Pemodelan dan Prediksi Produksi Padi Menggunakan Regresi Linear

Devi Wulandari[1], Rumini*[2]
[1,2]Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia
E-mail: *[1]devi@amikom.ac.id, [2]**rumini@amikom.ac.id**

Abstrak

Padi (Oryza Sativa) merupakan salah satu tanaman budidaya terpenting dalam peradaban manusia. Padi memiliki banyak varietas yang ditanam di sawah dan ladang, ketinggian 1.200 meter diatas permukaan laut. Tanaman padi merupakan bahan pangan utama yang hampir dari setengah penduduk dunia. Indonesia termasuk 10 besar negara produktivitas padi terbesar di dunia dengan rata-rata produksi mencapai 77,96 juta ton atau berkontribusi sebesar 10,28% terhadap total produksi padi dunia, namun pada lima tahun terakhir posisinya menurun. Pentingnya mengetahui beberapa aspek yang berpengaruh terhadap produksi padi. Maka dari itu, dilakukan penelitian untuk menganalisis data terkait dengan prediksi produksi padi dari data yang ada. Diperlukan algoritma untuk melihat prediksi produksi padi. Regresi linear merupakan Algoritma yang digunakan untuk prediksi dalam kasus ini. Berdasarkan model regresi didapatkan sebesar 82,6% produksi padi dipengaruhi oleh atribut luas panen, curah hujan, kelembaban, dan suhu rata-rata.

Kata Kunci*—Pemodelan, Regresi Linear, Prediksi, Padi*

1. PENDAHULUAN

Indonesia adalah negara besar yang terdiri dari puluhan ribu pulau, pulau-pulau ini menghasilkan berbagai macam sumber daya alam, diantaranya pulau Sumatera. Pulau Sumatera mempunyai lebih dari 50 persen lahan pertanian setiap provinsinya dengan komoditas pangan utama paling dominan adalah padi, sedangkan sisanya adalah jagung, kacang tanah, dan ubi. Hasil pertanian di Sumatera sangat rentan terhadap perubahan iklim karena dapat mempengaruhi pola tanam, waktu tanam, produksi dan kualitas hasil. Perubahan iklim dapat memberikan dampak negatif terhadap produksi bahan pokok tersebut. Apalagi bertambahnya suhu bumi akibat dampak dari pemanasan global yang akan mempengaruhi pola presipitasi, evaporasi, water-run off, kelembaban tanah, dan variasi iklim yang sangat fluktuatif secara keseluruhan dapat mengancam keberhasilan hasil produksi pertanian. Prediksi hasil pertanian komoditas bahan pangan banyak dipengaruhi oleh perubahan cuaca (climate change)[1].

Kekurangan pangan bisa menyebabkan kerawanan ekonomi, sosial, dan politik yang dapat menggoyahkan stabilitas nasional [2]. Peningkatkan produktivitas dan produksi padi harus terus dilakukan untuk meningkatkan pendapatan dan kesejahteraan petani serta menjamin ketahanan pangan [3].

Pengambilan data penelitian ini diperoleh dari dataset yang bersumber dari situs kaggle dan metode pengolahan datanya menggunakan regresi linear. Regresi Linear digunakan untuk melihat keeratan hubungan sebab-akibat yang terjadi antar variabel serta dapat digunakan untuk melakukan prediksi. Variabel yang digunakan pada penelitian ini adalah variabel luas panen, curah hujan, kelembaban, suhu rata - rata.

Pengambilan data diperoleh melalui website BPS pada kategori tanaman pangan utama dari 8 provinsi di pulau Sumatera yaitu Nanggroe Aceh Darussalam (NAD), Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu dan Lampung. Data yang digunakan adalah data dari tahun 1993 hingga tahun 2020 untuk dataset padi. Data memuat hasil produksi tahunan dan luas

panen atau luas lahan. Kemudian data perubahan cuaca diperoleh melalui website BMKG untuk data harian curah hujan, kelembaban, dan temperatur rata-rata atau suhu rata-rata dari tahun 1993 hingga tahun 2020[4].

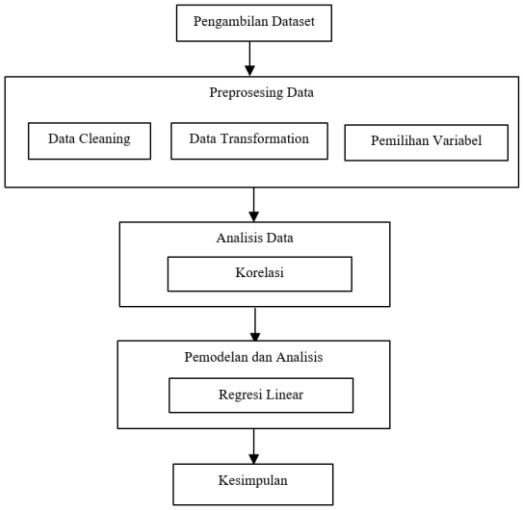
Dalam penelitian sebelumnya tentang penerapan model regresi bertatar dalam penentuan hasil tanaman sorgum didapatkan hasil bahwa dengan model regresi bertatar (stepwise analysis) tinggi tanaman saat panen, lebar malai dan panjang malai merupakan peubah bebas yang optimal dalam produksi biji sorgum dengan nilai koefisien korelasi dan determinasi adalah 0,82 dan 0,68. Uji regresi linier kenaikan 1 (satu) nilai variabel: panjang malai menaikkan produksi 0,033 ton, lebar malai menaikkan produksi 0,5 ton dan tinggi tanaman menaikkan produksi 0,06 ton. Analisis bertatar adalah prosedur memilih peubah bebas (xi) yang dominan untuk dijadikan input model regresi guna menduga besar peubah tak bebas (y) pada setiap satuan (xi)[5]. Sedangkan dalam penelitian dengan menggunakan metode yang sama yaitu dengan judul model prediksi kadar air media tanam menggunakan regresi linear berganda (studi kasus kebun tomat beef di Serenity farm mitra habibi garden) menghasilkan penelitian yaitu Model prediksi kadar air media tanam berhasil dikembangkan dengan menggunakan algoritma regresi linear berganda. Model dapat menghasilkan nilai prediksi kadar air media tanam 3 jam berikutnya dengan telah divalidasi sebanyak dua kali, yaitu pada tahap simulasi dan pengujian model. Diperoleh akurasi model berupa nilai R2 dan nilai RMSE, dimana pada tahap simulasi model sebesar 83,80% dan 1,81%, lalu pada tahap pengujian model sebesar 72,02% dan 1,74%. Akurasi model yang dihasilkan telah memenuhi kriteria model, sehingga membuktikan bahwa algoritma regresi linear berganda relevan digunakan untuk pengembangan model prediksi kadar air media tanam tanaman tomat beef di greenhouse Serenity Farm[6].

Berdasarkan penelitian sebelumnya bahwa prediksi ini sangat bermanfaat untuk dapat memberikan gambaran pengambilan keputusan dalam hal ini adalah untuk melakukan prediksi terhadap panen padi, dimana padi merupakan makanan kebutuhan negara Indonesia. Dengan adanya model regresi ini, maka akan dapat diketahui hal-hal apa yang berpengaruh terhadap besarnya hasil panen padi..

2. METODE PENELITIAN

2.1. Tahapan Penelitian

Tahapan Penelitian adalah sebagai berikut, pada gambar 1 :



Gambar 1. Alur Penelitian

2.2. Tahapan Pengumpulan Data

Penelitian ini menggunakan dataset yang bersumber dari situs Kaggle. Kaggle merupakan suatu layanan website penyedia beragam dataset dengan format.csv berupa tabel. Menggunakan sebanyak 225 set data yang dapat diunduh pada link https://www.kaggle.com/datasets/ardikasatria/datasettanamanpadisumatera dengan judul data tanaman padi Sumatera, dengan jumlah lima (5) variabel diantaranya produksi, luas panen, curah hujan, kelembapan, suhu rata-rata.

2.3. Preprocessing data

Preprocessing data dilakukan diantaranya data cleaning, data transformation dan pemilihan variabel.

1. Data Cleaning, dilakukan pengecekan data, data kosong atau data duplikat.
2. Data Transformation, mengamati apakah data sudah berdistribusi secara normal atau belum.
3. Pemilihan Variabel, dilakukan dengan cara memilih variabel apa saja yang digunakan dan variabel apa yang tidak digunakan dalam penelitian.

2.4. Analisis Data

Dalam analisis data dilakukan pengecekan apakah variabel-variabel yang akan digunakan memiliki hubungan atau saling berkorelasi.

Sejumlah penulis statistik membuat interval kategorisasi kekuatan hubungan korelasi, misalnya, membuat interval kekuatan hubungan sebagai berikut: 0 (Tidak ada Korelasi), 0.00-0.25 (Korelasi sangat lemah), 0.25-0.50 (Korelasi Cukup), 0.50-0.75 (korelasi kuat), 0.75-0.99 (korelasi sangat kuat), dan 1 (korelasi sempurna) [7].

2.5. Tahapan Review

Metode regresi merupakan sebuah metode statistik yang melakukan prediksi menggunakan pengembangan hubungan matematis antara variabel, yaitu variabel dependen (Y) dengan variabel independen (X). Variabel dependen merupakan variabel akibat atau variabel yang dipengaruhi, sedangkan variabel independen merupakan variabel sebab atau variabel yang mempengaruhi. Prediksi terhadap nilai variabel dependen dapat dilakukan jika variabel independennya diketahui.

Umumnya penjualan atau permintaan suatu produk dinyatakan sebagai variabel dependen yang besar atau nilainya dipengaruhi oleh variabel independen [8].

Metode regresi terdapat dua macam yaitu: Regresi Linear dan Regresi Non Linear. Regresi linear mempunyai model dengan 1 variabel bebas dan model dengan >1 variabel bebas (regresi linear berganda).

Sedangkan regresi non linear mempunyai model persamaan exponensial (ln) dan model persamaan berpangkat (log) [9].

Rumus untuk Regresi Linear dengan metode kuadrat terkecil atau sederhana adalah:

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} \quad (1)$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} \quad (2)$$

$$y = a + b.x \quad (3)$$

Keterangan :

Persamaan (1) adalah persamaan yang digunakan untuk mencari nilai a (konstanta) yang menunjukan besarnya nilai y apabila x = 0, dan b adalah besaran perubahan nilai y, x adalah variabel bebas [8].

Persamaan (2) adalah persamaan yang digunakan untuk mencari nilai b (koefisien regresi).

Persamaan (3) adalah persamaan yang digunakan untuk modul regresi. Y adalah variabel terikat/dependent. sedangkan x adalah variabel bebas/independent.

2.6. Kesimpulan

Dari hasil analisis yang didapatkan maka tahap terakhir yang dilakukan pada penelitian ini adalah penarikan kesimpulan.

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menerapkan algoritma Regresi untuk mendapatkan model regresi yang dapat digunakan untuk melakukan analisis prediksi produksi padi. Hasil dari penelitian ini berupa pengolahan kuantitatif dengan perhitungan yang dilakukan pada sebuah dataset publik.

3.1. Identifikasi Variabel

Data yang digunakan adalah data perbulan yang berjumlah 224 data. Variabel terikat (dependent) pada penelitian ini adalah produktivitas padi (kw/ha). Data produktivitas padi dihasilkan dari dua data utama yaitu data produksi dibagi dengan luas panen. Sedangkan variabel-variabel bebas (independent) nya yaitu luas panen, curah hujan, kelembaban, dan suhu rata-rata.

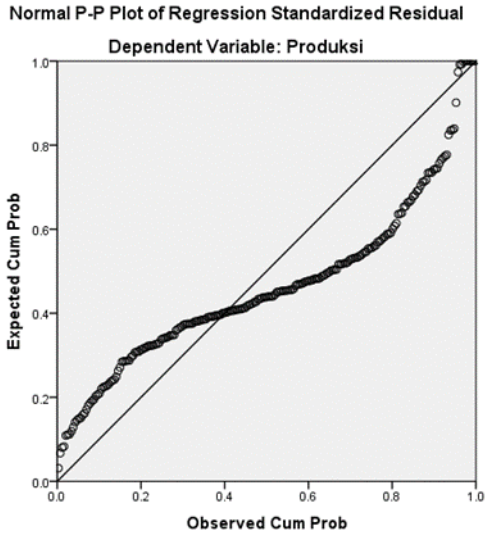
Berikut detail variabelnya.

Y = Produksi
X1 = Luas Panen
X2 = Curah Hujan
X3 = kelembaban
X4 = Suhu rata-rata

Y merupakan variabel terikat sedangkan X adalah variabel bebas.

3.1.1. Uji Normalitas

Uji yang digunakan untuk menentukan apakah suatu sampel data atau variabel tertentu berasal dari populasi yang memiliki distribusi normal atau tidak. Dari gambar 2, uji normalitas data produksi menggunakan grafik p-plot menunjukkan titik-titik data tersebar cenderung mengikuti garis diagonalnya sehingga model regresi memenuhi asumsi normalitas.



Gambar 2. Uji Normalitas data Produksi

3.1.2. Gambar dan Tabel

Pada tabel 1, berikut merupakan deskripsi bahwa jumlah data 224. Data minimum menunjukan jumlah data paling rendah dalam variabel, data maksimum menunjukkan data paling tinggi dalam variabel, sum menunjukkan jumlah, mean menunjukkan rata-rata nilai, skewness menunjukkan nilai data apakah data cenderung condong kekiri atau kekanan atau dapat juga simetris jika bernilai 0. Kurtosis menunjukan tingkat keruncingan distribusi data apakah data cenderung runcing, mendatar atau normal.

Tabel 1. Deskripsi Statistik

	N	Minimum	Maximum	Sum	Mean	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Produksi	224	42938	4881089	376252999	1679700.89	.549	.163	-.604	.324
Luas_Panen	224	63142	872737	83854393	374349.97	.429	.163	-.958	.324
Curah_Hujan	224	223	5522	549358	2452.49	.632	.163	.281	.324
Kelembapan	224	54	91	18133	80.95	-1.487	.163	3.723	.324
SuhuRataRata	224	22	30	6004	26.80	.062	.163	1.889	.324
Valid N (listwise)	224								

3.1.3. Gambar dan Tabel

Dari tabel 2 tentang korelasi/hubungan antar variabel berikut analisisnya :

Variabel produksi dengan luas panen memiliki hubungan yang sangat kuat karena nilai korelasi sebesar 0.906. dan hubungan paling rendah adalah hubungan antara produksi dengan curah hujan dan produksi dengan kelembaban karena skor korelasi adalah -0,042 dan -0,052.

Tabel 2. Korelasi/hubungan antar variable

Correlations

		Produksi	Luas_Panen	Curah_Hujan	Kelembapan	SuhuRataRata
Produksi	Pearson Correlation	1	.906**	-.042	-.052	.041
	Sig. (2-tailed)		.000	.530	.436	.540
	N	224	224	224	224	224
Luas_Panen	Pearson Correlation	.906**	1	-.093	-.061	.116
	Sig. (2-tailed)	.000		.166	.363	.084
	N	224	224	224	224	224
Curah_Hujan	Pearson Correlation	-.042	-.093	1	.056	-.229**
	Sig. (2-tailed)	.530	.166		.400	.001
	N	224	224	224	224	224
Kelembapan	Pearson Correlation	-.052	-.061	.056	1	-.408**
	Sig. (2-tailed)	.436	.363	.400		.000
	N	224	224	224	224	224
SuhuRataRata	Pearson Correlation	.041	.116	-.229**	-.408**	1
	Sig. (2-tailed)	.540	.084	.001	.000	
	N	224	224	224	224	224

**. Correlation is significant at the 0.01 level (2-tailed).

3.2. Pembahasan

3.2.1. Model Regresi

Berikut adalah model regresi yang didapatkan pada gambar 3.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.909[a]	.826	.822	489342.127

a. Predictors: (Constant), SuhuRataRata, Luas_Panen, Curah_Hujan, Kelembapan

b. Dependent Variable: Produksi

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2182767.952	1214986.811		1.797	.074
	Luas_Panen	4.564	.142	.915	32.118	.000
	Curah_Hujan	32.250	32.725	.029	.985	.325
	Kelembapan	-6206.889	7363.731	-.026	-.843	.400
	SuhuRataRata	-66720.285	30877.936	-.069	-2.161	.032

a. Dependent Variable: Produksi

Gambar 3. Data Analisis

Dari gambar 3 didapatkan model regresi sebagai berikut :

Y = 2182767, 952 + 4,56 X1 + 32,25 X2 - 6206,89 X3 - 66720,29 X4 + e

Dari model regresi diatas dapat dianalisis sebagai berikut :

Dari dataset didapatkan model regresi diatas, di daerah Sumatera untuk hasil produksi padi sebesar 2.182.767,95 tanpa dipengaruhi oleh variabel bebas (independent) yaitu luas panen, curah hujan, kelembaban, dan suhu rata-rata.

Hasil produksi padi dipengaruhi oleh variabel X1/luas panen, jika luas panen meningkat, maka produksi padi akan meningkat sebesar 4,56.

Hasil produksi padi dipengaruhi oleh variabel X2/curah hujan, jika curah hujan meningkat, maka produksi padi akan meningkat sebesar 32,25.

Hasil produksi padi dipengaruhi oleh variabel X3/kelembaban, jika kelembapan meningkat, maka produksi padi akan menurun sebesar 6.206,89.

Hasil produksi padi dipengaruhi oleh variabel X4/suhu rata-rata, jika suhu rata-rata meningkat, maka produksi padi akan menurun sebesar 66.720,29.

3.2.2. Analisis regresi

Jumlah produksi padi meningkat jika dipengaruhi oleh variabel luas panen dan curah hujan meningkat, jika kelembapan dan suhu rata-rata turun berpengaruh terhadap produksi padi yang turun.

3.2.3. Perbandingan Data Aktual dan Prediksi

Dari model regresi yang didapatkan, berikut merupakan perbandingan data nyata produksi dengan data prediksi :

Tabel 3. Tabel Data Actual Hasil Produksi dan Prediksi dari Model Regresi

Luas Panen	Curah hujan	Kelembapan	Suhu rata-rata	Data Actual	Data Predict
323,589.00	1,627.00	82.00	26.06	1,329,536.00	1,443,700.84
329,041.00	1,521.00	82.12	26.92	1,299,699.00	1,406,692.08
339,253.00	1,476.00	82.72	26.27	1,382,905.00	1,490,838.94
348,223.00	1,557.00	83.00	26.08	1,419,128.00	1,544,755.14
337,561.00	1,339.00	82.46	26.31	1,368,074.00	1,477,751.64
365,892.00	1,465.00	82.60	26.84	1,404,580.00	1,573,073.94
359,817.00	1,778.00	82.79	26.14	1,478,712.00	1,601,355.60
336,765.00	1,974.70	90.60	27.10	1,486,909.00	1,391,438.59
295,212.00	1,688.70	69.48	28.90	1,547,499.00	1,206,217.68
315,131.00	1,296.80	68.75	29.20	1,314,165.00	1,267,729.28
367,636.00	1,507.20	70.66	29.40	1,246,614.00	1,485,588.14
370,966.00	1,097.00	80.84	29.40	1,350,748.00	1,424,158.96
337,893.00	710.50	79.50	26.80	1,411,650.00	1,444,655.70
320,789.00	506.50	80.80	26.73	1,552,078.00	1,357,710.28
360,717.00	1,414.00	81.50	26.38	1,556,858.00	1,585,660.50
329,109.00	1,270.40	78.50	27.00	1,402,287.00	1,416,047.22
359,375.00	1,577.00	78.70	26.90	1,533,369.00	1,567,562.74
352,281.00	1,986.00	81.40	27.10	1,788,738.00	1,518,727.57
380,686.00	1,268.00	79.40	27.10	1,772,962.00	1,635,808.17
387,803.00	1,098.00	79.60	26.90	1,582,393.00	1,674,454.87
419,183.00	1,623.60	80.70	27.00	2,331,046.00	1,819,115.96

376,137.00	2,264.40	78.30	27.10	1,820,062.00	1,654,299.05
461,060.00	1,575.00	80.00	27.10	1,956,940.00	2,003,667.84
dan seterusnya					

4. KESIMPULAN

Berdasarkan model regresi yang didapat, sebesar 82,6% faktor-faktor produktivitas padi dapat dijelaskan oleh produksi, luas panen, luas tanam, curah hujan, dan hari hujan. Sedangkan sisanya 19,54% dapat dijelaskan oleh faktor-faktor lain yang tidak diteliti dalam penelitian ini. Variabel-variabel yang mempengaruhi peningkatan jumlah produktivitas padi yaitu variabel luas panen dan curah hujan, sedangkan variabel-variabel yang mempengaruhi penurunan jumlah produktivitas yaitu jika variabel kelembaban dan suhu rata-rata meningkat. Jumlah produksi padi akan meningkat jika dipengaruhi oleh variabel luas panen dan curah hujan meningkat, jika kelembapan dan suhu rata-rata turun berpengaruh terhadap produksi padi yang turun.

DAFTAR PUSTAKA

[1] https://bpiw.pu.go.id/uploads/publication/attachment/Buku_1Sumatera.pdf
[2] Suryana. A. 2002. Keragaan Perberasan Nasional. Dalam Pambudy et al. (Eds). Kebijakan Perberasan di Asia. Regional Meeting in Bangkok. 1October 2002.
[3] Irawan, B. 2004. Dinamika produktivitas dan kualitas budidaya padi sawah. Dalam Ekonomi Padi dan Beras Indonesia. Badan Litbang Pertanian. Deptan. 435 hal.
[4] https://www.kaggle.com/datasets/ardikasatria/datasettanamanpadisumatera
[5] Suwardi. Andayani, NN. Tabri, F. Aqil, M. 2020. Penerapan Model Regresi Bertatar dala Penentuan Hasil Tanaman Sorgum. UNS: Agrotechnology Research Journal.
[6] Listina, S.A. Sampurno, R.M. Ciptaningtyas D. Thoriq, A. 2022. Model Prediksi Kadar Air Media Tanam Menggunakan Regresi Linear Berganda (Studi Kasus Kebun Tomat Beef di Serenity Farm Mitra Habibi Garden). Jurnal Teknotan.
[7] I. A. Prabowo and H. Wijayanto, "Analisa Penggunaan Aplikasi Android Terhadap Penentuan Topik Tugas Akhir Mahasiswa," SNATIF, vol. 5, no. 1, 2018
[8] G. N. Ayuni and D. Fitrianah, "Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ," Jurnal telematika, vol. 14, no. 2, pp. 79--86, 2019.
[9] N. Nafi'iyah, "Perbandingan regresi linear, backpropagation dan fuzzy mamdani dalam prediksi harga emas," Prosiding SENIATI, vol. 2, no. 2, pp. 291--B, 2016.
[10] Tesa Nur Padilah dan Riza Ibnu Adam. 2019. Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang. FIBONACCI : Jurnal Pendidikan Matematika dan Matematika. Volume 5 No. 2 Bulan Desember Tahun 2019.
[11] Muhammad Faisal dan Ahmad Mutatkin Bakti. 2023. Implementasi Algoritma Monte Carlo Untuk Memprediksi Permintaan Aksesoris Mobil. JURIKOM (Jurnal Riset Komputer), Vol. 10 No. 2, April 2023.

[12] I. A. Prabowo and H. Wijayanto, "Analisa Penggunaan Aplikasi Android Terhadap Penentuan Topik Tugas Akhir Mahasiswa," SNATIF, vol. 5, no. 1, 2018.
[13] P. B. Utomo, E. Utami and S. Raharjo, "Implementasi Metode K-Nearest Neighbor Dan Regresi Linear Dalam Prediksi Harga Emas," Informasi Interaktif, vol. 4, no. 3, pp. 155--159, 2019.
[14] N. Nafi'iyah, "Perbandingan regresi linear, backpropagation dan fuzzy mamdani dalam prediksi harga emas," Prosiding SENIATI, vol. 2, no. 2, pp. 291--B, 2016.
[15] Y. Widiastuti, S. W. Sihwi and M. E. Sulistyo, "Decision Support System for House Purchasing using KNN (K-Nearest Neighbor) Method," ITSMART: Jurnal Teknologi dan Informasi, vol. 5, no. 1, pp. 43--49, 2016.
[16] C. A. Rahardja, T. Juardi and H. Agung, "Implementasi Algoritma K-Nearest Neighbor Pada Website Rekomendasi Laptop," Jurnal Buana Informatika, vol. 10, no. 1, pp. 75--84, 2019.
[17] M. Masruroh and K. F. Mauladi, "Penerapan Metode Regresi Linear Berganda dalam Sistem Prediksi Nilai Ujian Nasional Siswa SMP," Jurnal Teknika, vol. 12, no. 1, pp. 1--6, 2020.