

ANALISIS REGRESI LINEAR DAN MOVING AVERAGE DALAM MEMPREDIKSI DATA PENJUALAN SUPERMARKET

Nur Nafi'iyah¹, Eka Rakhmawati²

¹Dosen Program Studi Teknik Informatika, Universitas Islam Lamongan

²Mahasiswa Program Studi Teknik Informatika, Universitas Islam Lamongan

¹mynaff@unisla.ac.id, ²ekarakhmawati125@gmail.com

Abstrak

Penjualan di supermarket saat masa pandemi Covid 19 dengan saat normal akan berubah. Penjualan di masa normal kemungkinan selalu ramai dan banyak. Berbeda di masa pandemi, karena beberapa lapisan masyarakat mengalami kesulitan ekonomi. Supermarket yang tidak bisa mengatasi masalah ini kemungkinan ada yang rugi atau bisa juga harus tutup sementara waktu. Dari latar belakang tersebut Kami mengusulkan suatu analisa prediksi penjualan barang elektronik dan kesehatan di supermarket dengan metode Regresi Linear dan Moving Average. Data yang digunakan dalam penelitian ini diambil dari website Kaggle dataset. Data terdiri dari beberapa jenis barang, akan tetapi Kami hanya menggunakan kategori elektronik dan kesehatan. Tujuan penelitian ini, yaitu untuk menganalisa kedua algoritma dalam memprediksi data penjualan elektronik dan kesehatan. Tool yang digunakan untuk membantu menganalisa adalah python. Cara menganalisa kedua algoritma dengan MSE (Mean Square Error) dan RMSE. Hasil analisa menunjukkan bahwa metode Moving Average mempunyai kinerja yang paling baik dibuktikan dengan nilai MSE dan RMSE dari data elektronik adalah 57.603, 7.59, data kesehatan adalah 50.489, 7.106. Sedangkan metode Regresi Linear hasil prediksi pada data elektronik dan kesehatan nilai MSE dan RMSE adalah 114.79, 10.71, 59.965, 7.744.

Kata kunci : Analisa Algoritma, Moving Average, Regresi Linear, Data Kesehatan, Data Elektronik

Abstract

Sales in supermarkets during the Covid 19 pandemic will change normally. Sales in normal times may always be crowded and a lot. It is different during a pandemic because several layers of society experience economic difficulties. Supermarkets that cannot solve this problem may have a loss or may have to close temporarily. From this background, we propose a predictive analysis of sales of electronic and health goods in supermarkets using Linear Regression and Moving Average methods. The data used in this research were taken from the Kaggle dataset website. The data consists of several types of goods, but we use the electronics and health categories. The purpose of this research to analyze the two algorithms in predicting electronic sales and health data. The tool used to help analyze is python. How to analyze the two algorithms with MSE (Mean Square Error) and RMSE. The results of the analysis show that the Moving Average method has the best performance as evidenced by the MSE and RMSE values of electronic data which are 57,603, 7.59, health data are 50,489, 7,106. While the linear regression method prediction results on electronic data and health, the MSE and RMSE values are 114.79, 10.71, 59.965, 7.744.

Keywords: Algorithm Analysis, Moving Average, Linear Regression, Health Data, Electronic Data

1. Pendahuluan

Melakukan analisa data stok barang dengan metode Weight Moving Average dan Double Exponential Smoothing sangat penting [1]. Tujuan melakukan analisa agar tidak terjadi penyimpanan barang di gudang secara berlebihan, serta tidak sampai kekurangan stok di gudang. Jual beli barang kadang bisa ramai atau bisa juga sepi. Dalam kondisi pandemi Covid-19 penjualan di beberapa sektor mengalami penurunan serta bisa juga tidak laku. Peramalan penjualan salah satunya bertujuan untuk membantu menganalisa data penjualan tahun-tahun sebelumnya agar dapat memberikan informasi data

tahun berikutnya dan mengurangi tingkat ketidakpastian [2] [3].

Cara melakukan peramalan, analisa dan prediksi data penjualan dapat menggunakan beberapa metode, diantaranya Moving Average, Exponential Smoothin [1-3]. Metode Moving Average mempunyai beberapa macam dan memiliki tingkat prediksi yang baik. Misalnya jenis algoritma Moving Average adalah ARIMA (Autoregressive Integrated Moving Average) yang digunakan untuk memprediksi harga saham [4]. Metode Moving Average dalam penelitian Prapoyo juga dapat memberikan hasil prediksi jumlah mahasiswa yang baik [5]. Beberapa penelitian terkait peramalan

penjualan sering menggunakan metode Moving Average, ARIMA, Exponential Smoothing, dan juga membandingkan dari macam-macam metode Moving Average [6-10]. Penelitian Rahmadayanti membandingkan metode ARIMA dan Exponential Smoothing untuk memprediksi harga semen. Hasil penelitian Rahmadayanti bahwa algoritma ARIMA nilai MSE dari prediksi lebih kecil dibandingkan Exponential [6]. Penelitian lainnya menunjukkan bahwa algoritma Moving Average juga baik hasil peramalannya, terbukti dari prediksi produksi Kelapa Sawit nilai errornya 7.35% [7].

Prediksi data selain menggunakan Moving Average juga dapat dilakukan dengan Regresi Linear [11-13], Neural Network [13] atau Logika Fuzzy [13]. Metode Regresi Linear pada penelitian Fraticasari menunjukkan hasil prediksi dengan nilai error 1.5% [11]. Metode Regresi Linear saat dibandingkan dengan Logika Fuzzy atau Neural Network juga lebih baik Regresi Linear dalam memprediksi dengan tingkat error 0.1% [13].

Dari penelitian-penelitian sebelumnya bahwa penjualan yang akan datang dapat diprediksi dengan cara menganalisa data tahun sebelumnya. Tujuan melakukan prediksi data penjualan agar dapat mengantisipasi adanya tumpukan barang di gudang serta mengurangi kekurangan stok barang di gudang. Cara menganalisa data agar dapat memprediksi data penjualan dengan beberapa metode. Metode yang paling banyak digunakan serta memiliki nilai error yang rendah adalah Moving Average [1-3][5][7-9] dan Regresi Linear [11-13]. Dari permasalahan tersebut kami mengusulkan penelitian terkait analisa algoritma Regresi Linear dan Moving Average untuk memprediksi data penjualan di supermarket.

2. Metode Penelitian

2.1 Dataset

Data yang digunakan dalam penelitian ini diambil dari Kaggle dataset. Data yang digunakan mulai dari bulan Januari sampai Maret 2019, dan hanya kategori elektronik dan kesehatan. Data kesehatan awalnya sebanyak 152 baris seperti dalam Tabel 1. Tabel 1 kolom pertama adalah tanggal, bulan dan tahun, kolom kedua adalah jumlah penjualan di tanggal tersebut. Data elektronik awalnya sebanyak 170 baris, dengan model tabel dan kolom seperti Tabel 1.

Tabel 1. Contoh Dataset Awal Kategori Kesehatan

Date	Quantity
1/1/2019	2
1/2/2019	10
1/3/2019	5
1/3/2019	9
1/3/2019	2

1/3/2019	1
1/3/2019	8
1/4/2019	5
1/5/2019	7
1/5/2019	7
1/6/2019	9
1/10/2019	2
1/12/2019	2
1/12/2019	1
1/13/2019	9
1/14/2019	3
1/14/2019	1
1/14/2019	7
1/15/2019	10

Data dari masing-masing kategori elektronik dan kesehatan diolah terlebih dahulu. Pengolahan data terkait menjumlahkan kolom quantity sesuai tanggal yang sama. Misalkan data di Tabel 1 tanggal 12 Januari 2019 ada dua baris selanjutnya dijumlah menjadi 1 baris dengan jumlah 3 pada tanggal 12 Januari 2019. Misalkan data di Tabel 1 tanggal 14 Januari 2019 ada tiga baris dijumlahkan menjadi 1 baris dengan jumlah 11 pada tanggal tersebut. Hasil dari pengolahan data dari masing-masing kategori menjadi 73 baris data kesehatan, 75 baris data elektronik.

Data yang digunakan dalam metode Moving Average setiap kategori adalah 73 baris data kesehatan dan 75 baris data elektronik. data yang digunakan dalam metode Regresi Linear seperti dalam Tabel 2. Tabel 2 data masing-masing kategori kesehatan dan elektronik dibagi menjadi dua, yaitu training 85% dari total data seluruhnya dan testing 15% dari total data seluruhnya.

Tabel 2. Pembagian Dataset Metode Regresi Linear

Data	Training	Testing	Total
Kesehatan	62	11	73
Elektronik	64	11	75
Total	126	22	148

2.2 Penelitian yang Diusulkan

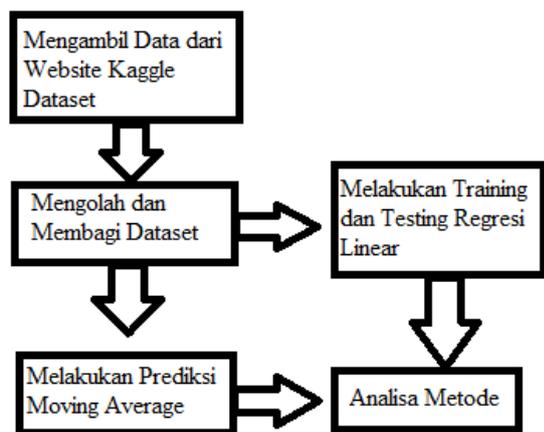
Penelitian ini mengusulkan analisa algoritma Regresi Linear dan Moving Average dalam memprediksi penjualan data kesehatan dan elektronik di supermarket. Data yang digunakan sebagai penelitian adalah data yang diambil dari website Kaggle Dataset. Usulan penelitian ini digambarkan dalam Gambar 1. Data awal masing-masing kategori kesehatan dan elektronik

diolah dengan dijumlahkan sesuai tanggal yang sama.

Metode Moving Average akan memprediksi data penjualan kategori kesehatan dan elektronik berdasarkan data penjualan tahun sebelumnya atau tanggal sebelumnya. Persamaan 1 merupakan proses perhitungan mencari prediksi penjualan. Metode Moving Average akan memprediksi data berdasarkan jumlah data atau window. Dalam penelitian ini Kami menggunakan prediksi per window 3.

$$\hat{x}_t = \frac{x_{t-1} + x_{t-2} + x_{t-3}}{3} \quad (1)$$

Keterangan Persamaan 1, \hat{x}_t data hasil prediksi priode t. $x_{t-1}, x_{t-2}, x_{t-3}$ merupakan data pada priode mundur ke-1, ke-2, dan ke-3.



Gambar 1. Diagram Proses Penelitian

Metode Regresi Linear dalam kasus ini akan memprediksi jumlah penjualan data elektronik dan kesehatan. Metode Regresi Linear kasus ini menggunakan tiga variabel. Variabel tanggal, bulan dan jumlah. Variabel dependen (terikat) atau yang akan diprediksi adalah jumlah. Variabel independen (bebas) atau variabel input adalah tanggal, bulan. Data dari masing-masing variabel seperti dalam Tabel 3. Tabel 3 menunjukkan bahwa kolom tanggal, bulan merupakan variabel input atau independen, dan kolom jumlah adalah variabel dependen atau yang akan diprediksi.

Tabel 3. Contoh Dataset Metode Regresi Linear

tanggal	bulan	jumlah
1	1	2
2	1	10
3	1	25
4	1	5
5	1	14
6	1	9

10	1	2
1	2	4
2	2	1
4	2	9
5	2	22
6	2	24
7	2	18
8	2	7
1	3	5
2	3	10
3	3	14
4	3	24
5	3	31
6	3	1

Metode Regresi Linear mempunyai dua tahapan penting, tahapan pertama adalah melakukan training dengan tujuan menghasilkan persamaan garis seperti Persamaan 2. Persamaan 2 adalah persamaan garis yang dihasilkan dari prediksi, dan Persamaan 3 cara mendapatkan persamaan garis. Tahapan kedua adalah ujicoba

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad (2)$$

$$n\beta_0 + \beta_1 \sum_{i=1}^n x_{i1} + \beta_2 \sum_{i=1}^n x_{i2} = \sum_{i=1}^n y_i$$

$$\beta_0 \sum_{i=1}^n x_{i1} + \beta_1 \sum_{i=1}^n x_{i1}^2 + \beta_2 \sum_{i=1}^n x_{i1} x_{i2} = \sum_{i=1}^n x_{i1} y_i \quad (3)$$

$$\beta_0 \sum_{i=1}^n x_{i2} + \beta_1 \sum_{i=1}^n x_{i2} x_{i1} + \beta_2 \sum_{i=1}^n x_{i2}^2 = \sum_{i=1}^n x_{i2} y_i$$

Persamaan 2 menjelaskan \hat{y} adalah data hasil prediksi, β_0 bilangan konstanta, dan β_1, β_2 adalah nilai koefisien dari variabel x_1, x_2 . x_1, x_2 secara berurutan adalah variabel tanggal, bulan.

3. Hasil dan Pembahasan

Penelitian ini melakukan analisa dua metode, yaitu Regresi Linear dan Moving Average. Proses analisa, training dan prediksi data menggunakan alat bantu Machine Learning bahasa pemrograman Python. Data awal dari penelitian ini seperti dalam Tabel 1. Data selanjutnya diolah dijumlahkan sesuai tanggal yang sama, data Tabel 4 merupakan hasil pengolahan data Tabel 1. Tabel 4 akan menjumlahkan tanggal 3 Januari 2019 yang awalnya ada 5 baris menjadi 1 baris dengan jumlah 25.

Tabel 4. Data Hasil Pengolahan

Date	jumlah
1/1/2019	2
1/2/2019	10

1/3/2019	25
1/4/2019	5
1/5/2019	14
1/6/2019	9
1/10/2019	2
1/12/2019	3
1/13/2019	9
1/14/2019	11
1/15/2019	17

Contoh proses perhitungan prediksi data Tabel 4 dengan metode Moving Average Persamaan 1 dalam Tabel 5. Moving Average akan melakukan prediksi data pada periode keempat berdasarkan data periode ke-3, ke-2, dan ke-1. Tabel 5 menunjukkan jika akan memprediksi data tanggal 4 Januari 2019, maka akan menghitung rata-rata dari data tanggal 3 Januari 2019, 2 Januari 2019, dan 1 Januari 2019.

Dalam penelitian ini Kami melakukan analisa akurasi dengan Mean Square Error (MSE) dan RMSE. Persamaan 4 adalah cara menghitung MSE, dan Persamaan 5 cara menghitung RMSE.

$$MSE = \frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

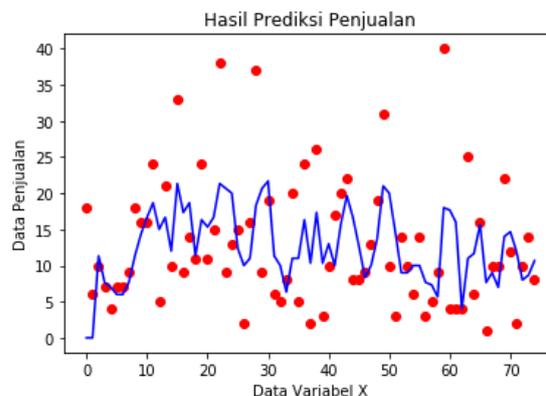
Tabel 5. Cara Memprediksi Data Kesehatan dengan Moving Average

date	jumlah	Prediksi MA-3
1/1/2019	2	0
1/2/2019	10	0
1/3/2019	25	0
1/4/2019	5	$=(2+10+25)/3=$ 12.33
1/5/2019	14	$=(10+25+5)/3=$ 13.33
1/6/2019	9	$=(25+5+14)/3=$ 14.67
1/10/2019	2	$=(5+14+9)/3=$ 9.33
1/12/2019	3	$=(14+9+2)/3=$ 8.33
1/13/2019	9	$=(9+2+3)/3=$ 4.67
1/14/2019	11	$=(2+3+9)/3=$ 4.67
1/15/2019	17	$=(3+9+11)/3=$ 7.67

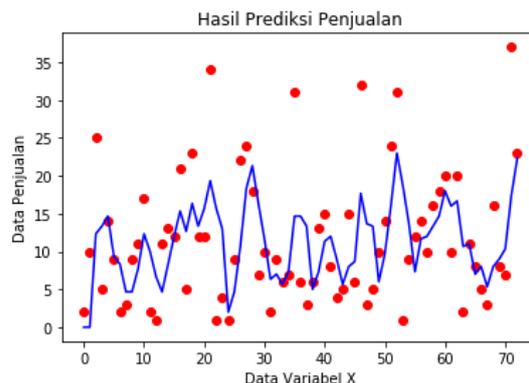
Tabel 6. Hasil Ujicoba Nilai MSE dan RMSE

Data	Nilai MA		Nilai RL	
	MSE	RMSE	MSE	RMSE
kesehatan	50.489	7.106	59.965	7.744
Elektronik	57.603	7.59	114.79	10.71

Penjelasan dari Persamaan 4 dan 5 adalah y adalah data aktual, dan \hat{y} adalah data hasil prediksi, N total data yang diujicoba. Hasil MSE dan RMSE dari metode Moving Average dan Regresi Linear kedua data dalam Tabel 6.



Gambar 2. Evaluasi Kinerja Moving Average Data Elektronik

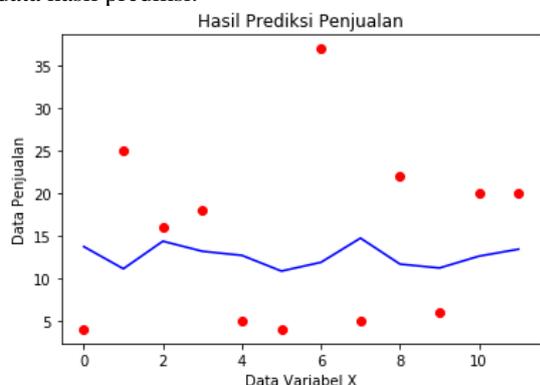


Gambar 3. Hasil Kinerja Prediksi Data Kesehatan Metode Moving Average

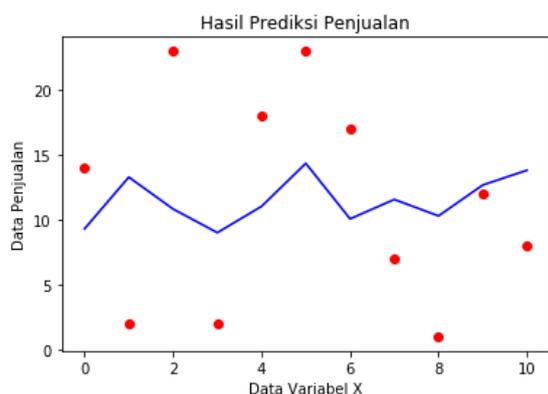
Tabel 6 kolom nilai MA merupakan hasil analisa error Moving Average kedua data, dan nilai RL merupakan hasil analisa error Regresi Linear. Hasil evaluasi dari kinerja metode Moving Average ada dalam Gambar 2 dan Gambar 3. Gambar 2 merupakan grafik hasil prediksi data elektronik, dan Gambar 3 merupakan grafik hasil prediksi data kesehatan.

Regresi Linear studi kasus ini akan melakukan training dengan menggunakan alat bantu atau Machine Learning bahasa pemrograman Python. Hasil dari training adalah persamaan garis, persamaan garis dari training data elektronik adalah $\hat{y} = 14.9265 + 0.09113 x_1 - 1.849 x_2$ dan persamaan garis dari training data kesehatan adalah $\hat{y} = 7.35 + 0.076 x_1 + 1.573 x_2$. Hasil kinerja metode Regresi Linear dalam memprediksi data elektronik dan kesehatan dalam Gambar 4 dan Gambar 5.

Penjelasan dari Gambar 2 sampai Gambar 5 adalah variabel x atau sumbu x merupakan nomor urutan data atau urutan tanggal, variabel y adalah data penjualan. Titik-titik merah merupakan data jumlah penjualan aktual dan garis biru merupakan data hasil prediksi.



Gambar 4. Hasil Kinerja Prediksi Data Elektronik Metode Regresi Linear



Gambar 5. Hasil Kinerja Prediksi Data Kesehatan Metode Regresi Linear

4. Kesimpulan

Hasil dari penelitian ini menunjukkan algoritma Moving Average lebih baik dalam memprediksi data kesehatan dan elektronik. Nilai MSE dan RMSE dari Moving Average data kesehatan adalah 50.489, 7.106. Nilai MSE dan RMSE dari Moving Average data elektronik adalah 57.603, 7.59. Hasil MSE dan RMSE dari metode Regresi Linear secara berurutan data kesehatan dan elektronik adalah 59.965, 7.744 dan 114.79, 10.71.

Daftar Pustaka:

[1] N. Nafi'iyah and A. W. Zannah, "Analisis Peramalan Stok Barang dengan Metode Weight Moving Average dan Double Exponential Smoothing pada Jovita Ms Glow Lamongan," *INSYST*, vol. 1, no. 1, pp. 39-42, 2019.

[2] C. Baktiar, A. Wibowo and R. Adipranata, "Pembuatan Sistem Peramalan Penjualan

Dengan Metode Weighted Moving Average dan Double Exponential Smoothing Pada UD Y," Universitas Kristen Petra, Surabaya.

[3] R. Rachman, "Penerapan Metode Moving Average dan Exponential Smoothing pada Peramalan Produksi Industri Garment," *JURNAL INFORMATIKA*, vol. 5, no. 1, pp. 211-220, 2018.

[4] D. Hatidja, "Penerapan Model ARIMA untuk Memprediksi Harga Saham PT. Telkom Tbk.," *Jurnal Ilmiah Sains*, vol. 11, no. 1, pp. 116-123, 2011.

[5] H. Prapcoyo, "Peramalan Jumlah Mahasiswa Menggunakan Moving Average," *TELEMATIKA*, vol. 15, no. 1, pp. 66-75, 2018.

[6] R. Rahmadayanti, B. Susilo and D. Puspitaningrum, "Perbandingan Keakuratan Metode Autoregressive Integrated Moving Average (ARIMA) dan Exponential Smoothing pada Peramalan Penjualan Semen di PT. Sinar Abadi," *Jurnal Rekursif*, vol. 3, no. 1, pp. 23-36, 2015.

[7] S. Agustian and H. Wibowo, "Perbandingan Metode Moving Average untuk Prediksi Hasil Produksi Kelapa Sawit," in *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri*, UIN Sultan Syarif Kasim Riau, 2019.

[8] A. Nurlifa and S. Kusumadewi, "Sistem Peramalan Jumlah Penjualan Menggunakan Metode Moving Average Pada Rumah Jilbab Zaky," *JURNAL INOVTEK POLBENG-SERI INFORMATIKA*, vol. 2, no. 1, pp. 18-25, 2017.

[9] S. S. Sundari, Susanto and W. Revianti, "Sistem Peramalan Persediaan Barang dengan Weight Moving Average di Toko The Kids 24," in *Konferensi Nasional Sistem & Informatika*, STMIK STIKOM Bali, 2015.

[10] K. F. Azriati, A. Hoyyi and M. A. Mukid, "Verifikasi Model ARIMA Musiman Menggunakan Peta Kendali Moving Range (Studi Kasus: Kecepatan Rata-rata Angin di Badan Meteorologi Klimatologi dan Geofisika Stasiun Meteorologi Maritim Semarang)," *JURNAL GAUSSIAN*, vol. 3, no. 4, pp. 701-710, 2014.

[11] S. Y. Fraticasari, D. E. Ratnawati and R. C. Wihandika, "Optimasi Pemodelan Regresi Linier Berganda Pada Prediksi Jumlah Kecelakaan Sepeda Motor dengan Algoritme Genetika," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, pp. 1932-1939, 2018.

[12] N. Nafi'iyah, "Penerapan Regresi Linear dalam Memprediksi Harga Jual Mobil Bekas," in *SENABAKTI*, Surabaya, 2015.

- [13] N. Nafi'iyah, "Perbandingan Regresi Linear, Backpropagation dan Fuzzy Mamdani dalam Memprediksi Harga Emas," in *SENIATI*, Malang, 2016.