

## PREDIKSI BERAT IKAN MENGGUNAKAN *MULTIPLE LINEAR REGRESSION*

Anindya Celena Khansa Kirana<sup>1</sup>, Annisa Nuri Nabila<sup>2</sup>, Rafid Ardito<sup>3</sup>, Imanuel Juventius Todo Gurning<sup>4</sup>

Program Studi Teknik Informatika, Jurusan Teknik Informatika

Email: <sup>1</sup>ackkvip@student.ub.ac.id, <sup>2</sup>ennoza@student.ub.ac.id, <sup>3</sup>vaxeria@student.ub.ac.id,

<sup>4</sup>imanuel@student.ub.ac.id

### Abstrak

Ikan merupakan komoditas yang diminati masyarakat sebagai sumber pangan dan nutrisi guna mencukupi kebutuhan gizi. Pembudidayaan ikan banyak dilakukan oleh masyarakat karena relatif mudah untuk dipelajari. Penerapan sains data dapat membantu pelaku usaha budidaya ikan dalam mengefisienkan waktu dan biaya produksi, salah satunya dengan cara menerapkan prediksi berat ikan untuk menentukan ikan siap panen. Penelitian ini menerapkan prosedur sains data dalam mengukur berat ikan dari tujuh jenis yang umum ditemukan di pasaran, dan menggunakan *multiple linear regression* sebagai metode prediksi. Metode regresi linear digunakan untuk memprediksi berat ikan berdasarkan fitur panjang vertikal, panjang diagonal, panjang silang, lebar, dan tinggi ikan melalui proses pelatihan model. Penelitian ini membandingkan model dengan menyertakan data *outlier* dan model dengan menghilangkan data *outlier*. Hasil dari perbandingan tersebut adalah nilai *root mean squared error* (RMSE) dari model dengan menghilangkan data *outlier* lebih kecil dan nilai R2 Square dari model lebih besar daripada model dengan menyertakan data *outlier*, dimana nilai RMSE pada data *train* adalah 100.64 dan pada data *test* adalah 89.22, sedangkan nilai R2 Square pada data *train* adalah 0.89 dan pada data *test* adalah 0.92.

**Kata kunci:** ikan, sains data, prediksi, *multiple linear regression*

---

### 1. PENDAHULUAN

Proses pemenuhan ketersediaan pangan secara mudah bagi masyarakat adalah syarat penting guna mencapai ketahanan pangan nasional. Ketersediaan pangan bagi warga negara sangat penting karena aspek tersebut merupakan salah satu faktor yang mempengaruhi stabilitas nasional. Salah satu unsur pemenuhan ketahanan pangan tersebut ialah Kemandirian Pangan ketika masyarakat dapat memproduksi sumber pangan secara mandiri (Bulog, 2014). Kemandirian pangan sebagai sumber pemenuhan gizi masyarakat merupakan aspek penting sebagai salah satu indikator keberhasilan pengembangan kualitas sumber daya manusia. Salah satu sumber gizi protein hewani adalah melalui pangan ikan, dengan beberapa keunggulan terhadap sumber protein hewani lainnya, seperti kandungan asam amino esensial, omega-3 guna meningkatkan kecerdasan anak, dan berbagai macam antioksidan (Jaya, 2012).

Beberapa metode budidaya perikanan dapat dilakukan, diantaranya adalah budidaya akuaponik, kandang, dan komposit. (Irawan, 2017). Metode budidaya ikan populer yang relatif mudah untuk dilakukan secara mandiri adalah budidaya ikan dalam ember atau budikamber, yang menggunakan media air untuk menghasilkan ikan dan sayuran dengan lahan terbatas. Metode budikamber dapat dilakukan dengan media ember 60 liter dan menghasilkan lele seberat 2.440 gram dan 42 ikat kangkung dalam

jangka waktu pemeliharaan selama 42 hari (Nursandi, 2018).

Budidaya ikan dalam skala kecil tetap membutuhkan perawatan dan pemantauan kondisi ikan. Rata-rata pertumbuhan panjang ikan dalam jangka waktu 42 hari sebesar  $\pm 7$  cm dan pertumbuhan bobot sebesar 18,45 gram, untuk ukuran tebar 5-12 cm dan 47,74 gram untuk ukuran tebar 12-14 cm (Nursandi, 2018). Proses pemantauan ikan dilakukan untuk mendapatkan berat optimal untuk mendapatkan hasil produksi yang optimal. Penerapan sains data dalam rangka efisiensi produksi ikan menjadi salah satu potensi pengembangan di masa modern, dengan berbagai metode yang dapat dimanfaatkan untuk memudahkan masyarakat. Salah satu metode prediksi berat ikan dapat menggunakan pengukuran panjang dan lebar objek pantau guna memperkirakan berat dengan metode regresi linier. Regresi linier digunakan dalam mendapatkan nilai variabel prediksi, dalam hal ini berat ikan, dengan menggunakan variabel prediktor dalam bentuk panjang, tinggi, dan lebar ikan.

Prediksi berat suatu objek dengan menggunakan pengukuran dimensi panjang dan tinggi objek tersebut telah dilakukan sebelumnya. Penelitian oleh (Firdaus & Hasan, 2019) memprediksi berat gedung dengan pengukuran panjang, lebar, dan tinggi gedung menggunakan aplikasi SAP 2000 serta *Artificial Neural Network* (ANN) dengan metode *Back Propagation*. Hasil prediksi ANN mendapatkan error tertinggi sebesar 38% dan terkecil 0,02%. Penelitian lain dilakukan oleh (Fitriyah, Setyawan & Masruri,

## 2 Proyek Akhir Mata Kuliah Pengantar Sains Data

2020) memprediksi berat buah *non axi-symmetric* dengan metode regresi linier. Prediksi berat didapatkan dengan pemrosesan gambar buah dan melakukan pengukuran panjang, lebar, dan tinggi dari objek buah berdasarkan pemrosesan pixel gambar, untuk kemudian menggunakan regresi linier dalam mendapatkan nilai prediksi berat buah. Hasil penelitian mendapatkan rata-rata error sebesar 7,089% dan rata-rata perbedaan perkiraan berat dan berat asli sebesar 16,322 gram. Penelitian dalam bidang perikanan pernah dilakukan oleh (Chen, Yang, Sun, Wang, Xu & Zhou, 2020) mengukur panjang ikan untuk memprediksi berat dan menentukan jumlah pakan optimal yang diperlukan. Penelitian menggunakan *Mind Evolutionary Algorithm* (MEA) untuk menentukan jumlah pakan dan *Back Propagation* untuk memprediksi berat ikan. Hasil penelitian menunjukkan tingkat error sebesar 6,17% untuk penentuan berat dan 6,89 untuk prediksi pakan optimal. Penelitian lain dalam bidang perikanan oleh (Fitriyah, 2020) dengan mengukur berat ikan dan sayuran secara otomatis pada budidaya ikan dalam ember menggunakan visi komputer dan metode regresi linier. Penelitian ini mengambil gambar ikan dan sayuran untuk diproses dengan visi komputer untuk mendapatkan fitur pada tanaman sayur dan ikan untuk segmentasi, kemudian menghasilkan jumlah pixel untuk dihitung sebagai panjang dan lebar tanaman dan ikan sebagai input variabel prediktor untuk model regresi linier. Hasil menunjukkan prediksi berat ikan memiliki tingkat error sebesar 4,50%.

Penelitian ini mengajukan proposisi penggunaan proses sains data dalam merancang penyelesaian masalah, dengan menggunakan panjang, lebar, dan berat dari tujuh jenis ikan yang umum ditemukan pada pasar ikan, dengan mengaplikasikan metode regresi linier sebagai model prediksi berat ikan.

## 2. METODOLOGI

Penelitian dilakukan dengan menerapkan metode sains data dalam menentukan penyelesaian masalah. Berikut merupakan diagram tahapan proses penelitian.

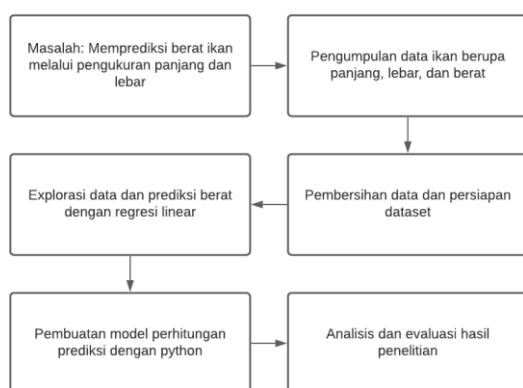


Diagram 1. Tahapan metode penelitian

### 2.1. Pengumpulan Data

Data ikan didapatkan dari SAS OnDemand for Academics melalui *repository* dataset Kaggle (<https://www.kaggle.com/aungpyaeap/fish-market>) menggunakan data dari 7 spesies ikan yang umum ditemukan pada pasar ikan. Jenis ikan yang didapatkan adalah *Bream*, *Roach*, *Whitefish*, *Parkki*, *Perch*, *Pike*, dan *Smelt*, beserta data panjang, lebar, tinggi, dan berat masing-masing ikan.

### 2.2. Persiapan Data

Pembersihan data dilakukan dengan menghilangkan nilai tidak mungkin, dalam kasus ini berupa data yang mengandung nilai berat 0. Selanjutnya dataset dibersihkan dari data *outlier* guna mengurangi nilai *error* dan meningkatkan efisiensi serta akurasi dari model prediksi yang digunakan.

### 2.3. Multiple Linear Regression

Penerapan prediksi berat ikan dilakukan dengan menggunakan metode *multiple linear regression* (MLR). Metode ini digunakan mengacu pada data yang mempunyai nilai kontinu dan referensi penelitian sebelumnya. Pada persamaan (1),  $y$  merupakan variabel dependen atau respon,  $a$  merupakan nilai *intercept* pada persamaan regresi,  $b$  merupakan koefisien *slope* dari nilai  $x$ , sedangkan  $x$  merupakan variabel independen atau prediktor.

$$y = a + b_1 * x_1 + \dots + b_n * x_n \quad (1)$$

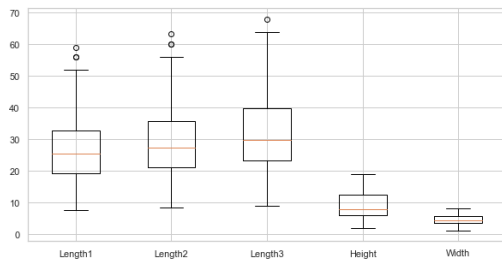
## 3. HASIL DAN PEMBAHASAN

Data yang digunakan mempunyai 6 variabel *independent*, yaitu: 1) Species yang merupakan spesies atau jenis ikan, 2) Length1 yang merupakan panjang ikan secara vertikal dalam cm, 3) Length2 yang merupakan panjang ikan secara diagonal dalam cm, 4) Length3 yang merupakan panjang ikan secara silang dalam cm, 5) Height yang merupakan tinggi ikan, serta 6) Width yang merupakan lebar diagonal ikan dalam cm. Sedangkan variabel *dependent* adalah Weight yang merupakan berat dalam gram. Tabel dibawah ini menunjukkan frekuensi pada masing-masing jenis ikan pada dataset, dimana persebaran jenis ikan tidak rata.

Tabel 1. Persebaran data dalam dataset

Spesies	Frekuensi
Pearch	56
Bream	35
Roach	20
Pike	17
Smelt	14
Parkki	11
Whitefish	6

*Outlier* merupakan nilai yang tidak umum dalam data dan nilai ini dapat mengubah kesimpulan dan hasil uji terhadap data. Terdapat banyak cara untuk mengidentifikasi *outlier*, salah satunya adalah dengan melakukan visualisasi menggunakan *boxplot*. Pada Gambar 1, *outlier* dapat ditemukan dalam dataset yang digunakan dan *outlier* dinyatakan oleh lingkaran kecil. Namun, terkadang *outlier* tidak perlu dihilangkan karena akan mengurangi informasi yang ada di dalam data.



Gambar 1. Boxplot pada dataset

Dalam penelitian ini, penulis akan menguji perbandingan antara model yang menyertakan data *outlier* dengan tidak menyertakan data *outlier*. Model tersebut selanjutnya akan dinyatakan dengan model 1 dan model 2.

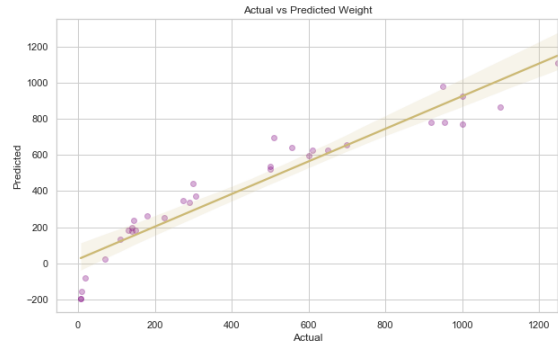
### 3.1. Model MLR Menggunakan Data *Outlier*

Dalam pemodelan, data yang digunakan adalah keseluruhan, termasuk yang terindikasi *outlier*. Dengan melakukan perhitungan pada persamaan regresi, didapatkan nilai *intercept* dari persamaan regresi dan *coefficient* dari variabel.

Tabel 2. Nilai *coefficient* dan *intercept* dari model 1

Istilah	Variabel	Nilai
<i>Coefficient</i>	Length1	3304.05
	Length2	-1128.98
	Length3	-1025.27
	Height	360.43
	Width	240.38
<i>Intercept</i>		-274.16

Jika diinterpretasikan dari Tabel 1, apabila panjang vertikal ikan sebesar 23.2 cm, panjang diagonal sebesar 25.4 cm, panjang silang sebesar 30 cm, tinggi sebesar 11.5 cm, lebar sebesar 4 cm, dan jenis ikan adalah Bream, maka berat ikan adalah sekitar 83.89 gram. Dengan menggunakan persamaan (1), hasil regresi dari model 1 adalah sebagai berikut.



Gambar 2. Hasil MLR pada model 1

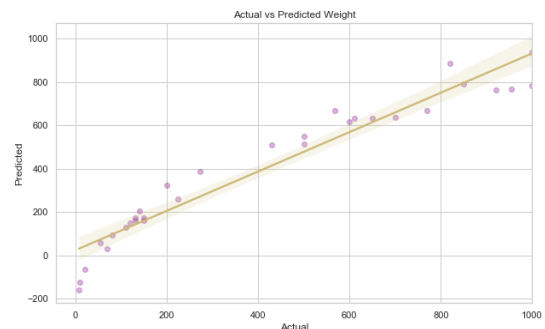
### 3.2. Model MLR dengan Menghapus Data *Outlier*

Pada dataset yang digunakan, terdapat 3 kolom data *outlier* yang dihilangkan.

Tabel 3. Nilai *coefficient* dan *intercept* dari model 2

Istilah	Variabel	Nilai
<i>Coefficient</i>	Length1	5392.87
	Length2	-3002.79
	Length3	-1878.68
	Height	616.74
	Width	359.49
<i>Intercept</i>		-244.4

Interpretasi dari Tabel 2 adalah apabila panjang vertikal ikan sebesar 23.2 cm, panjang diagonal sebesar 25.4 cm, panjang silang sebesar 30 cm, tinggi sebesar 11.5 cm, lebar sebesar 4 cm, dan jenis ikan adalah Bream, maka berat ikan adalah sekitar 113.71 gram. Dengan menggunakan persamaan (1), hasil regresi dari model 2 adalah sebagai berikut.



Gambar 3. Hasil MLR pada model 2

### 3.1. Hasil Evaluasi Model MLR

Untuk mengevaluasi model dengan membandingkan nilai aktual dan hasil prediksi, penulis menggunakan *Root Mean Squared Error* (RMSE) dan *R2 Score*. Pada persamaan (2),  $n$  merupakan jumlah data,  $\hat{y}_i$  merupakan nilai prediksi, dan  $y_i$  merupakan nilai aktual. Sedangkan pada persamaan (3),  $SS_{res}$  merupakan kuadrat total dari *residual errors* dan  $SS_{tot}$  merupakan total *error*,

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

#### 4 Proyek Akhir Mata Kuliah Pengantar Sains Data

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (3)$$

Dari hasil perbandingan model pada Tabel 4, dapat diketahui bahwa nilai RMSE pada model 2 lebih kecil dari model 1 dan nilai R2 Score lebih besar dari model 1. Hal tersebut berarti bahwa model dengan menghilangkan *outlier* lebih baik daripada dengan tidak menghilangkan *outlier*. Perbandingan RMSE antara data *train* dan data *test* tidak besar, hal ini diartikan sebagai model cukup baik.

Tabel 4. Perbandingan nilai evaluasi pada kedua model

Data	Model	Metrics Evaluasi	
		RMSE	R2 Score
Train	Model 1	123.00	0.87
	Model 2	100.64	0.89
Test	Model 1	113.26	0.90
	Model 2	89.22	0.92

#### DAFTAR RUJUKAN

- BULOG, 2014. Pengertian Ketahanan Pangan [online]. Tersedia di: <<http://www.bulog.co.id/ketahananpangan.php>> [Diakses 15 Desember 2020]
- JAYA, I., 2012. Pangan Ikan, Pangan Rakyat Negara Kepulauan, [online] Tersedia di: <<https://202.124.205.241/bitstream/handle/123456789/58449>> [Diakses 15 Desember 2020]
- IRAWAN, H., 2017. Penerapan teknologi Pen Culture Pada Budidaya Perikanan Perairan Dalam Dan Dangkal Di Perairan Laut Kepulauan Natuna. Intek Akuakultur, 1(2), pp.43-54, [online] Tersedia di: <<http://ojs.umrah.ac.id/index.php/intek/article/view/257>> [Diakses 15 Desember 2020]
- NURSANDI, J., 2018. Budidaya Ikan Daam Ember “Budikamber” dengan Akuaponik di Lahan Sempit. Prosiding Seminar Nasional Pengembangan Teknologi Pertanian, pp.129-136, [online] Tersedia di: <<https://jurnal.polinela.ac.id/index.php/PROSIDING/article/view/1150/778>> [Diakses 15 Desember 2020]
- FIRDAUSA, F., dan HASAN, A., 2020. Prediksi dan Analisis Berat Gedung dengan Structural Analysis Program 2000 (SAP 2000) dengan Metode Artificial Neural Network. Jurna Deformasi, 5(1), pp.1-10, [online] Tersedia di: <<https://jurnal.univpgri-palembang.ac.id/index.php/deformasi/article/view/4236>> [Diakses 15 Desember 2020]
- FITRIYAH, H., SETIAWAN, E., dan MASRURI, M.R.R., 2020. Applying Linear Regression to Estimate Weight of Non Axi-Symmetric Fruit. Journal of Information Technology and Computer Science, 5(2), pp.160-167, [online] Tersedia di: <<http://jitecs.ub.ac.id/index.php/jitecs/article/view/163>> [Diakses 15 Desember 2020]
- CHEN, L., YANG, X., SUN, C., WANG, Y., XU, D., dan ZHOU, C., 2020. Feed Intake Prediction Model for Group Fish Using The MEA-BP Neural Network in Intensive Aquaculture. Information Processing in Agriculture, 7(2), pp.261-271, [online] Tersedia di: <<https://www.sciencedirect.com/science/article/pii/S2214317319300964>> [Diakses 15 Desember 2020]
- FITRIYAH, H., 2020. Pengukuran Panjang-BBerat Ikan dan Sayuran pada Budikdamber (Budi Daya Ikan dalam Ember) Menggunakan Visi Komputer dan Regresi Linier.. Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan), 4(1), pp.8-14, [online] Tersedia di: <<http://jurnal.tau.ac.id/index.php/siskom-kb/article/view/166>> [Diakses 15 Desember 2020]
- BUDIMAN, I., AKHLAKULKARIMAH, A., 2015. Aplikasi Data Mining Menggunakan Multiple Linear Regression Untuk Pengenalan Pola Curah Hujan. Tersedia melalui: Kumpulan Jurnal Ilmu Komputer <[klik.ulm.ac.id](http://klik.ulm.ac.id)> [Diakses 16 Desember 2020]
- PUTRI, W., V., SAPUTRA, R., RAYENDRA, R., MUSTAKIM. 2017. Penerapan Multiple Regression dalam Pendugaan Awal Kelulusan Mahasiswa. Tersedia melalui: Seminar Nasional Teknologi Informasi Komunikasi dan Industri <[ejournal.uin-suska.ac.id](http://ejournal.uin-suska.ac.id)> [Diakses 16 Desember 2020]
- POLLET, T.V., VAN DER MEIJ, L. To Remove or not to Remove: the Impact of Outlier Handling on Significance Testing in Testosterone Data. Adaptive Human Behavior and Physiology 3, 43–60 (2017). <https://doi.org/10.1007/s40750-016-0050-z>
- COUSINEAU, D., 2011. Outliers detection and treatment : A review . International Journal of Psychological Research, 3(1), pp.58–67.
- TALLÓN-BALLESTEROS, A.J. dan RIQUELME, J.C., 2014. Deleting or keeping outliers for classifier training? 2014 6th World Congress on Nature and Biologically Inspired Computing, NaBIC 2014, pp.281–286.
- ALEXANDER, D. L. J.; TROPSHA, A.; WINKLER, DAVID A. (2015). Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. Journal of Chemical Information and Modeling, 55(7), 1316–1322. doi:10.1021/acs.jcim.5b00206