

# PREDIKSI JARAK TEMPUH KAPAL MOTOR SANGIANG MENGUNAKAN *SUPERVISED MACHINE LEARNING*

Afrioni Roma Rio<sup>1\*</sup>, Berton Maruli Siahaan<sup>2</sup>

<sup>1,2</sup>Jurusan Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sam Ratulangi  
Kampus UNSRAT, Bahu, Manado, 95115

\*Email: [afrioni@unsrat.ac.id](mailto:afrioni@unsrat.ac.id)

## Abstract

*As a maritime nation with thousands of islands and a vast sea area, sea transportation is the most effective transportation used by the people of Indonesia. A motorboat is one type of maritime transportation that is used to move people or commodities. In this article, we'll discuss predicting the daily mileage of one of the motorboats, the Sangiang, which travels from Bitung to Ternate. Three independent variables, "Anchor Time (minutes)", "Speed (knots/hour)", and "Sailing Time (minutes)", are used in supervised machine learning techniques to estimate the daily mileage (mile). Of the various methods evaluated, the multiple regression model was found to be the most accurate at forecasting the Sangiang motorboat's daily mileage.*

**Keyword:** *machine learning, motorboat, daily mileage*

## Abstrak

Indonesia sebagai negara maritim, yang memiliki ribuan pulau dan wilayah laut yang luas melebihi wilayah daratannya, maka transportasi laut adalah transportasi yang paling efektif digunakan oleh masyarakat Indonesia. Salah satu transportasi laut yang digunakan untuk mengangkut barang, ataupun orang adalah kapal motor (KM). Pada artikel ini akan dibahas prediksi jarak tempuh harian salah satu kapal motor, yaitu kapal motor Sangiang dengan rute dari Bitung ke Ternate. Prediksi jarak tempuh harian (mile) menggunakan metode *supervised machine learning* dengan tiga variabel bebas yaitu 'Waktu Jangkar (menit)', 'Kecepatan (knot/jam)', dan 'Waktu Berlayar (menit)'. Dari beberapa metode yang digunakan maka diperoleh model *multiple regression* sebagai model terbaik untuk memprediksi jarak tempuh kapal motor Sangiang.

**Kata Kunci:** *machine learning, kapal motor, jarak tempuh*

## 1. Pendahuluan

Indonesia merupakan negara yang luas, terdiri dari 17.508 pulau yang tersebar diseluruh wilayah Indonesia terbentang sepanjang 5.000 km. Indonesia merupakan negara pesisir terbesar kedua di Kawasan Samudera Hindia (*Indian Ocean Region*/IOR) dan terletak di persimpangan Samudera Hindia dan Pasifik. Indonesia juga berbagi batas maritim sepanjang hampir 1.300 km di Samudra Hindia dengan lima negara yaitu Australia, India, Malaysia, Thailand, dan Singapura. Total area yang dikelilingi oleh laut empat kali lebih besar dibandingkan dengan area daratan. Selain keunggulan geografis, Indonesia juga memiliki sejarah maritim yang kental. Dari zaman kerajaan kuno Sriwijaya dan Majapahit yang memiliki armada yang kuat dan menguasai lautan yang luas [1].

Sebagai negara maritim, maka alat transportasi laut merupakan moda transportasi yang paling efektif untuk pengangkutan orang ataupun barang. Salah satu alat transportasi laut adalah kapal motor biasa disingkat KM, pada artikel ini akan membahas tentang prediksi jarak tempuh harian sebuah kapal motor yang bernama kapal

motor Sangiang dengan rute Bitung ke Ternate. Metode prediksi yang digunakan adalah metode *supervised machine learning*. Metode ini adalah salah satu bagian dari prediksi pembelajaran mesin (*machine learning*) dengan memprediksi variabel terikat (*dependent variable*) dari suatu data, menggunakan data latih yang ada, dan variabel-variabel lain yang mempengaruhinya (*independent variable*) [2]–[4]. *Supervised machine learning* terdiri dari dua tipe, yaitu regresi, dan klasifikasi. Regresi pada umumnya digunakan untuk memprediksi suatu luaran yang berupa angka, misalnya memprediksi harga, memprediksi tekanan darah, banyaknya hasil tani yang diperoleh setiap musimnya, dan lain sebagainya. Sedangkan klasifikasi digunakan untuk memprediksi hasil luaran berupa kelompok-kelompok atau grup.

## 2. Metode Penelitian

Prediksi jarak tempuh KM Sangiang dengan *supervised machine learning* menggunakan bahasa pemrograman Python [5], dengan beberapa *package* dan *library* diantaranya adalah:

- Numpy yang digunakan untuk kalkulasi numerik

pada suatu *array* atau matriks [6]

- Pandas digunakan untuk analisa data dalam bentuk *dataframe* [7];
- Matplotlib dan Seaborn digunakan untuk visualisasi data dalam bentuk grafik [8][9];
- Scikit-learn digunakan untuk proses *machine learning* [10].

Beberapa metode *supervised machine learning* yang digunakan adalah sebagai berikut:

- *Multiple regression* [2]–[4]  
Persamaan *multiple regression* adalah sebagai berikut:

$$h_{\theta}(X) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n,$$

dengan  $h_{\theta}(X)$  adalah variable yang akan diprediksi,  $\theta_0$  adalah *intercept* (titik potong pada sumbu y), dan  $\theta_i x_i$  adalah koefisien regresi untuk variabel ke  $i$ .

- *k-nearest neighbors (k-NN) regression* [2], [11]  
Suatu model prediksi dengan mengambil nilai rata-rata dari daerah sekitarnya. Jika  $k = 1$ , maka output hanya ditetapkan pada satu data di daerah sekitar data prediksi.
- *least absolute shrinkage and selection operator (Lasso) regression* [2]

### 3. Hasil dan Pembahasan

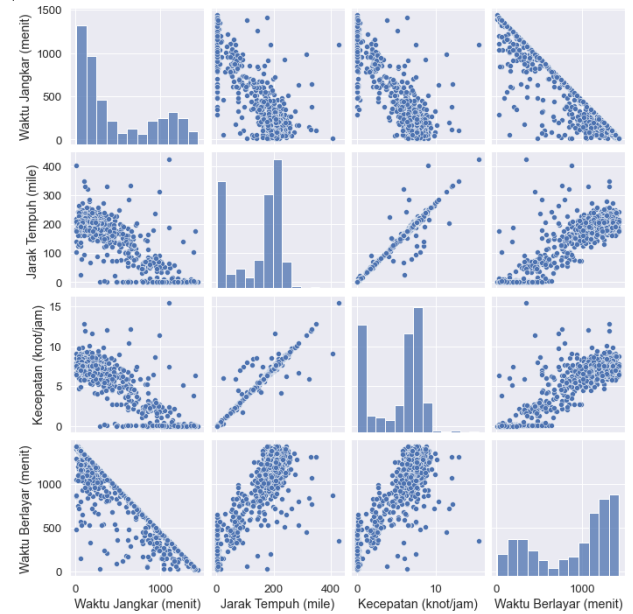
Data yang digunakan untuk memprediksi jarak tempuh KM Sangiang adalah data harian dari tahun 2018 hingga 2022, dengan menggunakan tiga variabel sebagai variabel bebas. Ketiga variabel tersebut adalah Waktu Jangkar (menit), Kecepatan (knot/jam), dan Waktu Berlayar (menit) (lihat Tabel 1).

**Tabel 1.** Data harian KM Sangiang.

Tanggal	Waktu Jangkar (menit)	Kecepatan (knot/jam)	Waktu Berlayar (menit)	Jarak Tempuh (mile)
2018-09-28	30	8.81	1425	242.707513
2018-09-29	285	6.21	1155	171.125573
2018-09-30	255	7.74	1185	213.068116
2018-10-01	105	7.8	1305	210.023398
...	...	...	...	...
2022-04-14	210	6.41	1230	176.407227
2022-04-15	195	7.54	1245	207.475777
2022-04-16	240	7.64	1185	207.972874
2022-04-17	510	4.8	930	131.854926
2022-04-18	690	4.63	750	127.878152

Dari data yang ada, dilihat pola dari datanya apakah linear, polinomial, dsb. Dari hasil grafik pada Gambar 1, hubungan antara ketiga variabel memiliki hubungan yang linear dengan jarak tempuh kapal. Hubungan antara Waktu Jangkar (menit) berbanding terbalik secara linear. Semakin lama waktu jangkar maka semakin pendek jarak tempuhnya. Sedangkan kedua parameter sisanya, yaitu Kecepatan (knot/jam), dan Waktu Berlayar (menit) berbanding lurus dengan jarak tempuhnya. Hal ini berarti

semakin cepat dan semakin lama kapal berlayar maka jarak tempuhnya juga semakin jauh.



**Gambar 1.** Grafik relasi antara jarak tempuh dengan waktu jangkar, kecepatan, dan waktu berlayar kapal motor Sangiang.

Untuk melihat suatu linearitas dari masing-masing variabel dapat dilihat dari korelasinya (lihat Gambar 2). Korelasi dari setiap variabel akan semakin kuat apabila semakin mendekati nilai 1 mutlak ( $\pm 1$ ), yang mana tanda positif menandakan berbanding lurus dan negatif berbanding terbalik. Metode korelasi yang digunakan adalah metode Pearson [12], dimana variabel Waktu Jangkar (menit) memiliki korelasi yang negatif dengan Jarak Tempuh (mile) atau berbanding terbalik. Parameter yang lain yaitu Kecepatan (knot/jam) dan Waktu Berlayar (menit) memiliki korelasi positif atau berbanding lurus.



**Gambar 2.** Korelasi jarak tempuh dengan waktu jangkar, kecepatan dan waktu berlayar kapal motor Sangiang.

Dalam membuat model *machine learning*, perlu diperhatikan seberapa baik suatu model dalam

memprediksi. Salah satu caranya adalah dengan melihat nilai erornya. Dalam artikel ini, nilai eror yang digunakan adalah *root mean squared error* (RMSE). Ketika nilai eror semakin kecil maka model yang digunakan semakin baik. Sedangkan untuk melihat seberapa *fit* model terhadap variabel yang akan diprediksi dapat menggunakan *R-squared*, dimana ketika nilai *R-squared* semakin mendekati nilai 1 maka semakin *fit* model kita, namun perlu diperhatikan jika nilai 1 biasanya adalah model yang *overfitting*. Untuk masing-masing nilai eror dan *R-squared* dapat dilihat pada Tabel 2, dimana dari evaluasi model, diperoleh model *multiple regression* sebagai model terbaik.

**Tabel 2.** Model *supervised machine learning* dan evaluasinya.

Model	Detail	Root Mean Squared Error (RMSE)	R-squared (test)	5-Fold Cross Validation
Multiple Regression	<i>selected features</i>	9.93	0.988	0.97
Lasso Regression	<i>alpha=1, all features</i>	9.867	0.988	0.969
KNN Regression	<i>k=15, all features</i>	39.963	0.804	0.8
KNN Regression	<i>k=25, all features</i>	40.436	0.8	0.796
KNN Regression	<i>k=27, all features</i>	40.474	0.799	0.796
Lasso Regression	<i>alpha=100, all features</i>	43.222	0.771	0.76
Lasso Regression	<i>alpha=1000, all features</i>	43.168	0.772	0.758

Dari model *multiple regression* diperoleh persamaan fitting sebagai berikut:

$$h_{\theta}(X) = -19.23 + 0.0135x_1 + 26.5798x_2 + 0.0195x_3$$

dengan  $x_1$  adalah Waktu Jangkar (menit),  $x_2$  adalah Kecepatan (knot/jam),  $x_3$  adalah Waktu Berlayar (menit), dan  $h_{\theta}(X)$  adalah prediksi Jarak Tempuh (mile). Dari persamaan fitting diatas maka jarak tempuh harian dari KM Sangiang dapat diprediksi (lihat Tabel 3).

**Tabel 3.** Prediksi jarak tempuh KM Sangiang menggunakan metode *multiple regression*.

Tanggal	Waktu Jangkar (menit)	Kecepatan (knot/jam)	Waktu Berlayar (menit)	Jarak Tempuh (mile)	Prediksi Jarak Tempuh (mile)
2018-09-28	30	8.81	1425	242.707513	243.185567
2018-09-29	285	6.21	1155	171.125573	172.233235
2018-09-30	255	7.74	1185	213.068116	213.082852
2018-10-01	105	7.8	1305	210.023398	215.004279
...	...	...	...	...	...
2022-04-14	210	6.41	1230	176.407227	178.00561
2022-04-15	195	7.54	1245	207.475777	208.132034
2022-04-16	240	7.64	1185	207.972874	210.223059
2022-04-17	510	4.8	930	131.854926	133.386497
2022-04-18	690	4.63	750	127.878152	127.772527

## Kesimpulan

Telah dilakukan prediksi jarak tempuh kapal harian dari KM Sangiang dengan rute Bitung ke Ternate menggunakan metode *supervised learning*. Dari beberapa model yang digunakan, diperoleh bahwa model *multiple regression* sebagai model terbaik. Dengan memasukkan variabel bebas yaitu Waktu Jangkar (menit), Kecepatan (knot/jam) dan Waktu Berlayar (menit), maka Jarak Tempuh (mile) harian KM Sangiang dapat diprediksi dengan cukup akurat. Metode ini diharapkan dapat digunakan untuk memprediksi jarak tempuh kapal untuk kapal motor yang lain dengan rute yang berbeda.

## Daftar Pustaka

- [1] B. Suseto, Z. Othman, and F. M. Razalli, "The need to reform Indonesia's maritime strategy: A review," *Indones J Geogr*, vol. 50, no. 2, pp. 145–153, 2018.
- [2] R. Choudhary and H. K. Gianey, "Comprehensive review on supervised machine learning algorithms," in *2017 International Conference on Machine Learning and Data Science (MLDS)*, 2017, pp. 37–43.
- [3] V. Nasteski, "An overview of the supervised machine learning methods," *Horizons. b*, vol. 4, pp. 51–62, 2017.
- [4] S. B. Kotsiantis, I. Zaharakis, P. Pintelas, and others, "Supervised machine learning: A review of classification techniques," *Emerging artificial intelligence applications in computer engineering*, vol. 160, no. 1, pp. 3–24, 2007.
- [5] M. F. Sanner and others, "Python: a programming language for software integration and development," *J Mol Graph Model*, vol. 17, no. 1, pp. 57–61, 1999.
- [6] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.
- [7] W. McKinney and others, "pandas: a foundational Python library for data analysis and statistics," *Python for high performance and scientific computing*, vol. 14, no. 9, pp. 1–9, 2011.
- [8] P. Barrett, J. Hunter, J. T. Miller, J.-C. Hsu, and P. Greenfield, "matplotlib—A Portable Python Plotting Package," in *Astronomical data analysis software and systems XIV*, 2005, vol. 347, p. 91.
- [9] M. L. Waskom, "Seaborn: statistical data visualization," *J Open Source Softw*, vol. 6, no. 60, p. 3021, 2021.
- [10] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [11] S. Kohli, G. T. Godwin, and S. Urolagin, "Sales prediction using linear and KNN regression," in *Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019*, 2021, pp. 321–329.
- [12] I. Cohen *et al.*, "Pearson correlation coefficient," *Noise reduction in speech processing*, pp. 1–4, 2009.