

Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda

Adji Prasetyo¹, Salahuddin², Amirullah^{3*}

^{1,2,3} Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe

¹prasyoadji080@gmail.com

²salahuddintik@pnl.ac.id

^{3*}amir@pnl.ac.id

Abstrak— Sektor industri di Indonesia khususnya daerah Kabupaten Aceh Utara banyak didominasi oleh tanaman kelapa sawit, sektor industri tersebut merupakan komoditas utama dalam perekonomian masyarakat. PT. Perkebunan Nusantara I (PTPN I) merupakan perusahaan yang bergerak dalam bidang industri kelapa sawit berupaya untuk meramalkan perencanaan demi menjaga kestabilan serta meningkatkan hasil produksi kelapa sawit. Untuk menunjang upaya tersebut dibutuhkan penentuan kebijakan, pengawasan, serta perencanaan yang baik dalam memprediksi hasil produksi kelapa sawit kedepannya, pembelajaran mesin merupakan salah satu teknik yang dapat diterapkan untuk melakukan prediksi hasil produksi dimasa yang akan datang. Metode peramalan yang digunakan pada penelitian ini adalah Regresi Linier Berganda dengan variabel *predictor* yang digunakan dalam memprediksi adalah bulan, curah hujan, umur luas lahan, jumlah pokok, jumlah tandan, rerata berat, dengan variabel terikat yaitu hasil produksi kelapa sawit. Sumber dari data yang digunakan berasal dari PT. Perkebunan Nusantara I dan sebuah data eksternal berupa curah hujan yang bersumber dari Badan Meteorologi, Klimatologi, dan Geofisika. *Data training* berjumlah 180 dan *data testing* sebanyak 20% dari *data training*. Hasil penelitian menunjukkan persamaan Regresi Linier Berganda yang didapatkan yaitu $Y = -415337,95 + 1073,82208X_1 + 3736,68741X_2 + -15306,629X_3 + -621,89932X_4 + 11,7449262X_5 + 7,47948459X_6 + 33441,5621X_7$ dengan *Mean Absolute Percentage error* (MAPE) sebesar 14.28%.

Kata kunci— Pembelajaran Mesin, Prediksi Produksi Kelapa Sawit, Regresi Linier Berganda.

Abstract— The industrial sector in Indonesia, especially the North Aceh Regency area, is dominated by oil palm plantations, the industrial sector is the main commodity in the community's economy. PT. Perkebunan Nusantara I (PTPN I) is a company engaged in the palm oil industry trying to predict plans to maintain stability and increase palm oil production. To support these efforts, it is necessary to determine policies, supervision, and good planning in predicting future palm oil production results, machine learning is one technique that can be applied to predict future production results. The forecasting method used in this study is Multiple Linear Regression with predictor variables used in predicting are month, rainfall, age of land area, number of principals, number of bunches, average weight, with the dependent variable being the yield of oil palm production. The source of the data used comes from PT. Perkebunan Nusantara I and an external data in the form of rainfall sourced from the Meteorology, Climatology and Geophysics Agency. The training data is 180 and the testing data is 20% of the training data. The results showed that the Multiple Linear Regression equation obtained was $Y = -415337,95 + 1073,82208X_1 + 3736,68741X_2 + -15306,629X_3 + -621,89932X_4 + 11,7449262X_5 + 7,47948459X_6 + 33441,5621X_7$ with Mean Absolute Percentage error (MAPE) of 14.28%.

Keywords— Machine Learning, Palm Oil Production Prediction, Multiple Linear Regression.

I. PENDAHULUAN

Sektor industri kelapa sawit Indonesia terkhususnya di daerah Kabupaten Aceh Utara banyak didominasi oleh tanaman kelapa sawit. Bahkan, sektor industri tersebut merupakan komoditas utama dalam perekonomian masyarakat. PT Perkebunan Nusantara I (PTPN I) merupakan salah satu perusahaan terbesar Badan Usaha Milik Negara (BUMN) yang berada di Kabupaten Aceh Utara bergerak dalam bidang industri kelapa sawit. PTPN I berupaya untuk terus menjaga kestabilan permintaan pasar dengan berbagai cara salah satunya dengan meningkatkan hasil produksi kelapa sawit, hal ini dilakukan karena kelapa sawit merupakan sumber utama dari berbagai macam produk kecantikan, makanan dan industri lainnya, seperti dijadikan sebagai bahan bakar energi terbarukan.

Untuk menunjang upaya tersebut dibutuhkan penentuan kebijakan, pengawasan, serta perencanaan yang baik dalam memprediksi hasil produksi kelapa sawit kedepannya, *machine learning* merupakan salah satu teknik yang dapat digunakan untuk melakukan prediksi produksi di masa yang akan datang. Salah satu algoritma dalam *machine learning* yaitu Regresi Linier Berganda yang akan digunakan dalam menemukan pola

hubungan antar faktor-faktor yang mempengaruhi hasil produksi serta digunakan dalam memprediksi hasil produksi untuk masa panen bulan berikutnya.

Machine learning merupakan bagian dari ilmu kecerdasan buatan yang memfokuskan pada pembangunan dan studi dari sebuah sistem agar mampu belajar dari data-data yang diperolehnya. *machine learning* merupakan bidang studi yang memberikan kemampuan program komputer untuk belajar tanpa harus diprogram secara eksplisit[1].

Suatu data dibutuhkan agar dapat menerapkan teknik-teknik *machine learning*, tanpa adanya data maka algoritma *machine learning* tidak dapat bekerja, data yang ada biasanya akan dibagi menjadi dua yaitu data latih dan data uji, data latih digunakan untuk melatih model algoritma, sedangkan data uji digunakan untuk mengetahui performa model algoritma yang telah dilatih sebelumnya ketika menemukan data baru yang belum pernah dilihat.

Forecasting merupakan teknik untuk menduga, meramalkan suatu kondisi di masa yang akan datang berdasarkan pada kondisi masa lalu dan sekarang, sehingga dapat dengan tepat meramalkan suatu tindakan atau keputusan. Fungsi dari *forecasting* adalah sebagai tindakan dasar seseorang pada saat akan melakukan suatu perencanaan,

misalnya perencanaan produksi dan persediaan, anggaran, jumlah penjualan, serta perencanaan kebutuhan bahan baku[2].

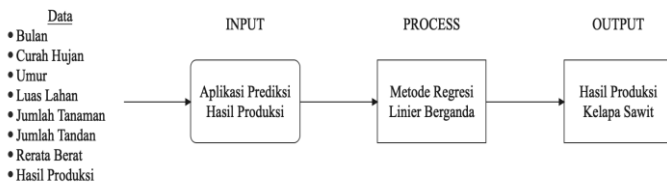
Berdasarkan permasalahan yaitu bagaimana perancangan sistem prediksi hasil produksi kelapa sawit menggunakan metode Regresi Linier Berganda agar dapat melakukan prediksi dalam kurun waktu periode bulan.

Adapun tujuan dari penelitian ini membuat sistem yang mampu mempelajari data yang diberikan sehingga dapat melakukan proses prediksi secara otomatis, dan dapat memberikan gambaran masa depan produksi kelapa sawit untuk masa panen dibulan berikutnya.

II. METODOLOGI PENELITIAN

A. Perancangan Sistem

Perancangan sistem merupakan proses memperbaiki atau merancang sistem yang sudah ada sehingga sistem menjadi lebih baik dari sebelumnya dan dapat melakukan pekerjaan secara efektif dan efisien, terdapat beberapa tahapan dalam proses perancangan yaitu berupa rancangan input, rancangan proses, dan rancangan output.



Gambar 1. Rancangan Sistem

B. Preprocessing

Preprocessing data merupakan teknik awal data mining untuk mengubah data mentah atau biasa dikenal dengan raw data yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih dan bisa digunakan untuk pengolahan selanjutnya. Proses ini bisa juga disebut dengan langkah awal untuk mengambil semua informasi yang tersedia dengan cara membersihkan, memfilter, dan menggabungkan data-data tersebut. 3 masalah umum yang diselesaikan dalam tahap preprocessing adalah menangani missing value, data noise, dan data yang tidak konsisten. Missing value merupakan data yang tidak akurat karena informasi yang hilang menyebabkan informasi yang ada di dalamnya tidak relevan. Missing value sering terjadi ketika ada masalah dalam proses pengumpulan, seperti kesalahan dalam entry data atau masalah dalam penggunaan biometrik. Data noise berisi data yang salah dan pencilan yang dapat ditemukan di kumpulan data. Pencilan dan data salah ini berisi informasi yang tidak berarti. Beberapa penyebab adanya data noise adalah karena kesalahan manusia berupa kesalahan pemberian label dan masalah lain selama pengumpulan data. Inkonsisten data terjadi ketika seseorang menyimpan file yang berisi data yang sama dengan format yang berbeda-beda. Beberapa inkonsisten data adalah duplikasi dalam format yang berbeda, kesalahan pada kode nama, dan lain sebagainya[3].

C. Regresi Linier Berganda

Regresi Linier Berganda merupakan model regresi linier yang melibatkan lebih dari satu variabel bebas atau independen yang biasanya diwakili oleh variabel X dan satu variabel

respon atau dependen yang biasanya diwakili oleh variabel Y, teknik tersebut digunakan untuk ada tidaknya pengaruh signifikan dua atau lebih variabel bebas (X1, X2, X3..., Xn) terhadap variabel terikat (Y)[4].

Secara matematis bentuk umum persamaan regresi linier berganda dapat ditulis pada persamaan 1:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n \quad (1)$$

Keterangan:

- Y = Variabel yang diprediksi (*response*)
- X = Variabel bebas yang diketahui (*predictor*)
- b0 = *Intercept*
- b = *Coefficients* regresi

Regresi Linier Berganda merupakan suatu analisis regresi yang mendeskripsikan hubungan antara peubah respon (Y) dengan faktor-faktor yang mempengaruhi yang jumlahnya lebih dari satu predictor (X1, X2, X3, X4, X5, X6, dan X7), ketika suatu keluaran yang dicari dalam bentuk numerik dan semua atribut yang dimiliki adalah numerik maka regresi linier adalah teknik yang tepat untuk menyelesaikannya. Berikut merupakan rumus dari regresi dengan bobot yang telah ditentukan:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7$$

Dimana b0 merupakan *intercept* dan b1, b2, b3, b4, b5, b6, dan b7 adalah *coefficients* yang dihitung dari data *training*. Untuk menghitung *intercept* dan *coefficients* regresi dapat menggunakan Metode Kuadrat Terkecil (*Least Square Method*), namun dapat disederhanakan dengan persamaan matriks sebagai berikut.

$$A = \begin{pmatrix} n & \sum X_1 & \sum X_2 & \sum X_3 & \sum X_4 & \sum X_5 & \sum X_6 & \sum X_7 \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 & \sum X_1X_3 & \sum X_1X_4 & \sum X_1X_5 & \sum X_1X_6 & \sum X_1X_7 \\ \sum X_2 & \sum X_2X_1 & \sum X_2^2 & \sum X_2X_3 & \sum X_2X_4 & \sum X_2X_5 & \sum X_2X_6 & \sum X_2X_7 \\ \sum X_3 & \sum X_3X_1 & \sum X_3X_2 & \sum X_3^2 & \sum X_3X_4 & \sum X_3X_5 & \sum X_3X_6 & \sum X_3X_7 \\ \sum X_4 & \sum X_4X_1 & \sum X_4X_2 & \sum X_4X_3 & \sum X_4^2 & \sum X_4X_5 & \sum X_4X_6 & \sum X_4X_7 \\ \sum X_5 & \sum X_5X_1 & \sum X_5X_2 & \sum X_5X_3 & \sum X_5X_4 & \sum X_5^2 & \sum X_5X_6 & \sum X_5X_7 \\ \sum X_6 & \sum X_6X_1 & \sum X_6X_2 & \sum X_6X_3 & \sum X_6X_4 & \sum X_6X_5 & \sum X_6^2 & \sum X_6X_7 \\ \sum X_7 & \sum X_7X_1 & \sum X_7X_2 & \sum X_7X_3 & \sum X_7X_4 & \sum X_7X_5 & \sum X_7X_6 & \sum X_7^2 \end{pmatrix}$$

$$b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{pmatrix} \quad H = \begin{pmatrix} \sum Y & Y \\ \sum X_1 & Y \\ \sum X_2 & Y \\ \sum X_3 & Y \\ \sum X_4 & Y \\ \sum X_5 & Y \\ \sum X_6 & Y \\ \sum X_7 & Y \end{pmatrix} \quad (2)$$

dengan:

A = matriks (diketahui)
H = vektor kolom (diketahui)
b = vektor kolom (tidak diketahui)

variabel b dapat di selesaikan dengan cara sebagai berikut ;

$$Ab = H$$

$$b = A^{-1} H$$

$$MAPE = \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (3)$$

Keterangan:

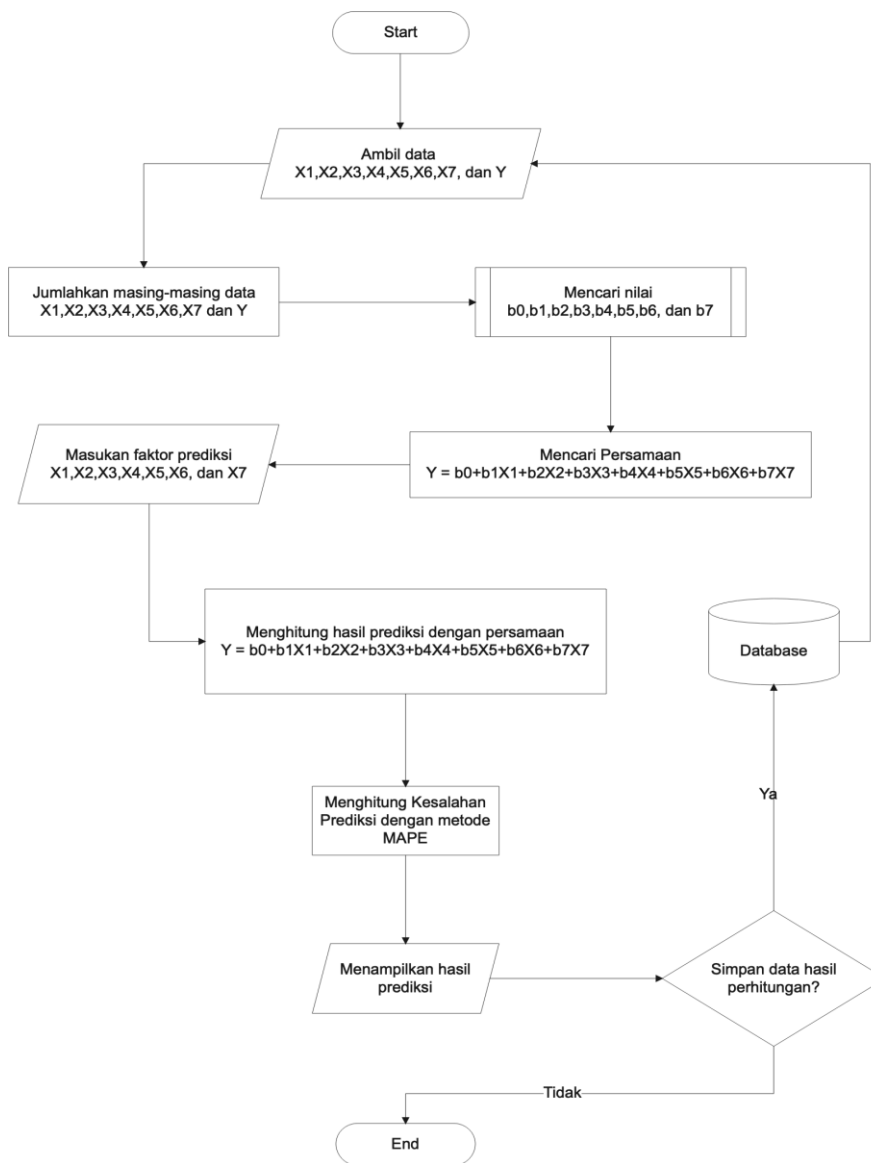
y_i = Nilai aktual
 \hat{y}_i = Nilai prediksi
n = Jumlah data

D. Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) merupakan metode yang digunakan untuk mencari perhitungan nilai rata-rata perbedaan absolut yang ada di antara nilai prediksi dan nilai realisasi yang disebutkan sebagai hasil persentas. Nilai MAPE dapat dihitung dengan menggunakan persamaan 3[5].

E. Flowchart Sistem

Agar dapat memahami alur algoritma regresi linier berganda yang ada pada sistem, dapat dilihat seperti gambar 2[6].



Gambar 2. Flowchart Sistem

Pada tahap implementasi aplikasi akan ditambahkan fitur *flowchart* pada gambar 2. Berikut adalah penjelasan dari prediksi dengan menggunakan aturan-aturan sesuai dengan tahapan-tahapan tersebut:

- *Start*, merupakan titik penanda awal akan dimulainya suatu proses.
- Sistem mengambil variabel *predictor* dan variabel *response* dari *database* yang digunakan untuk *data training* dan *data testing*.
- Menjumlahkan seluruh data dari variabel *predictor* dan variabel *response* pada masing-masing kolom.
- Mencari nilai *coefficients* dan *intercept* dari persamaan regresi linier berganda.
- Setelah nilai seluruh variabel dalam persamaan diketahui, maka prediksi dapat dilakukan dengan memasukan masing-masing variabel *predictor* (faktor prediksi).
- Menampilkan hasil prediksi serta menyimpannya kedalam *database* agar dijadikan penalaran sistem.
- *End*, merupakan titik berakhirnya suatu.

III. HASIL DAN PEMBAHASAN

Proses melakukan prediksi membutuhkan data *training* yang digunakan sebagai penalaran sistem sehingga memperoleh persamaan regresi linier berganda. Selanjutnya pencarian *coefficients* dan *intercept* dilakukan sehingga memperoleh nilai matriks yang diketahui (A) dan nilai vektor kolom diketahui (H) menggunakan persamaan (1).

TABEL I
DATA TRAINING

No	Bulan = (X1)	Cuaca = (X2)	Umur = (X3)	Luas (Ha) = (X4)	Jml Pohon = (X5)	Tandan = (X6)	Berat (Kg) = (X7)	Produksi (Kg) = (Y)
1	1	-10	6	409	54112	20205	7,12	143859,6
2	2	0	6	409	54112	53132	8,79	467030,28
3	3	10	6	409	54112	70101	9,12	639321,12
4	4	0	6	409	54112	48100	8,15	392015
5	5	-10	6	409	54112	29657	6,92	205226,44
6	6	0	6	409	54112	50750	8,23	417672,5
...
...
...
175	7	10	20	966,5	127970	64662	29,59	1694519,113
176	8	10	20	966,5	127970	59379	29,49	1655896,504
177	9	10	20	966,5	127970	55613	29,45	1629558,78
178	10	0	20	966,5	127970	53726	29,41	1576826,256
179	11	0	20	966,5	127970	43312	29,31	1503325,803
180	12	0	20	966,5	127970	46074	29,38	1524979,375

$$A = \begin{bmatrix} 180 & 1170 & -60 & 2340 & 44601,6 & 562946 & 4005571 & 3414,09 \\ 1170 & 9750 & -360 & 15210 & 289910,4 & 36591516 & 26819952 & 22249,07 \\ -60 & -360 & 10400 & -100 & 10842 & 1605530 & 7022030 & 394,5 \\ 2340 & 15210 & -100 & 33780 & 576541,2 & 73429956 & 43801259 & 48972,7 \\ 44601,6 & 289910,4 & 10842 & 576541,2 & 21885390,5 & 2796479749 & 1801579609 & 846292,659 \\ 562946 & 36591516 & 1605530 & 73429956 & 2796479749 & 3,59E+11 & 2,28E+11 & 108053343 \\ 4005571 & 26819952 & 7022030 & 43801259 & 1801579609 & 2,28E+11 & 1,89E+11 & 65425958,9 \\ 3414,09 & 22249,07 & 394,5 & 48972,7 & 846292,659 & 108053343 & 65425958,9 & 71390,8421 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 0,1895288 & -0,001964471 & 0,00275373 & -0,0018306 & 2,61485E-06 & -0,000179177 & -2,529E-06 & -0,00672687 \\ -0,0019645 & 0,000477968 & 2,4893E-05 & 6,2691E-05 & 1,02425E-08 & 5,84467E-07 & -2,607E-08 & -9,6697E-05 \\ 0,00275373 & 2,48928E-05 & 0,00015698 & 0,0001174 & 2,13563E-08 & 1,26719E-06 & -5,839E-08 & -0,00021468 \\ -0,0018306 & 6,2691E-05 & 0,0001174 & 0,00736316 & 9,54263E-07 & -0,000135566 & 1,9032E-07 & -0,00499531 \end{bmatrix}$$

2,6149E-06	1,02425E-08	2,1356E-08	9,5426E-07	8,04166E-10	-1,05842E-07	2,7685E-11	-7,7079E-07
-0,0001792	5,84467E-07	1,2672E-06	-0,0001356	-1,0584E-07	1,44036E-05	-8,909E-09	9,8992E-05
-2,529E-06	-2,6068E-08	-5,839E-08	1,9032E-07	2,76848E-11	-8,90889E-09	7,3964E-11	-5,2399E-09
-0,0067269	-9,66971E-05	-0,0002147	-0,0049953	-7,7079E-07	9,89921E-05	-5,24E-09	0,00379164

$$H = \begin{bmatrix} 72965871,73 \\ 484477978 \\ 142753973 \\ 996223700 \\ 34012360230 \\ 4,38E+12 \\ 2,88E+12 \\ 1477301019 \end{bmatrix} \quad b = \begin{bmatrix} -415337,95 \\ 1073,82208 \\ 3736,68741 \\ -15306,629 \\ -621,89932 \\ 11,7449262 \\ 7,47948459 \\ 33441,5621 \end{bmatrix}$$

Setelah *coeficients* dan *intercept* regresi telah diketahui maka persamaan Regresi Linier Berganda yang didapatkan adalah $Y = -415337,95 + 1073,82208X_1 + 3736,68741X_2 + -15306,629X_3 + -621,89932X_4 + 11,7449262X_5 + 7,47948459X_6 + 33441,5621X_7$.

F. Pengujian Sistem

Untuk mengetahui seberapa akurat sistem dalam melakukan perlu dilakukannya sebuah pengujian dengan menggunakan data *testing* dan dilakukan sebanyak 5 kali pengujian. Data *actual* merupakan data variabel *response* hasil produksi sebenarnya yang di ambil dari data *testing* dan disimbolkan dengan variabel *y* sedangkan data *predict* adalah data hasil prediksi produksi dengan menggunakan variabel *predictor* yang sama pada data *actual* dan disimbolkan dengan variabel *y*.

Untuk dapat menghitung kesalahan rata-rata pada prediksi produksi kelapa sawit dapat diselesaikan dengan menggunakan persamaan (3).

Tahapan pengujian dilakukan sebanyak 5 kali iterasi dengan data *testing* sebesar 20% dari data *training* yang digunakan sehingga setiap iterasi perhitungan dilakukan data *testing* yang digunakan sebanyak 36 data yang disajikan dalam bentuk tabel seperti berikut:

TABEL II
AKURASI PREDIKSI

	Jumlah Data	MAPE
Pengujian Ke-1	36	10,5332232 %
Pengujian Ke- 2	36	33,3883688 %
Pengujian Ke- 3	36	25,2833608 %
Pengujian Ke- 4	36	1,010999722 %
Pengujian Ke- 5	36	1,16465767 %
Rata-Rata		14,28 %

IV. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan di PT. Perkebunan Nusantara I (PTPN I) dan hasil pengembangan sistem prediksi produksi kelapa sawit dengan teknik *Machine learning*. Sistem tersebut dapat memprediksi hasil produksi kelapa sawit dalam periode bulan dengan persentase ketepatan akurasi sistem yang diperoleh dengan menggunakan metode *Mean Absolute Percentage Error* sebesar 14.28 %.

REFERENSI

- [1] Z. A. Fikriya, M. I. Irawan, and S. Soetrisno., "Implementasi Extreme Learning Machine untuk Pengenalan Objek Citra Digital," *J. Sains dan Seni ITS*, vol. 6, no. 1, p. 18, 2017, doi: 10.12962/j23373520.v6i1.21754.
- [2] T. Khotimah and R. Nindiyasari, "Forecasting Dengan Metode Regresi Linier Pada Sistem Penunjang Keputusan Untuk Memprediksi Jumlah Penjualan Batik (Studi Kasus Kub Sarwo Endah Batik Tulis Lasem)," *J. Mantik Penusa*, vol. 1, no. 1, pp. 71–92, 2017.
- [3] S. A. Alasadi and W. S. Bhaya, "Review of data preprocessing techniques in data mining," *J. Eng. Appl. Sci.*, vol. 12, no. 16, 2017, doi: 10.3923/jeasci.2017.4102.4107.
- [4] J. Adhiva, S. A. Putri, and S. G. Setyorini, "Prediksi Hasil Produksi Kelapa Sawit Menggunakan Model Regresi Pada PT . Perkebunan Nusantara V," pp. 155–162, 2020.
- [5] I. Nabillah and I. Ranggadara, "Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut," vol. 5, no. 2, pp. 250–255, 2020, doi: 10.33633/joins.v5i2.3900.
- [6] Verawati and P. D. Liksha, "Aplikasi Akuntansi Pengolahan Data Jasa Service," *J. Sist. Inf. Akunt.*, vol. 1, no. 1, p. 3, 2018.