based on the region and the types of available commodities. This study uses three regression methods, namely Linear Regression, Support Vector Regression (SVR), and Gradient Boosting Machine (GBM). Data visualization is done using bar diagrams, and results are validated using the determination coefficient R^2 which is then analyzed descriptively. The Sumenep region has the highest level of fish consumption in East Java during the period 2018-2020. Whereas the commodities with the highest consumption are Tuna, Tongkol, Cakalang (TTC) Diawetkan. The GBM method showed its best performance and proved to be the most effective and accurate in predicting the level of fish consumption in East Java with a perfect determination coefficient (0,9999), compared to Linear Regression (0,8755) and SVR (0,9825).

Keywords- *Data Mining, Gradient Boosting Machine, Fish Consumption, Regression*

PENDAHULUAN

Ekosistem laut di Indonesia sangat besar dan memiliki potensi untuk dikembangkan lebih lanjut. Sebagai negara kepulauan, Indonesia kaya akan sumber daya perairan dan beragam jenis ikan. Indonesia yang terdiri dari 14.572 pulau, dengan wilayah laut seluas 6.315.222 km² dan garis pantai sepanjang 99.093 km memiliki potensi perikanan yang besar, mencapai 7,3 juta ton ikan per tahun [1]. Jawa Timur merupakan salah satu provinsi dengan potensi sumber daya perikanan yang tinggi, menurut data statistik KKP tahun 2023 provinsi ini memiliki produksi perikanan tangkap tertinggi di Indonesia, dengan total mencapai 534.397 ton.

Sektor perikanan mempunyai tiga peran penting, antara lain sebagai sumber pertumbuhan ekonomi, sumber pangan, dan penyedia lapangan kerja [2][3][4]. Sebagai salah satu sumber pangan, ikan memiliki peran penting dalam kepastian ketersediaan pangan di negara ini. Hal ini dikarenakan ikan memiliki protein dan gizi yang berkualitas tinggi untuk kesehatan tubuh manusia [5]. Konsumsi ikan merupakan salah satu faktor yang dapat mempengaruhi status gizi seseorang, keluarga atau masyarakat [6]. Setiap individu memiliki pola konsumsi yang bervariasi, namun secara umum, orang akan memprioritaskan pemenuhan kebutuhan pokok sebelum memenuhi kebutuhan lainnya. Tingkat kebutuhan konsumsi setiap orang atau keluarga dipengaruhi oleh pendapatan mereka. Pendapatan seseorang akan berdampak pada tingkat konsumsi mereka, semakin tinggi pendapatan, semakin banyak barang yang dikonsumsi [7].

Meskipun Indonesia memiliki potensi sumber

daya ikan yang besar, tingkat konsumsi ikan di negara ini masih tergolong rendah [8]. Beberapa faktor yang menyebabkan rendahnya konsumsi ikan antara lain kesadaran masyarakat Indonesia yang masih kurang terhadap manfaat dari mengonsumsi ikan, kurang meratanya distribusi ikan dan kurang optimalnya sarana dan prasarana yang ada di Indonesia [9]. Jumlah konsumsi ikan perkapita yang ada di Indonesia dipengaruhi beberapa faktor seperti ekonomi, sosial, dan demografi turut berperan dalam menentukan jumlah konsumsi ikan perkapita [10]. Di antara tujuh pulau besar di Indonesia, Pulau Jawa memiliki tingkat konsumsi ikan terendah [11]. Oleh sebab itu, penelitian ini bertujuan untuk menganalisis tingkat konsumsi ikan di Provinsi Jawa Timur dengan mengidentifikasi wilayah serta komoditas ikan yang memiliki tingkat konsumsi tertinggi hingga terendah. Penelitian ini akan membandingkan tiga metode regresi, yaitu Regresi Linear, *Support Vector Regression* (SVR), dan *Gradient Boosting Machine* (GBM), untuk menentukan metode yang paling efektif dan akurat dalam analisis tingkat konsumsi ikan di Jawa Timur.

METODE PENELITIAN

Dalam penelitian ini, digunakan 3 metode regresi yaitu Regresi Linear, *Support Vector Regression* (SVR), dan *Gradient Boosting Machine* (GBM) untuk menganalisis tingkat konsumsi ikan yang ada di Jawa Timur. Adapun untuk alur dalam penelitian dapat dilihat pada diagram berikut.



Gambar 1. Alur Penelitian

Pengumpulan Data

Data diperoleh dari Badan Pusat Statistik berupa data rata-rata konsumsi per kapita seminggu menurut komoditas ikan per kabupaten/kota di Jawa Timur dari tahun 2018-2020 yang terdiri dari 29 wilayah dan 36 komoditas.

Tabel 1. Dataset Tingkat Konsumsi Ikan di Jawa Timur

No	Wilayah	Jenis Ikan	Tahun	Tingkat Konsumsi
1	Pacitan	Bandeng	2018	0,008
2	Pacitan	Bandeng Diawetkan	2018	0,008
3	Pacitan	Baronang	2018	0,000
4	Pacitan	Bawal	2018	0,000
5	Pacitan	Cumi-Cumi	2018	0,002
...
3128	Sumenep	TTC Diawetkan	2020	0,516
3129	Sumenep	Udang diawetkan	2020	0,006
3130	Sumenep	Udang segar	2020	0,011
3131	Sumenep	Rebon	2020	0,000
3132	Sumenep	Lobster	2020	0,010

Pembersihan Data

Data yang sudah diperoleh selanjutnya akan dilakukan pembersihan dengan menghilangkan variabel yang tidak diperlukan agar data dapat digunakan [12]. Hal ini dilakukan untuk mengantisipasi terjadinya *error* akibat kekosongan data pada saat proses pengolahan data nantinya.

Pengolahan Data Menggunakan Metode Regresi

Terdapat tiga metode regresi yang digunakan dalam analisis tingkat konsumsi ikan di Jawa Timur, yaitu Regresi Linear, Support Vector Regression (SVR), dan Gradient Boosting Machine (GBM). Regresi Linear

adalah algoritma yang dapat digunakan untuk membuat peramalan berdasarkan data sebelumnya dengan menggambarkan hubungan antara variabel terikat dan variabel bebas [13]. Formula matematis dari regresi linear adalah :

$$\hat{y} = \beta_0 + \beta_1 x \quad (1)$$

di mana :

\hat{y} : prediksi tingkat konsumsi, β_0 adalah intersep, dan β_1 : koefisien regresi.

Support Vector Regression (SVR) adalah pengembangan algoritma dari metode SVM yang digunakan untuk membuat prediksi dengan tujuan mencari sebuah fungsi $\square(\square)$ dan dapat meminimalkan kesalahan prediksi dalam suatu margin toleransi tertentu dalam hal ini SVR berupaya memaksimalkan margin antara data observasi dan hyperplane regresi, dengan harapan meminimalkan kesalahan prediksi [14]. Formula matematis dari SVR dengan kernel RBF adalah:

$$f(x) = \sum_{i=1}^n a_i k(x_i, x) + b \quad (2)$$

di mana :

a_i : koefisien model,
 $k(x_i, x)$: fungsi kernel yang mengukur kesamaan antara titik data, dan b adalah bias atau intersep.

Gradient Boosting Machine (GBM) merupakan algoritma yang menggabungkan prediksi dari beberapa decision tree secara bertahap untuk meningkatkan akurasi mode [15]. Formula matematis dari GBM secara umum dapat dinyatakan sebagai :

$$F_m(x) = F_{m-1}(x) + \lambda h_m(x) \quad (3)$$

di mana

$F_m(x)$: model pada iterasi ke-m,
 $F_{m-1}(x)$: model sebelumnya,

λ : learning rate,

$h_m(x)$: decision tree yang baru dibangun pada iterasi m.

GBM bertujuan untuk mengurangi kesalahan prediksi secara iteratif dengan menambahkan model-model baru yang fokus pada kesalahan yang masih ada dari model sebelumnya. Metode ini dikembangkan dari *Recurrent Neural Network* (RNN) dengan tujuan agar setiap *recurrent* unit mampu secara adaptif menangkap hubungan (dependensi) dalam berbagai skala waktu [16].

Pada situasi prediksi dapat dipastikan mengandung tingkat ketidakpastian yang disebut dengan *error* yang mana sumber penyimpangan pada prediksi memiliki banyak unsur, salah satunya tidak mampunya suatu model dalam mengenali unsur lain

dalam deret data mempengaruhi besarnya penyimpangan dalam prediksi [17]. Untuk mengevaluasi akurasi dan prediksi kinerja model regresi yang berbeda digunakan dua indeks evaluasi:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

Visualisasi Data

Data divisualisasikan menggunakan diagram batang untuk menunjukkan tingkat konsumsi ikan berdasarkan komoditas dan wilayah di Jawa Timur serta untuk melihat perbandingan dari ketiga metode regresi.

Validasi Hasil

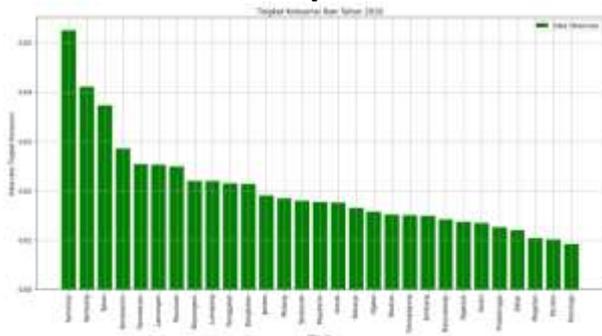
Pada tahap ini dilakukan validasi serta pengujian dari ketiga metode regresi berdasarkan koefisien determinasi R² yang diperoleh.

Analisis Hasil Pengolahan Data

Analisis dilakukan secara deskriptif dengan membandingkan hasil data diagram yang menunjukkan kenaikan atau penurunan tingkat konsumsi ikan berdasarkan wilayah dan komoditas setiap tahunnya di Jawa Timur serta hasil koefisien determinasi R² yang didapat dari ketiga metode regresi.

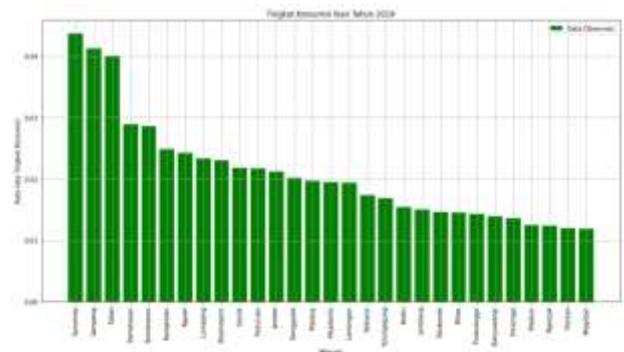
HASIL DAN PEMBAHASAN

Analisis Berdasarkan Wilayah



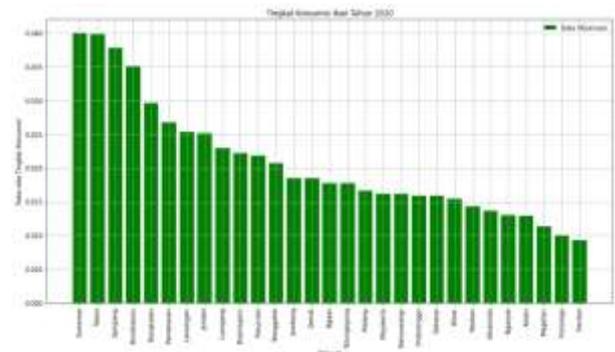
Gambar 2. Diagram Tingkat Konsumsi Ikan Berdasarkan Wilayah di Jawa Timur tahun 2018

Pada tahun 2018, Sumenep menempati posisi sebagai wilayah yang memiliki tingkat konsumsi ikan tertinggi di Jawa Timur dengan indeks lebih dari 0.05 diikuti Sampang dan Tuban yang berada cukup jauh di bawah Sumenep. Sedangkan wilayah dengan tingkat konsumsi ikan terendah tahun 2018 di Jawa Timur adalah Ponorogo.



Gambar 3. Diagram Tingkat Konsumsi Ikan Berdasarkan Wilayah di Jawa Timur tahun 2019

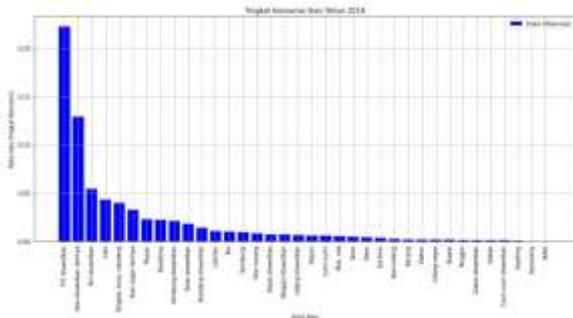
Pada tahun 2019, Sumenep sedikit mengalami penurunan indeks menjadi berkisar 0.04 akan tetapi Sumenep masih menjadi wilayah dengan tingkat konsumsi ikan tertinggi di Jawa Timur diikuti Sampang serta Tuban yang mengalami peningkatan indeks mencapai 0.04 yang membuat jarak ketiganya tidak terlalu jauh. Sedangkan untuk wilayah dengan tingkat konsumsi ikan terendah di tahun 2019 ditempati oleh Magetan. Lalu Ponorogo yang pada tahun 2018 menjadi wilayah dengan tingkat konsumsi ikan terendah telah mengalami peningkatan di tahun 2019 dengan indeks lebih dari 0.01.



Gambar 4. Diagram Tingkat Konsumsi Ikan Berdasarkan Wilayah di Jawa Timur tahun 2020

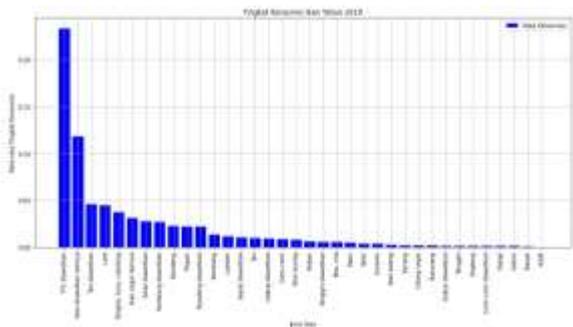
Pada tahun 2020, Sumenep mengalami penurunan indeks yang membuatnya sejajar dengan Tuban sebagai wilayah yang memiliki tingkat konsumsi ikan tertinggi di Jawa Timur diikuti Sampang yang juga mengalami penurunan indeks. Sementara wilayah dengan tingkat konsumsi ikan terendah di Jawa Timur tahun 2020 ditempati oleh Pacitan.

Analisis Berdasarkan Jenis Komoditas Ikan



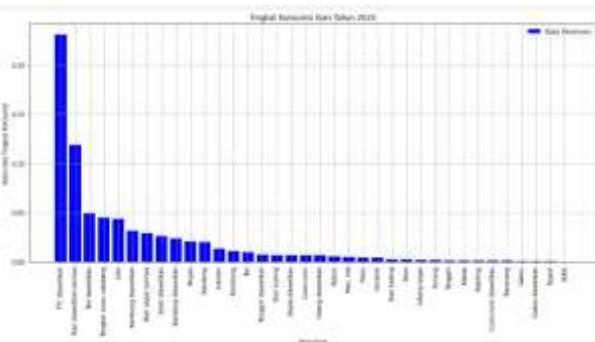
Gambar 5. Diagram Tingkat Konsumsi Ikan Berdasarkan Komoditas di Jawa Timur tahun 2018

Berdasarkan diagram pada tahun 2018, terlihat bahwa TTC (Tuna, Tongkol, Cakalang) Diawetkan memiliki tingkat konsumsi tertinggi di Jawa Timur dengan rata-rata konsumsi di atas 0,20. Kemudian diikuti Ikan Diawetkan Lainnya dan Teri Diawetkan, juga menunjukkan tingkat konsumsi yang tinggi, meskipun lebih rendah dibandingkan dengan TTC Diawetkan. Sedangkan tingkat konsumsi Ikan Baronang menempati urutan terendah pada tahun 2018.



Gambar 6. Diagram Tingkat Konsumsi Ikan Berdasarkan Komoditas di Jawa Timur tahun 2019

Berdasarkan diagram pada tahun 2019, TTC Diawetkan masih berada pada tingkat konsumsi tertinggi di Jawa Timur dan tidak mengalami penurunan maupun kenaikan indeks berdasarkan data grafik 2018. Namun terjadi penurunan indeks sebesar 0.12 untuk Ikan Diawetkan Lainnya dan 0.04 untuk Teri Diawetkan. Sedangkan terjadi perubahan tingkat konsumsi jenis ikan terendah pada tahun 2019 di Jawa Timur menjadi jenis Ikan Bawal.

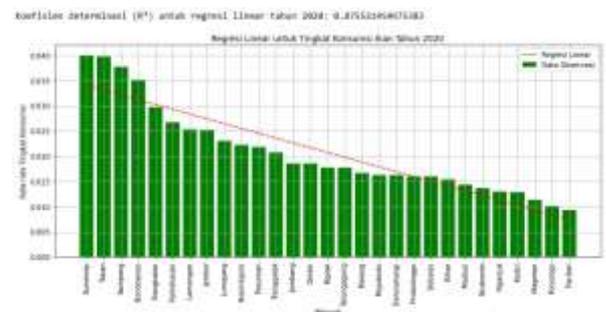


Gambar 7. Diagram Tingkat Konsumsi Ikan Berdasarkan Komoditas di Jawa Timur tahun 2020

Berdasarkan diagram pada tahun 2020, TTC Diawetkan masih berada pada tingkat konsumsi tertinggi di Jawa Timur dan tidak mengalami penurunan maupun kenaikan indeks selama 2 tahun terakhir. Namun terjadi kenaikan indeks sebesar 0.05 untuk Teri Diawetkan. Sedangkan untuk tingkat konsumsi jenis ikan terendah pada tahun 2020 di Jawa Timur tidak mengalami perubahan.

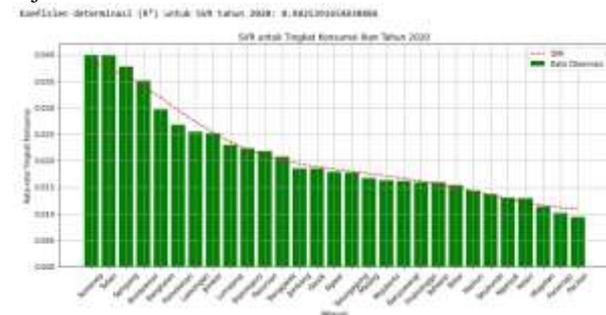
Analisis Metode Regresi

Diagram tingkat konsumsi ikan berdasarkan wilayah di Jawa Timur tahun 2020 dijadikan sampel perbandingan untuk 3 metode regresi yaitu Regresi Linear, Support Vector regression (SVR) dan Gradient Boosting Machine (GBM).



Gambar 8. Diagram Metode Regresi Linear

Pada analisis tiga metode regresi yaitu regresi linear yang dievaluasi berdasarkan koefisien determinasi R^2 pada tahun 2020, menunjukkan hasil bahwa Regresi Linear memberikan koefisien determinasi sebesar 0.8755, hal ini menandakan sekitar 87.55% variasi dalam tingkat konsumsi ikan dapat dijelaskan oleh model ini.



Gambar 9. Diagram Metode Regresi SVR

Sementara itu pada metode SVR yang dievaluasi berdasarkan koefisien determinasi R^2 pada tahun 2020, menunjukkan performa yang lebih baik dengan koefisien determinasi sebesar 0.9825, menunjukkan bahwa model ini mampu menjelaskan sekitar 98.25% variasi dalam data.

Prediksi Penjualan Dan Cashflow Pada Aplikasi Point Of Sales Kafe Xyz," 2021..

- [14] Y. W. Syaifudin, P. Y. Saputra, dan T. Fatmawati, "Prediksi Kuantitas Hasil Budidaya Ikan Konsumsi Menggunakan Penerapan Metode Regresi Data," *INFORMAL: Informatics Journal*, vol. 9, no. 1, pp. 39-48, 2024
- [15] F. A. Nugraha, N. H. Harani, dan R. Habibi, "Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning," Kreatif, 2020.
- [16] L. Lumbaa, "Implementasi Metode SVM dan Gardiant Boost Dalam Kalsifikasi Bahasa Daerah," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 2, pp. 908-915, 2022.
- [17] A. N. Safira, B. Warsito, dan A. Rusgiyono, "Analisis Support Vector Regression (SVR) dengan Algoritma Grid Search Time Series Cross Validation untuk Prediksi Jumlah Kasus Terkonfirmasi COVID-19 di Indonesia," *JURNAL GAUSSIAN*, vol. 11, no. 4, pp. 512-521, 2023..