



PREDIKSI PENJUALAN *HANDPHONE* DI TOKO X MENGGUNAKAN ALGORITMA REGRESI LINEAR

Yubi Aqsho Ramadhan¹, Ahmad Faqih², Gifthera Dwilestari³

^{1,2,3} Teknik Informatika, STMIK IKMI Cirebon
Cirebon, Jawa Barat, Indonesia, 45131

yubiaqsho@gmail.com, ahmadfaqih367@gmail.com, gifthera.ikmi@gmail.com

Abstract

Mobile phones or smartphones have become a basic necessity in life. This device has transformed how we communicate and access information, and entertainment, making our lives easier and more comfortable. Store X sells mobile phones from various brands. The purpose of this research is to predict the sales of handphoned in the next three months based on the sales data of brand X. Linear regression is used as the prediction method, with the number of handphoned sold as the Y variable and the sales period as the X variable. RMSE (Root Mean Squared Error) and Relative Error are used to evaluate the prediction results. The predicted sales for the Entry category in the first month are 84 units, 86 units in the second month, and 88 units in the third month, while for the Mid category, 28 teams are sold in the first month, 29 teams in the second month, and 30 units in the third month. The RMSE evaluation result for the Entry category is 10.36, while the Relative Error value is 19.11%, and the RMSE value for the Mid category is 7.50, while the Relative Error value is 32.97%. The prediction of handphoned sales using this linear regression method can be classified as sufficient or usable.

Keywords: Linear Regression, Prediction, RMSE, Relative Error, Sales

Abstrak

Pada saat ini *handphone* atau *smartphone* menjadi kebutuhan pokok dalam kehidupan. Perangkat ini telah merubah cara berkomunikasi, mengakses informasi, dan hiburan, *handphone* telah membuat hidup kita lebih mudah dan nyaman. Toko X menjual *handphone* dari berbagai macam merk. Tujuan dari penelitian ini adalah untuk memprediksi penjualan *handphone* pada 3 bulan berikutnya pada data penjualan *handphone* merk X. Regresi linear digunakan sebagai metode prediksi dengan jumlah *handphone* yang terjual sebagai variabel Y dan periode penjualan *handphone* sebagai variabel X. RMSE (*Root Mean Squared Error*) dan *Relative Error* digunakan untuk mengevaluasi hasil prediksi. Hasil prediksi kategori *Entry* pada bulan pertama terjual 84 unit, pada bulan kedua 86 unit dan bulan ketiga 88 unit dan kategori *Mid* pada bulan pertama terjual unit 28, pada bulan kedua 29 unit dan bulan ketiga 30 unit. Hasil evaluasi nilai RMSE pada kategori *Entry* 10.36, sedangkan nilai *Relative Error* kategori *Entry* 19.11% dan Nilai RMSE kategori *Mid* 7.50, sedangkan Nilai *Relative Error* kategori *Mid* 32.97%. Prediksi penjualan *handphone* menggunakan metode regresi linear ini dapat dikatakan tergolong dalam kategori cukup atau bisa digunakan.

Kata kunci: Prediksi, Penjualan, RMSE, *Relative Error*, Regresi Linear

1. PENDAHULUAN

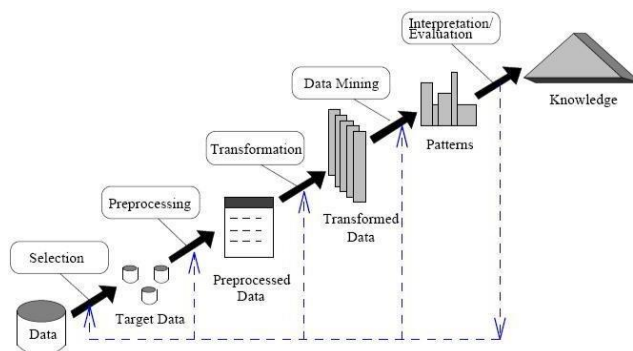
Handphone atau *smartphone* sangat memudahkan kehidupan sehari-hari, termasuk dalam belanja, mencari informasi, dan berkomunikasi [1]. Setiap penjualan di toko X dikumpulkan menjadi data, yang dapat diolah dengan data mining untuk menghasilkan pengetahuan baru dan memperbaiki pengambilan keputusan di masa depan [2]. Penelitian prediksi penjualan toko X dengan merek X pada 2021-2022 untuk pengambilan keputusan. Prediksi atau *Forecasting* adalah ramalan terhadap kejadian masa depan

yang membantu dalam pengambilan keputusan[3]. Pada penelitian ini prediksi atau *Forecasting* bersifat *time series*, data urutan waktu atau "*time series data*" merujuk pada data yang terkumpul, tercatat atau diamati secara berurutan sesuai dengan waktu [4]. Regresi linear, metode statistik untuk *time series*, mempelajari hubungan antara variabel *dependent* Y dan variabel *independent* X, digunakan untuk memprediksi nilai Y berdasarkan nilai yang diberikan pada X[5]. Toko X mengalami hambatan *cashflow* karena stok merek X yang masih belum terjual. Studi *data mining*

dengan regresi linear pada data penjualan *handphone* merek X di toko X dan dievaluasi dengan 2 indikator penilaian yaitu RMSE dan *Relative Error*. RMSE adalah akar kuadrat rata-rata *error* dari suatu metode yang digunakan untuk mengevaluasi kesesuaian data sebenarnya dengan model prediksi. Semakin mendekati 0 nilai RMSE, semakin akurat prediksinya[6]. *Relative error* adalah perbandingan kesalahan pada nilai aktual dengan kesalahan yang diukur, yang dinyatakan dalam satuan persen. Semakin kecil persentase kesalahan, semakin akurat prediksinya. Keputusan mengenai akurasi jika *Relative Error* $\leq 25\%$ dianggap akurat, sedangkan nilai $> 25\%$ tetapi $\leq 50\%$ perlu dipertimbangkan jika ingin digunakan, dan estimasi dengan nilai $> 50\%$ dianggap tidak akurat[7]. diharapkan dapat memberikan prediksi penjualan dan memandu keputusan terbaik dalam menentukan stok *handphone*.

2. METODE PENELITIAN

Penelitian ini menerapkan jenis penelitian deskriptif menggunakan pendekatan kuantitatif. Deskriptif kuantitatif berfungsi untuk menentukan jumlah setiap variabel dimana satu atau lebih variabel bebas, tanpa membentuk suatu ikatan atau perbedaan dengan variabel yang lain dengan tujuan memberikan penjelasan yang objektif tentang suatu kondisi[8]. Dengan menggunakan Regresi Linear. Regresi Linear merupakan metode prediksi yang mempelajari ikatan antara variabel *dependent* dan variabel *independent*. Algoritma regresi linear berfungsi untuk mengukur kekuatan hubungan dua variabel, dan menunjukkan alur hubungan antara variabel *dependent* dengan variabel *independent* [9]. KDD (*Knowledge Discovery in Database*) digunakan dalam menganalisis data pada penelitian ini. Gambar tahapan KDD bisa dilihat pada Gambar1 berikut.



Gambar 1. Tahapan KDD

2.1 KDD (*Knowledge Discovery in Database*)

Proses mencari hubungan antar variabel menggunakan *data mining* dan *Knowledge Discovery in Databases* (KDD) untuk menggali data tersembunyi dari informasi besar. *Data mining* adalah tahapan dalam KDD, yang juga terdiri dari *Selection*, *Preprocessing*, *Transformation*, *Evaluation*, dan *Knowledge*. KDD dan *data mining* berbeda konsep namun terkait [10]. Pada penelitian ini dataset tentang penjualan *handphone* dengan merk X di toko X sebanyak 730 data dari

tahun 2021- 2022 menggunakan 1 atribut tanggal dan 22 atribut tipe *handphone* dan 1 atribut jumlah penjualan.

2.2 Tahapan Penelitian

a. *Data Selection*

Sebelum memasuki tahap pencarian informasi dalam proses KDD, langkah penting yang harus dilakukan adalah memilih data yang relevan dari berbagai data operasional yang tersedia. Tahap ini dikenal sebagai *Data Selection* [11]. Pada proses ini dilakukan proses penggalian informasi. Atribut yang dipilih dari *dataset* adalah periode, kategori *Entry* dan *Mid* data penjualan 2021-2022. Hasil *Data Selection* dapat dilihat pada Tabel 1 berikut.

Tabel 1. Hasil *Data Selection*

Periode	Entry	Midt
Jan-21	46	5
Feb-21	48	12
Mar-21	53	8
Apr-21	57	10
May-21	46	18
Jun-21	43	17
Jul-21	46	14
Aug-21	62	21
Sep-21	67	22
Oct-21	69	19
Nov-21	59	18
Dec-21	63	26
Jan-22	61	0
Feb-22	59	0
Mar-22	55	5
Apr-22	46	15
May-22	56	19
Jun-22	43	31
Jul-22	48	22
Aug-22	69	25
Sep-22	66	29
Oct-22	108	35
Nov-22	98	28
Dec-22	118	27

b. *Preprocessing*

Proses *preprocessing* mencakup beberapa langkah, antara lain menghilangkan data yang tidak relevan atau duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan dalam data seperti kesalahan penulisan atau ketik[12]. Pada tahap ini tidak dilakukan *preprocessing* data, karena tidak ada kesalahan pada data

c. *Transformation*

Transformation bertujuan untuk mengurangi kerumitan data, memfasilitasi pengubahan data, serta mencegah salah satu variabel bebas mendominasi variabel bebas lainnya

yang dapat berdampak pada hasil prediksi [13]. Pada atribut periode ubah menjadi atribut nomer, untuk membantu memudahkan koordinasi data yang akan diproses. Hasil *Transformation* dapat dilihat pada Tabel 2 berikut.

Tabel 2. Hasil *Transformation*

No	Entry	Midt
-11	46	5
-10	48	12
-9	53	8
-8	57	10
-7	46	18
-6	43	17
-5	46	14
-4	62	21
-3	67	22
-2	69	19
-1	59	18
0	63	26
0	61	0
1	59	0
2	55	5
3	46	15
4	56	19
5	43	31
6	48	22
7	69	25
8	66	29
9	108	35
10	98	28
11	118	27

d. Data mining

Tahap kelima dalam proses KDD adalah langkah di mana tugas *Data mining* yang sesuai dipilih sesuai dengan tujuan yang telah ditetapkan pada tahap awal untuk menggali informasi dari data[14]. Pada proses ini menggunakan aplikasi Rapidminer, dilakukan proses pemodelan pada setiap kategori *Entry* dan *Midt* untuk menentukan nilai persamaan algoritma regresi linear, setelah didapat kan hasil persamaan $Y = a+b(X)$ dari setiap kategori, selanjutnya dilakukan proses prediksi penjualan 3 bulan berikutnya dari setiap kategori, setelah mendapatkan hasil prediksi dilakukan perhitungan evaluasi dari hasil prediksi dengan menggunakan RMSE dan *Relative Error* untuk mengetahui akurasi hasil prediksi.

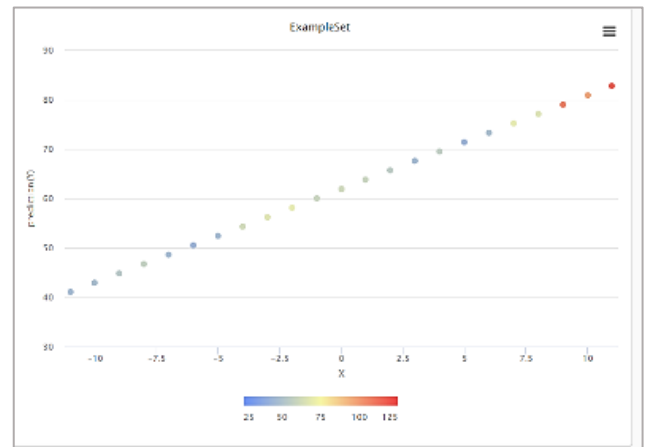
e. Evaluation

Diperlukan penyajian pola informasi hasil dari *data mining* dalam format yang dapat dipahami dengan mudah oleh para pihak yang berkepentingan. *Evaluation* merupakan tahap penting dalam proses KDD yang melibatkan pemeriksaan apakah pola atau informasi yang ditemukan sesuai dengan fakta atau hipotesis yang telah ada sebelumnya [15].

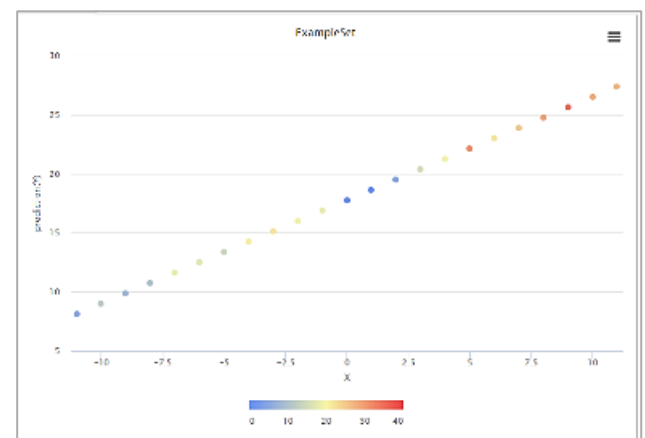
Ditetapkan hasil evaluasi dari hail yang telah diprediksi, Nilai RMSE dan *Relative Error* bisa dikatakan sangat baik jika mendekati 0 atau prediksi yang mendekati nilai aktual.

f. Knowledge

Berdasarkan hasil validitas prediksi penjualan *handphone* pada masing-masing kategori ditampilkan *plot Scatter Bubble* hasil prediksi algoritma regresi linear, diketahui hasil prediksi penjualan *handphone* memiliki nilai yang konstan. Hasil *Plot Scatter* dapat dilihat pada Gambar 2 dan 3 berikut.



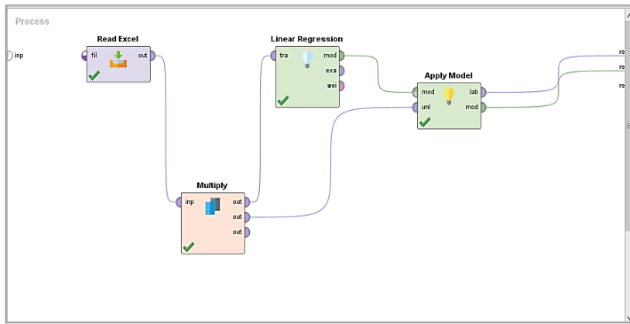
Gambar 2. *Plot Scatter Entry*



Gambar 3. *Plot Scatter Mid*

3. HASIL DAN PEMBAHASAN

Pengujian didapatkan beberapa model *data mining* Regresi Linear menggunakan *tools* Rapidminer dalam menentukan persamaan regresi linear, memprediksi penjualan periode berikutnya dan menentukan hasil evaluasi dari hasil prediksi menggunakan RMSE dan *Relative Error*. Desain menentukan persamaan dapat dilihat pada Gambar 4 berikut.



Gambar 4. Desain Menentukan Persamaan

Berdasarkan proses perhitungan persamaan nilai a dan b didapatkan kategori *Entry* $Y = 61.92 + 1.90(X)$ dan kategori *Mid* $Y = 17.75 + 0.87(X)$. Hasil pemodelan dapat dilihat pada Gambar 5 dan 6 berikut.

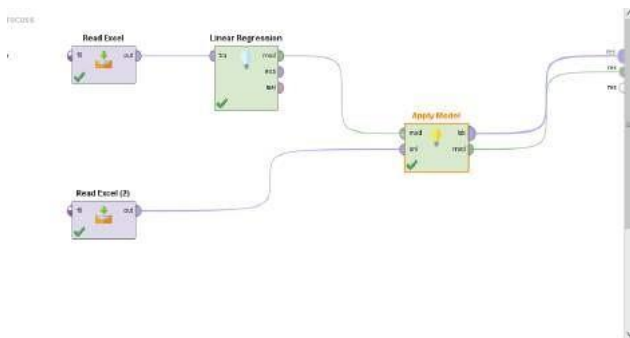
Attribute	Coefficient
X	1.897
(Intercept)	61.917

Gambar 5. Hasil Pemodelan *Entry*

Attribute	Coefficient
X	0.875
(Intercept)	17.750

Gambar 6. Hasil Pemodelan *Mid*

Setelah persamaan didapatkan dilakukan prediksi penjualan *handphone* pada 3 bulan dengan nilai $X = 12, 13, 14$. Desain prediksi data penjualan dapat dilihat pada gambar 7 berikut.



Gambar 7. Desain Prediksi data penjualan

Didapat kan kategori *Entry* pada bulan pertama terjual 84 unit, pada bulan kedua 86 unit dan bulan ketiga 88 unit, sedangkan kategori *Mid* pada bulan pertama terjual 28 unit, pada bulan kedua 29 unit dan bulan ketiga 30 unit. Hasil prediksi dapat dilihat pada Gambar 8 dan 9 berikut.

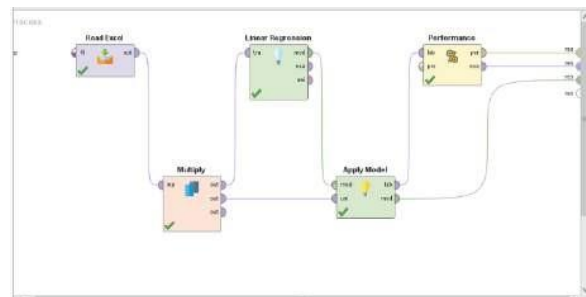
Row No.	Y	prediction(Y)	X
1	?	84.683	12
2	?	86.581	13
3	?	88.478	14

Gambar 8. Hasil Prediksi *Entry*

Row No.	Y	prediction(Y)	X
1	?	28.256	12
2	?	29.131	13
3	?	30.007	14

Gambar 9. Hasil Prediksi *Mid*

Setelah hasil prediksi didapatkan dilakukan evaluasi menggunakan 2 indikator penilaian RMSE dan *Relative Error*. Desain Evaluasi RMSE dan *Relative Error* dapat dilihat pada Gambar 10 berikut.



Gambar 10. Desain Evaluasi RMSE dan *Relative Error*

Didapatkan nilai RMSE kategori *Entry* 14.99 dan *Relative Error* 19.11% sedangkan pada kategori *Mid* nilai RMSE 7.50 dan *Relative Error* 32.97%. Hasil Evaluasi RMSE dan *Relative Error* dapat dilihat pada Gambar 11 dan 12 berikut.

root_mean_squared_error	relative_error
root_mean_squared_error: 14.990 +/- 0.000	relative_error: 19.11% +/- 15.66%

Gambar 11. Hasil Evaluasi RMSE dan *Relative Error* *Entry*

root_mean_squared_error	relative_error
root_mean_squared_error: 7.502 +/- 0.000	relative_error: 32.97% +/- 57.94%

Gambar 12. Hasil Evaluasi RMSE dan *Relative Error* *Mid*

4. KESIMPULAN

Pada penelitian ini diusulkan algoritma regresi linear untuk memprediksi penjualan *handphone* di toko X, setelah melewati proses tahapan KDD (*Knowledge Discovery in Data Base*) dan *data mining* didapatkan hasil dari prediksi penjualan 3 bulan berikutnya, dengan nilai evaluasi yang bisa disimpulkan cukup atau bisa digunakan hasil dari persamaan regresi linear, karena hasil evaluasi RMSE dan *Relative Error* cukup mendekati nilai 0.

DAFTAR PUSTAKA

- [1] E. Yuliana dan S. Nurfiani, "Pengaruh Penggunaan Handphone Terhadap Prestasi Belajar Siswa Pada Mata Pelajaran SKI (Sejarah Kebudayaan Islam) Kelas XI di MA NW Dames Tahun pelajaran 2020/2021," *Jurnal Al-Amin*, vol. 6, 2021.
- [2] D. S. O. Panggabean, E. Buulolo, dan N. Silalahi, "Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda," *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, hlm. 56, Feb 2020, doi: 10.30865/jurikom.v7i1.1947.
- [3] Y. Kurnia Hadi, M. Julian Syaputra, dan D. Setiawan, "Peramalan Penjualan Obat Generik Melalui Time Series Forecasting Model Pada Perusahaan Farmasi di Tangerang: Studi Kasus," *Journal Industrial Engineering & Management Research (JIEMAR)*, vol. 1, no. 2, hlm. 2722–8878, doi: 10.7777/jiemar.v1i2.
- [4] M. N. Arridho dan Y. Astuti, "Penerapan Metode Single Exponential Smoothing untuk Memprediksi Penjualan Katering pada Kedai Pojok Kedaung Implementation Single Exponential Smoothing Method For Sales Catering Prediction At Kedai Pojok Kedaung," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 2, no. 02, hlm. 35–44, 2020.
- [5] E. Kwok dan W. Susanti, "Penerapan Metode Regresi Linier dalam Aplikasi Sistem Peramalan Jumlah Bahan Baku untuk Produksi Tahu," *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi*, vol. 1, no. 2, hlm. 121–128, 2019.
- [6] F. H. Hamdanah dan D. Fitrihanah, "Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikro, Kecil, dan Menengah," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 10, no. 1, hlm. 23, Apr 2021, doi: 10.23887/janapati.v10i1.31035.
- [7] B. Pradito dan D. Silvi Purnia, "Komparasi Algoritma Linear Regression Dan Neural Network Untuk Memprediksi Nilai Kurs Mata Uang," *Jurnal Sains dan Manajemen*, vol. 10, no. 2, 2022.
- [8] D. S. Purnia, H. Muhajir, M. F. Adiwisastro, dan D. Supriadi, "Pengukuran Kesenjangan Digital Menggunakan Metode Deskriptif Berbasis Website," *Jurnal Sains dan Manajemen*, vol. 8, no. 2, 2020.
- [9] A. Anggrawan, N. Azmi, U. Bumigora, dan I. Anthonyangrawan, "Prediksi Penjualan Produk Unilever Menggunakan Metode Regresi Linear," *Jurnal Bumigora Information Technology (BITE)*, vol. 4, no. 2, hlm. 123–132, 2022, doi: 10.30812/bite.v4i2.2416.
- [10] F. Ginting, E. Buulolo, dan E. R. Siagian, "Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang)," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, hlm. 274–279, Nov 2019, doi: 10.30865/komik.v3i1.1602.
- [11] E. Yanti, M. Yetri, dan F. Taufik, "Penerapan Data Mining Untuk Mengestimasi Biaya Pembangunan Di Desa Puang Aja Biaya Pembangunan Di Desa Puang Aja Regresi Linear Berganda," *Jurnal CyberTech*, 2022, [Daring]. Tersedia pada: <https://ojs.trigunadharma.ac.id/>
- [12] A. Tameliza Sitompul, M. Yetri, dan R. Mahyuni, "Data Mining Mengestimasi Jumlah Tonase Kelapa Sawit Dengan Metode Regresi Linear Berganda," *Jurnal Sistem Informasi TGD*, vol. 2, no. 1, hlm. 148, 2023, [Daring]. Tersedia pada: <https://ojs.trigunadharma.ac.id/index.php/jsi>
- [13] M. N. Pangestu, M. Jajuli, dan U. Enri, "Prediksi Harga Kartu Grafis NVIDIA Berdasarkan Pengaruh Harga Cryptocurrency Menggunakan Support Vector Regression," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 17, hlm. 280–287, 2022, doi: 10.5281/zenodo.7076540.
- [14] I. Sugiyarto, R. Irawan, dan D. Rosiyadi, "Pengelompokan Dampak Gempa Bumi dan Kerusakan Pada Wilayah Berpotensi Gempa di Provinsi Sumatera Barat," *Journal of Students Research in Computer Science*, vol. 2, no. 2, hlm. 211–222, 2021, [Daring]. Tersedia pada: <http://ejurnal.ubharajaya.ac.id/index.php/JSRCS>
- [15] M. Fajri, A. Amin, Y. N. Kunang, dan S. D. Purnamasari, "Penerapan Model Regression Untuk Prediksi Cuaca Wilayah seberang Ulu 1 Palembang," *Bina Darma Conference on Computer Science*, hlm. 185–197, 2019.