

## Metode Regresi Linier Berganda dan SVR dalam Menentukan Tingkat Pengaruh Cuaca Terhadap Produktivitas Padi di Indonesia

Risma Yulistiani<sup>1</sup>, Ilham Ramadhan<sup>2</sup>, Qahtan Said<sup>3</sup>, Mayanda Mega Santoni<sup>4</sup>

Fakultas Ilmu Komputer  
Universitas Pembangunan “Veteran” Jakarta, DKI Jakarta  
email: yulistianir@gmail.com

Jl. RS Fatmawati, Pondok Labu, Cilandak, Jakarta Selatan, DKI Jakarta 12450, Indonesia

### Abstrak

Kebutuhan masyarakat Indonesia terhadap beras yang dihasilkan dari padi menjadikan produktivitas padi tiap tahunnya penting untuk diperhatikan. Salah satu masalah yang menyangkut produktivitas padi di Indonesia adalah perubahan cuaca yang tidak menentu. Untuk mengatasi permasalahan tersebut, dilakukan penelitian untuk memprediksi produktivitas padi menggunakan Metode Regresi Linier Berganda dan SVR (*Support Vector Regression*). Data pengujian adalah data sektoral dari situs BPS tahun 2000-2015. Pemilihan atribut dibatasi pada data yang memungkinkan memiliki hubungan antara pengaruh cuaca dan produktivitas padi, yaitu curah hujan, kelembapan udara dan kecepatan angin. Hasil pengujian menunjukkan SVR dan Regresi Linier memiliki tingkat akurasi yang hampir sama. Dari hasil evaluasi Regresi Linier menunjukkan tingkat galat sebesar RMSE (*Root Mean Square Error*) sebesar 26.4763 dan MAE (*Mean Absolut Error*) sebesar 21,7043, sedangkan SVR menunjukkan tingkat galat RMSE sebesar 25.8503 dan MAE sebesar 20,9812.

Kata Kunci: Regresi Linier Berganda, SVR, Prediksi, Produktifitas Padi.

### 1. PENDAHULUAN

Sumber pangan merupakan salah satu faktor penting dalam kehidupan. Bagi Indonesia sebagai negara agraris, tidak hanya sebagai pemenuh kebutuhan pokok sehari-hari, namun sumber daya pangan juga mempengaruhi perekonomian masyarakat. Salah satu sumber pangan yang selalu dibutuhkan masyarakat adalah padi. Dalam kenyataannya, meskipun Indonesia memiliki banyak sektor pertanian, banyak faktor yang mempengaruhi tingkat produktivitas padi di Indonesia. Pergantian cuaca yang tidak stabil menjadi salah satu masalah pangan terpenting di Indonesia. Beberapa tahun belakangan, perubahan iklim terjadi sangat ekstrim, sehingga berpengaruh pada produktivitas padi di Indonesia.

Menurut Hosang, P.R., dkk. (2012) Perubahan dan anomali iklim mempengaruhi kemampuan dan dinamika produksi pertanian. Perubahan iklim yang dominan mempengaruhi ketahanan pangan, yaitu pergeseran musim hujan dan kemarau yang sangat mempengaruhi pola dan waktu tanam tanaman semusim yang umumnya ialah tanaman pangan. Dalam kasus ini, pengaruh cuaca terhadap produktivitas padi, diantaranya disebabkan oleh curah hujan, kecepatan angin dan kelembapan udara.

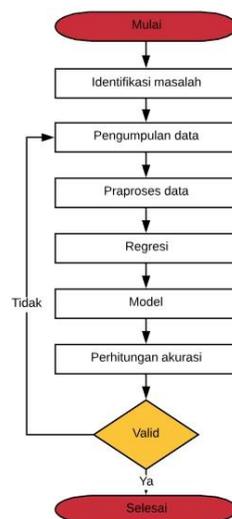
Berbagai cara dilakukan pemerintah untuk meningkatkan produktivitas padi di Indonesia, namun keterbatasan teknologi mengakibatkan masih kurangnya penanganan yang memadai untuk meningkatkan produktivitas padi. Data mining, dapat dimanfaatkan untuk menangani permasalahan tersebut, yaitu dengan metode prediksi menggunakan regresi linier. Prediksi produktivitas padi menggunakan metode regresi linier diharapkan mampu menganalisis pengaruh antara variabel-variabel cuaca dengan produktivitas padi. Menurut Petrus Katemba dan Rosita Koro Djoh (2017) Manfaat dari regresi linier diantaranya analisis regresi lebih akurat dalam melakukan analisis

korelasi, karena dengan analisis ini kesulitan dalam menunjukkan tingkat perubahan suatu variabel terhadap variabel lainnya (slop) dapat ditentukan. Dengan analisis regresi peramalan atau perkiraan nilai variabel terikat pada nilai variabel bebas lebih akurat.

Tujuan melakukan prediksi pada produktivitas padi untuk melihat berapa besar pengaruh cuaca terhadap produktivitas padi, berdasarkan perubahan cuaca yang terjadi di Indonesia. Menurut Hendy Tannady dan Fan Andrew (2013), secara rata-rata, metode peramalan terbaik adalah regresi linier, namun apabila diturunkan kedalam kategori varian, akan terlihat bahwa setiap varian data memiliki metode peramalan terbaik.

## 2. METODOLOGI PENELITIAN

Pada penelitian ini, menggunakan beberapa tahapan yaitu proses pengumpulan data, praproses data, menentukan variabel dependen dan independen, serta proses menganalisis data menggunakan WEKA (*Waikato Environment of Knowledge Analysis*). Bagan rancangan penelitian, dapat dilihat pada Gambar 1.



Gambar 1: Flowchart

### 2.1 Identifikasi Masalah

Penelitian ini mengangkat masalah pengaruh faktor cuaca untuk memprediksi produktivitas padi di Indonesia menggunakan metode regresi linier berganda dan SVR.

### 2.2 Proses Pengumpulan Data

Pada proses ini, data dikumpulkan melalui situs resmi Badan Pusat Statistik (<http://www.bps.go.id>). Data yang digunakan meliputi 3 atribut faktor cuaca yang mempengaruhi produktivitas padi di Indonesia. Data yang diambil berdasarkan data sektoral dari pangkalan BMKG. Data diambil berdasarkan setiap provinsi di Indonesia dalam kurun waktu tahun 2000-2015. Data asli kemudian dikumpulkan dalam satu file berbentuk excel dan csv. Keterangan atribut dapat dilihat pada Tabel 1.

Tabel 1: Atribut dan Kelas

No	Atribut	Kode
1	Curah Hujan per tahun berdasarkan data pangkalan BMKG	X1
2	Kecepatan Angin berdasarkan (satuan Knot)	X2
3	Kelembapan Udara	X3
4	Persentase Produktivitas Padi	Y

### 2.3 Praproses Data

Praproses dimulai dengan proses *cleaning* data. Proses *cleaning* dilakukan untuk membersihkan noise dan *missing value*. Data tersebut kemudian di normalisasi, tahap ini dilakukan untuk menyetarakan nilai *value* pada setiap atribut dan kelas. Pada proses ini, didapatkan data sebanyak 462 data *training* dan 33 data *testing*.

Proses *Cleaning* dan *missing value* data dilakukan dengan membuang data yang tidak memiliki *value*, termasuk semua data pada tahun 2004 karena tidak ditemukan data pada tahun tersebut.

Setelah proses diatas, normalisasi dilakukan dengan metode Min-Max, metode ini dipilih karena paling sesuai dengan ketersediaan data. Adapun batasan nilai dari Min-Max baru adalah 0-100. Range ini didapat dari ukuran sebaran data setiap atribut yang dapat dilihat pada Tabel 2.

Tabel 2: Mean dan Nilai Min-Max Data

	X1	X2	X3	Y
Nilai Mean	2 196.42	5.49	78.88	42.17
Nilai Min	1.50	0.07	54.20	22.44
Nilai Max	5652.00	19.25	97.80	61.88

### 2.4 Pembagian Data

Pada tahap ini, data dibagi menjadi dua bagian, yaitu data latih (*data training*) dan data uji (*data testing*). Data *training* merupakan data yang digunakan sebagai data yang nantinya akan dipelajari menggunakan regresi sebagai model data prediksi. Sedangkan, data *testing* digunakan untuk menguji model prediksi dari data *training*. Pembagian dilakukan dengan metode *K-Cross Fold Validation*. Menurut Refaeilzadeh et al. (2008), *K-Cross Fold Validation* merupakan sebuah metode membagi himpunan contoh secara acak menjadi *k* himpunan bagian (subset).

### 2.5 Pemodelan Data

Data yang telah dibagi, kemudian dimodelkan menggunakan regresi linier berganda dan *SVR* (*Support Vector Regression*). Tujuan dari pemodelan data adalah untuk melihat hubungan keterkaitan antara atribut dan kelas pada data yang akan diprediksi sebelum masuk proses pengujian data. Analisis regresi linier berganda adalah salah satu metode statistika yang digunakan untuk mengetahui hubungan fungsional sebuah variabel tidak bebas (*dependent variable*) dengan dua atau lebih variabel bebas (*independent variable*), Neter (1997). Model regresi linier berganda dapat dihitung melalui persamaan berikut.

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n$$

Sedangkan, Ide dasar dari *SVR* adalah memetakan data ke dalam ruang vector berdimensi tinggi melalui pemetaan non linier dan melakukan regresi linier dalam ruang tersebut. Metode ini juga dapat diaplikasikan langsung ke dalam regresi, dengan menjaga karakteristik algoritma margin maksimal. Fungsi nonlinier dikembangkan dengan menggunakan fungsi kernel yang membentuk ruang vektor berdimensi tinggi dimana sistem dikontrol melalui parameter yang tidak tergantung pada dimensi ruang vektor. Dalam WEKA, *SVR* memiliki fungsi sebagai *SMOreg*.

### 2.6 Pengujian Data

Setelah proses pemodelan data, model data diuji keakuratan dan tingkat eror prediksi terhadap data sebenarnya. Tujuannya melihat seberapa berpengaruh setiap atribut dengan prediksi yang dilakukan. Kemudian dilakukan proses evaluasi untuk melihat berapa tingkat galat dan akurasi

penelitian yang dilakukan, dengan *Root Mean Square Error (RMSE)* dan *Mean Absolut 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.

### 3. HASIL DAN PEMBAHASAN

Bentuk perbandingan hasil prediksi regresi linier dan SVR terhadap nilai aktual disajikan pada Tabel 3. Prediksi produktivitas padi tahun 2015 dengan menggunakan variabel independen curah hujan, kecepatan angin dan kelembapan udara.

*Tabel 3: Perbandingan Nilai Prediksi*

Provinsi	Tahun	Y (nilai sebenarnya)	Y' (Regresi Linier)	Y' (SVR)
Aceh	2015	70.53	36.88	33.16
Sumatera Utara	2015	73.53	29.26	41.11
Sumatera Barat	2015	69.74	16.10	45.30
Riau	2015	35.07	31.33	36.45
Jambi	2015	54.62	34.19	35.20
Sumatera Selatan	2015	65.72	31.57	36.37
Bengkulu	2015	56.17	27.98	37.99
Lampung	2015	72.89	43.73	27.66
Kep. Bangka Belitung	2015	0.00	28.44	39.71
Kep. Riau	2015	34.64	21.86	43.70
DKI Jakarta	2015	84.25	52.71	19.02
Jawa Barat	2015	97.66	47.78	22.79
Jawa Tengah	2015	95.19	55.24	17.80
DI Yogyakarta	2015	96.21	46.09	25.40
Jawa Timur	2015	97.43	35.75	32.49
Banten	2015	85.93	51.23	22.64
Bali	2015	100.00	37.57	33.47
Nusa Tenggara Barat	2015	73.45	33.02	37.21
Nusa Tenggara Timur	2015	32.48	37.63	32.43
Kalimantan Barat	2015	16.67	25.04	40.35
Kalimantan Tengah	2015	31.10	32.42	34.10
Kalimantan Selatan	2015	48.41	34.15	33.37
Kalimantan Timur	2015	46.70	39.45	30.08
Kalimantan Utara	2015	11.25	28.63	38.33
Sulawesi Utara	2015	66.68	42.64	27.72
Sulawesi Tengah	2015	65.46	60.56	16.54
Sulawesi Selatan	2015	75.24	34.13	30.79
Sulawesi Tenggara	2015	61.64	40.31	30.86
Gorontalo	2015	83.13	51.01	23.55
Sulawesi Barat	2015	67.60	49.94	23.69

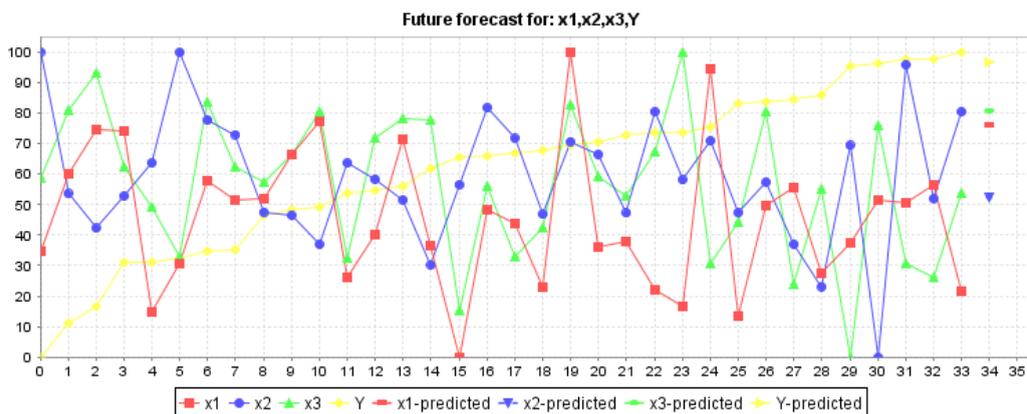
Maluku	2015	83.66	29.78	38.13
Maluku Utara	2015	31.20	44.89	28.24
Papua Barat	2015	49.05	29.99	36.12
Papua	2015	53.70	48.21	24.60

Merujuk pada Tabel diatas, diperoleh hasil koefisien korelasi antar variabel. Koefisien Korelasi atau *Correlation Coefficient* adalah nilai yang menunjukkan seberapa kuat tingkat hubungan atau korelasi antar variabel-variabel independen dengan variabel dependennya. Hasil koefisien korelasi setiap metode dtunjukkan pada Tabel 4.

Tabel 4: Nilai Koefisien Korelasi

Metode	Koefisien Korelasi
Regresi Linier	-0,3215
SVR	0,1145

Perbedaan antara prediksi menggunakan linier berganda dan SVR dapat dilihat pada gambar. Gambar 2 menjelaskan prediksi menggunakan Linier berganda dan SVR.



Gambar 1: Menggunakan Regresi Linier dan SVR

### 3.1 Evaluasi

Dari pegujian diatas dilihat kembali nilai galat dan efisiensi algoritma yang digunakan menggunakan RMSE (*Root Mean Square Error*) dan MAE (*Mean Absolute Error*) yang ditunjukkan pada Tabel 5. Nilai MAE menunjukkan nilai rata-rata galat dari keseluruhan data. Sedangkan RMSE menunjukkan nilai prediksi total dibandingkan nilai total aktualnya yang terjadi di seluruh Indonesia.

Tabel 5: Perbandingan Nilai Evaluasi

	Regresi Linier	SVR
MAE	21.7043	20.9812
RMSE	26.4763	25.8503

## 4. KESIMPULAN

Dari hasil yang diperoleh, dapat dilihat bahwa perbedaan yang didapat dari prediksi dan nilai aktualnya masih sangat tinggi, hal ini dikarenakan tingkat koefisien korelasinya. Tingkat koefisien korelasi ini dilihat berdasarkan seberapa kuat hubungan antar variabel. Dalam percobaan

menggunakan regresi linier, diperoleh koefisien korelasi sebesar -0,3215 atau sebesar -32,15%. Sedangkan metode SVR memiliki tingkat koefisien korelasi sebesar 0,1145 atau 11,45%. Hal tersebut menunjukkan metode SVR lebih baik karena memiliki korelasi antar variabel lebih besar. Metode memiliki tingkat rata-rata galat absolut (MAE) lebih rendah dibandingkan regresi linier, perbandingan nilai MAE antara keduanya yaitu 0.7231. Dan tingkat RMSE yang lebih rendah 0.6672 dibandingkan metode regresi linier. Hal ini menjelaskan bahwa metode SVR lebih baik digunakan dalam penelitian ini karena memiliki tingkat eror yang lebih rendah. Maka, dapat ditarik kesimpulan bahwa pengaruh cuaca dalam memprediksi produktivitas padi adalah 11,45%.

#### Referensi

- Hosang, Peter Rene, J. Tatuh, Johannes E.X. Rogi. 2012. "*Analisis Dampak Perubahan Iklim Terhadap Produksi Beras Provinsi Sulawesi Utara Tahun 2013-2030*". Vol. 8. Nomor 03. Eugenia
- Katamba, Petrus, Rosita Koro Djoh. 2017. "*Prediksi Tingkat Produksi Kopi Menggunakan Regresi Linier*". Vol. 03, Nomor 1. FLASH
- Tannady, Hendy, Fan Andrew. 2013. "*Analisis Perbandingan Metode Regresi Linier dan Exponential Smoothing dalam Parameter Tingkat Error*". Vol. 02. UKRIDA
- Soetanto & Maria, U.A. (2010). "*Modul Pelatihan Peningkatan Akurasi Prakiraan Musim*". Jakarta: BMKG.
- Badan Pusat Statistik ([www.bps.go.id](http://www.bps.go.id))