

Penerapan Algoritma Regresi Linear untuk Prediksi Produksi Padi dengan *Mean Absolute Error* (MAE)

Andik Adi Suryanto¹, Asfan Muqtadir²

^{1,2}Jurusan Teknik Informatika, FT, Universitas PGRI Ronggolawe
Email: ¹andikadisuryanto@gmail.com, ²asfanme@gmail.com

Abstrak

Padi adalah salah satu kebutuhan pokok untuk memenuhi kebutuhan karbohidrat bagi penduduk. Dengan meningkatnya pertumbuhan penduduk tiap tahunnya dan kegiatan sosial ekonomi yang menyertainya kebutuhan produksi padi makin meningkat pula berbanding lurus jumlah penduduk dan kegiatan ekonomi. Regresi linear merupakan suatu cara mengukur data prediksi melalui garis lurus sebagai gambaran hubungan korelasi diantara 2 variabel atau lebih. Regresi linear digunakan sebagai teknik mempelajari bagaimana hubungan variabel-variabel pada proses peramalan data. *Mean Absolute Error* (MAE) adalah dua diantara banyak metode untuk mengukur tingkat keakuratan suatu model peramalan. Nilai MAE merepresentasikan rata – rata kesalahan (*error*) absolut antara hasil peramalan dengan nilai sebenarnya. Dengan menggunakan algoritma regresi linear dapat memberikan nilai prediksi produksi padi dengan 2 variabel jumlah pertumbuhan penduduk dan jumlah produksi padi pertahun, sedangkan keakuratan dari hasil perhitungan prediksi menggunakan metode *Mean Absolute Error* (MAE) yang digunakan untuk mengukur tingkat keakuratan suatu model peramalan.

Kata Kunci: Produksi, padi, regresi linear, *Mean Absolute Error*, *Mean Absolute Percentage Error*

Abstract

Rice is one of the basic needs to meet carbohydrate needs for the population. With increasing population growth each year and the socioeconomic activities that accompany the need for rice production is also increasing directly proportional to the population and economic activity. Linear Regression is a way of measuring predictive data through a straight line as an illustration of the correlation relationship between 2 or more variables. Linear regression is used as a technique to study how variables relate to the data forecasting process. Mean Absolute Error (MAE) are two of the many methods to measure the accuracy of a forecasting model. The MAE value represents the average absolute error between the forecasting results and the actual value. By using a linear regression algorithm can provide predictive value of rice production with 2 variables the amount of population growth and the amount of rice production per year, while the accuracy of the results of prediction calculations using the Mean Absolute Error (MAE) method that is used to measure the accuracy of a forecasting model.

Keywords: Production, rice, linear regression, *Mean Absolute Error*, *Mean Absolute Percentage Error*

1. PENDAHULUAN

Regresi linear merupakan metode yang digunakan untuk mengukur 2 variabel atau lebih, cara mengukur datanya dengan menggunakan variabel *dependen* dan variabel *independent*. Korelasi antar 2 variabel tersebut melalui garis lurus sebagai gambaran hubungan korelasinya [1].

Padi merupakan salah satu kebutuhan pangan yang digunakan untuk memenuhi kebutuhan karbohidrat bagi penduduk. Dengan perkembangan jumlah penduduk meningkat tiap tahun, maka kebutuhan produksi meningkat pula berbanding lurus dengan jumlah pertumbuhan penduduk [3].

Berdasarkan permasalahan tersebut peneliti menggunakan salah satu metode peramalan yaitu regresi linear untuk memprediksi kebutuhan produksi padi. Untuk keakuratan hasil prediksi menggunakan *Mean Absolute Error* dan *Mean Absolute Error Percentage Error*.

2. METODE

2.1 Regresi Linear

Fungsi Prediksi merupakan bagian dari fungsi minor data mining. Prediksi adalah suatu cara untuk mencari kemungkinan hasil dari suatu hal pada masa akan datang. Pada teknologi data mining, metode peramalan digunakan sebagai alat bantu memprediksi kemungkinan akan datang berdasarkan bukti-bukti temuan pada data. Proses prediksi berhubungan erat dengan perhitungan matematik dan statistik.

Langkah prediksi dapat dilakukan menggunakan perhitungan rumus *regresi linear*. Ada 2 jenis rumus *regresi* sebagai langkah proses analisis prediksi yaitu, *regresi linear* sederhana dan *regresi linear* berganda. Cara membaca hubungan korelasi data hipotesis hasil analisis prediksi dilakukan melalui analisis statistik.

Regresi Linear merupakan suatu cara mengukur data prediksi melalui garis lurus sebagai gambaran hubungan korelasi diantara 2 variabel atau lebih [1]. Prediksi *regresi linear*, digunakan sebagai teknik mempelajari bagaimana hubungan variabel-variabel pada proses peramalan data.

Variabel adalah besaran yang berubah-ubah nilainya. Ada 2 tipe variabel dalam *regresi linear* yaitu: variabel pemberi pengaruh dan variabel terpengaruh. Variabel pemberi pengaruh digunakan sebagai sebab, sedangkan variabel terpengaruh sebagai akibat.

2.2 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah dua diantara banyak metode untuk mengukur tingkat keakuratan suatu model peramalan. Nilai MAE merepresentasikan rata – rata kesalahan (*error*) absolut antara hasil peramalan dengan nilai sebenarnya [4].

Secara matematis MAE didefinisikan sebagai berikut,

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (1)$$

dimana:

f_i adalah nilai hasil peramalan,

y_i adalah nilai sebenarnya, dan

n adalah jumlah data.

Berdasarkan formula 1 di atas, MAE secara intuitif menghitung rata – rata *error* dengan memberikan bobot yang sama untuk seluruh data ($i = 1 \dots n$). Untuk evaluasi model peramalan, MAE lebih intuitif dalam memberikan rata – rata *error* dari keseluruhan data. Dalam kasus ini pemilihan MAE menjadi tepat karena memberikan bobot yang sama untuk seluruh data.

Dalam kasus klasifikasi biner, dimana hanya terdapat dua kelas dengan label kelas 0 dan 1, dalam penggunaan MAE. Hal ini karena nilai *error* hanya mempunyai dua kemungkinan, 0 jika prediksi benar dan 1 jika prediksi kelas berbeda dengan kelas sebenarnya.

2.3 Mean Absolute Percentage Error (MAE)

Mean Absolute Percentage Error digunakan untuk mengindikasikan kesalahan dalam peramalan dengan hasil model peramalan dibandingkan dengan nilai aktual [5]. *Mean absolute percentage Error* (MAPE) penggunaan dengan menghitung kesalahan absolut pada tiap periode dibagi dengan nilai observasi yang nyata untuk periode itu. Lalu melakukan rata-rata kesalahan persentase absolut tersebut. Metode ini berguna jika besar atau ukuran variabel ramalan itu penting dalam mengevaluasi ketepatan ramalan.

$$MAPE = \frac{\sum \frac{|e_i|}{X_i} \times 100\%}{n} = \frac{\sum \frac{|X_i - F_i|}{X_i} \times 100\%}{n} \quad (2)$$

Tabel klasifikasi keakuratan dari MAPE dapat disajikan pada Tabel 2.1.

<i>MAPE</i>	Daya perkiraan
< 0,1	Sangat akurat
0,1 – 0,2	Akurat
0,2 – 0,5	Kurang
> 0,5	Tidak Akurat

2.4 Padi

Padi merupakan bahan pokok sehari-hari pada mayoritas kebanyakan di negara Indonesia. Pada dikenal sebagai sumber karbohidrat. Bagi penduduk Indonesia padi merupakan makanan pokok dan tidak dapat digantikan oleh bahan makanan yang lain[2]. Komoniditas pemenuhan kaborhidrat bagi penduduk dapat diperoleh pada hasil pertanian yaitu padi yang terus meningkat tiap tahunnya akibat dari peningkatan jumlah penduduk serta berkembangnya kegiatan ekonomi dan sosial [3].

3. HASIL DAN PEMBAHASAN

3.1 Hasil Regresi Linear

Penyelesaian penggunaan regresi linear sederhana untuk memprediksi kebutuhan produksi padi seperti tahapan dibawah ini:

1. Penentuan tujuan
Penggunaan metode regresi linear sederhana ini untuk memprediksi jumlah produksi padi.
2. Identifikasi variabel penyebab dan akibat
Dari data yang diperoleh digolongkan menjadi 2 variabel faktor penyebab (X) dan variabel akibat (Y) sebagai berikut:
Variabel faktor penyebab (X) = Jumlah penduduk
Variabel akibat (Y) = Jumlah Produksi Padi
3. Pengumpulan data
Data yang digunakan mulai tahun 2000 sampai 2017, disajikan pada Tabel 2.

Tabel 2. Tabel Jumlah Penduduk dan Produksi Padi

Tahun	Jumlah Penduduk Registrasi (X)	Jumlah Produksi Padi Dalam Kuintal (Y)
2000	1021920	3898622
2001	1027546	3802048
2002	1035341	3849083
2003	1058979	3820119
2004	1042718	3959801
2005	1074109	4075379
2006	1124508	4075390
2007	1127416	4585170
2008	1137708	4543510
2009	1139052	4710640
2010	1259996	5115100
2011	1258816	5176300
2012	1290394	5890660
2013	1288975	4888390
2014	1291665	5376630
2015	1304080	5463100
2016	1315155	5843060

Tahun	Jumlah Penduduk Registrasi (X)	Jumlah Produksi Padi Dalam Kuintal (Y)
2017	1267886	5894210

4. Model persamaan regresi linear sederhana

$$Y = a + bX \quad (3)$$

5. Hasil prediksi terhadap variabel penyebab atau akibat produksi padi tiap tahun dari tahun 2000 - 2017 menggunakan regresi linear sederhana dapat dilihat pada Tabel 3.

Tabel 3. Hasil Prediksi pertahun

Tahun	Jumlah Penduduk Registrasi (X)	Jumlah Produksi Padi Dalam Kuintal (Y)	Nilai Prediksi Produksi Padi dalam Kuintal
2000	1021920	3898622	0
2001	1027546	3802048	3802048
2002	1035341	3849083	3828080
2003	1058979	3820119	3814568
2004	1042718	3959801	3863785
2005	1074109	4075379	3996009
2006	1124508	4075390	4092447
2007	1127416	4585170	4320132
2008	1137708	4543510	4433516
2009	1139052	4710640	4514351
2010	1259996	5115100	5178435
2011	1258816	5176300	5173376
2012	1290394	5890660	5539802
2013	1288975	4888390	5368272
2014	1291665	5376630	5382631
2015	1304080	5463100	5457309
2016	1315155	5843060	5577583
2017	1267886	5894210	5347763

Tabel 3 merupakan hasil prediksi dari persamaan regresi linear. Contohnya pada tahun 2017 hasil prediksi 5347763 kuintal hari produksi padinya.

3.2 Hasil *Mean Absolute Error* (MAE) dan *Mean Absolute Percentage Error* (MAPE)

Penggunaan *Mean Absolute Error* dan *Absoluter Percentage Error* dalam penelitian ini untuk menguji hasil prediksi dengan menggunakan regresi linear sederhana. Adapun hasil perhitungan dapat dilihat pada Tabel 4.

Tabel 4. Hasil MAE dan MAPE

No	Tahun	Jumlah Penduduk Registrasi (X)	Jumlah Produksi Padi Dalam Kuintal (Y)	Nilai Prediksi (y1)	(y-y1)	MAE (y-y1 / y)
1	2000	1021920	3898622	0	3898622	1,00000
2	2001	1027546	3802048	3802048	0	0,00000
3	2002	1035341	3849083	3828080	21003	0,00546
4	2003	1058979	3820119	3814568	5551	0,00145
5	2004	1042718	3959801	3863785	96016	0,02425
6	2005	1074109	4075379	3996009	79370	0,01948
7	2006	1124508	4075390	4092447	17057	0,00419
8	2007	1127416	4585170	4320132	265038	0,05780
9	2008	1137708	4543510	4433516	109994	0,02421
10	2009	1139052	4710640	4514351	196289	0,04167
11	2010	1259996	5115100	5178435	63335	0,01238
12	2011	1258816	5176300	5173376	2924	0,00056
13	2012	1290394	5890660	5539802	350858	0,05956
14	2013	1288975	4888390	5368272	479882	0,09817
15	2014	1291665	5376630	5382631	6001	0,00112
16	2015	1304080	5463100	5457309	5791	0,00106
17	2016	1315155	5843060	5577583	265477	0,04543
18	2017	1267886	5894210	5347763	546447	0,09271
Total MAE						1,48950
MAPE						0,08762

Pada Tabel 4 diperoleh hasil perhitungan *Mean Absolute Error* (MAE) dengan nilai 1,48950 sedangkan nilai *Mean Absolute Percentage Error* (MAPE) dengan nilai 0,08762. Berdasarkan pada Tabel 1 klasifikasi keakuratan *Mean Absolute Percentage Error* (MAPE) jadi hasil pengajuan keakuratan dipeoleh hasil sangat akurat dengan nilai 0,08762.

4. SIMPULAN

Dengan menggunakan metode regresi linear sederhana dapat memperoleh hasil perhitungan produksi padi pada tahun 2017 prediksi produksi padi sebesar 5347763 kuintal. Pengujian keakuratan peramalan atau prediksi metode regresi linear menggunakan metode *Mean Absolute Error* (MEA) dan *Mean Absolute Percentage Error* (MAPE). Hasil perhitungan *Mean Absolute Percentage Error* (MAPE) sebesar 0,08762.

5. REFERENSI

- [1] Susanto, S., Suryadi, D. 2010. *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. CV ANDI Offset, Yogyakarta.
- [2] Aak. 1990. *Budidaya Tanaman Padi*. Kanisius, Yogyakarta.
- [3] Yusuf, A. 2010. *Teknologi Budidaya Pada Sawah Mendukung SL-PTT.BPTP, Sumatera Utara*.
- [4] Subagyo, P. 1986. *Forecasting Konsep dan Aplikasi*. Yogyakarta, BPPE UGM.
- [5] Pakaja. 2012. Peramalan Penjualan Mobil Menggunakan Jaringan Syaraf Tiruan dan Certainty Factor. *EECCIS*. Vol. 6(1).
- [6] Shang, L. 2012. Forecasting Agricultural Output with an Improved Grey Forecasting Model Based on the Genetic Algorithm. *Journal of Computers and Electronics in Agriculture*. Vol. 85: 33–39.