

Estimasi Ketersediaan Lahan Pertanian Beras Menggunakan Algoritma Regresi Linear

Evi Dewi Sri Mulyani ^{1,*}, Rijal Aji Jatnika ²

^{1,2} Teknik Informatika; STMIK Tasikmalaya; Jalan R.E. Martadinata No. 272A, Panyingkiran, Indihiang, Kota Tasikmalaya, Jawa Barat 46151; Email : eviajadech@gmail.com, rijaljatnika@yandex.com

* Korespondensi: email: eviajadech@gmail.com

Diterima: Diterima: 7 Oktober 2019; Review: 21 Oktober 2019; Disetujui: 4 November 2019

Cara sitasi: Mulyani EDS, Jatnika RA. 2019. Estimasi Ketersediaan Lahan Pertanian Beras Menggunakan Algoritma Regresi Linear. Information System For Educators and Professionals. 4 (1): 75 – 84.

Abstrak: Sektor pertanian memiliki multifungsi yang mencakup aspek produksi atau ketahanan pangan, peningkatan kesejahteraan petani atau pengentasan kemiskinan, dan menjaga kelestarian lingkungan hidup. Untuk memberikan dampak yang baik dalam pengembangan lahan agar pembagian lahan merata dan hasil produksi dari hasil pertanian menjadi baik diperlukan sebuah metode untuk mengukur ketersediaan lahan yang ada dan mampu membuat estimasi kedepannya dalam mempersiapkan lahan yang cukup untuk produksi yang lebih baik. Maka dibuatlah metode yang dapat memperkirakan luas lahan untuk produksi di tahun-tahun selanjutnya. Regresi Linear adalah salah satu metode pencari keputusan dan perhitungan untuk memodelkan hubungan antara variabel terkait Y dan satu variabel bebas yaitu X. fungsi dari regresi adalah melakukan prediksi berdasarkan data-data yang telah dimiliki sebelumnya. Dalam jurnal acuan sebelumnya penggunaan regresi linear untuk memprediksi pertumbuhan produksi sampah dan perkembangan penduduk dari tahun ke tahun. Dalam jurnal tersebut menggunakan 2 variabel sebagai acuannya. Penelitian ini telah dilakukan dengan regresi linear dimana hasil dari estimasi lahan pertanian nya adalah 685.547,51 M². Pengujiannya telah dilakukan dengan menggunakan aplikasi weka untuk melihat hasil akurasi dan prediksi serta melihat eror yang ditemui. Nilai eror yang didapatkan adalah 12,47% sehingga dapat disimpulkan bahwa pengujian ini memiliki tingkat akurasi yang baik.

Kata kunci: Data mining, Estimasi lahan, Regresi linier

Abstract: *The agricultural sector has a multifunctionality that covers aspects of production or food security, increasing the welfare of poverty or poverty alleviation, and protecting the environment. To provide a good impact on land development so that the division of land is evenly distributed and the production of agricultural products to be good requires a method to measure the availability of existing land and be able to make future estimates in preparing sufficient land for better production. Then made a method that can estimate the area of land for production in subsequent years. Linear Regression is a method of decision making and calculation to model the relationship between Y related variables and one independent variable, X. The function of regression is to make predictions based on data that has been previously owned. In the previous reference journals the use of linear regression was to predict the growth of waste production and population development from year to year. In the journal uses 2 variables as a reference. This research has been carried out with linear regression where the results of the estimated agricultural land is 685,547.51 M². The test has been carried out using the Weka application to see the results of accuracy and prediction and to see errors encountered. The error value obtained is 12.47% so it can be concluded that this test has a good level of accuracy.*

Keywords: Data mining, Land estimation, Linear regression

1. Pendahuluan

Sektor pertanian memiliki multifungsi yang mencakup aspek produksi atau ketahanan pangan, peningkatan kesejahteraan petani atau pengentasan kemiskinan, dan menjaga kelestarian lingkungan hidup. Bagi daerah yang mengutamakan pertanian sebagai mata pencahariannya perlu dipertimbangkan dalam penelitian kebijakan struktur intensif sektor pertanian. Pengembangan lahan pertanian abadi akan dapat diwujudkan jika sektor pertanian dengan nilai multifungsinya dapat memberikan manfaat bagi peningkatan kesejahteraan petani dan pengentasan kemiskinan [Warih, Eggy Inaidi Andana; Rahayu, 2014].

Untuk memberikan dampak yang baik dalam pengembangan lahan agar pembagian lahan merata dan hasil produksi dari hasil pertanian menjadi baik diperlukan sebuah metode untuk mengukur ketersediaan lahan yang ada dan mampu membuat estimasi kedepannya dalam mempersiapkan lahan yang cukup untuk produksi yang lebih baik. Maka dibuatlah metode yang dapat memperkirakan luas lahan untuk produksi di tahun-tahun selanjutnya. Dalam jurnal sebelumnya membahas tentang estimasi tempat pembuangan sampah dengan pertumbuhan penduduk sekitar selama 5 tahun dari 2012-2017 menggunakan regresi linear sederhana, yang dapat disimpulkan pertumbuhan penduduk naik hingga mencapai 0,735% dan dapat dijadikan acuan untuk memprediksi pertumbuhan penduduk pada tahun 2018-2025 sebagai data untuk mengetahui perkembangan volume sampah dari tahun 2018 sampai dengan 2025 sehingga dapat diperkirakan bahwa pada tahun 2025 penduduk tumbuh sebesar 201484 jiwa dan estimasi sampah adalah 36.052.326 ton dan lahan yang dibutuhkan mencapai 20.434,9 M². Dalam jurnal sebelumnya peneliti menggunakan 2 variabel sebagai pembanding, yaitu lahan dan total sampah yang diproduksi oleh penduduk dan menghasilkan persamaan linear yang bertemu pada kebutuhan lahan pembuangan sampah [Yanto, 2018].

Pada Penelitian Yang Lainnya Tentang Mengetahui Menggunakan Tingkat Kekuatan Beton Yang Dihasilkan Dengan Metode Estimasi Menggunakan Regresi Linear dimana disana terdapat 8 variabel perhitungan yaitu semen, biji besi, air, abu terbang, superplasticizer, batu krikil, pasir, dan umur. Semakin banyak variabel yang digunakan serta penambahan variabel pendukung dapat membuat hasil estimasi lebih tepat dan akurat. Dalam jurnal ini juga mempunyai 2 tujuan yang berbeda pertama, untuk menguji signifikansi nilai koefisien korelasi. Kedua yaitu menguji signifikansi regresi. Setiap peubah (x) yang mempunyai koefisien regresi yang tidak signifikan secara statistik harus dibuang dari model [Fikri, 2013].

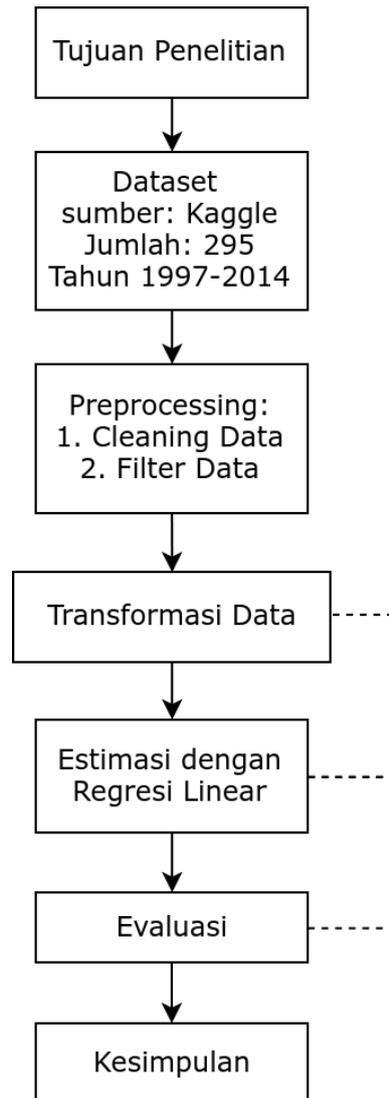
Pada penelitian sebelumnya peneliti menggunakan *software* untuk membantu perhitungan data yaitu *rapidminer* sehingga hasilnya lebih akurat dan dapat digunakan dalam data berskala besar. Namun pada jurnal pembanding lainnya, Peneliti tidak menggunakan *software* tertentu untuk membantu perhitungan data, dan langsung dilakukan secara manual mengingat objek adalah beton yang dibuat secara bersamaan sehingga ketahanan dari setiap beton dapat disamakan jangka waktunya [Kristanto, 2017].

Banyak algoritma yang dapat digunakan selain regresi linear seperti *Trend Moment*, *SVM*, dan *Neural Network* namun pada penelitian ini berfokus dengan penggunaan metode Estimasi menggunakan algoritma linear regresi sederhana dengan dataset dari India yang bertujuan untuk mengetahui seberapa banyak lahan yang diperlukan untuk memproduksi beras, sehingga dapat memprediksi hasil produksi untuk beberapa tahun kedepan. Perhitungan ini akan mengacu pada variabel lahan dan produksi yang dilakukan dari tahun 1997 sampai 2015. Hasil dari estimasi data ini dapat dijadikan bahan pembelajaran untuk dapat memahami metode perhitungan estimasi menggunakan algoritma linear regresi sederhana [Nafi'iyah, 2016].

2. Metode Penelitian

Tahapan penelitian yang dilakukan yaitu dengan dataset tentang lahan pertanian di India sebanyak 2256 *record*, data tersebut sudah melewati proses *cleaning* sehingga total datasets adalah 293. Untuk kerangka penelitian dapat digambarkan dengan struktur pada *flowchart* gambar 1.

Metode regresi Linear ini menggali pengetahuan dari data atau informasi yang dimiliki [Warih, Eggy Inaidi Andana; Rahayu, 2014], dengan data mining ini berisi pencarian pola yang diinginkan pada dataset, sehingga dapat menemukan suatu kesimpulan tentang sebuah data.



Sumber: Hasil Penelitian (2019)

Gambar 1. Kerangka Penelitian

Tahap-tahap data mining yang akan dilakukan adalah pertama dengan mengumpulkan dataset yang didapatkan dari kaggle, dengan judul lahan pertanian dan hasil produksi dari tahun 1997-2014, kedua yaitu tahap *preproccesing* diantaranya adalah *Cleaning Data* dari data yang tidak lengkap dan rancu, karena ada 92 data yang missing. Selanjutnya tahap *filter* data untuk memilih record yang diperlukan untuk penelitian ini. Transformasi Data, mengakumulasikan sejumlah data lahan dan produksi tahunan menjadi total data lahan per-tahun. Sehingga semua lahan dan produksi dari tahun tertentu digabungkan dalam satu tahun. Simple regresi linear, perhitungan estimasi dengan metode regresi linear. selanjutnya adalah tahap evaluasi, dari hasil yang telah dilakukan perhitungan dengan regresi linear dan menghitung estimasi lahan dan hasil produksi. Terakhir merupakan kesimpulan tentang estimasi lahan dan hasil produksi pada tahun yang ditentukan[Popy Purnamasari; Wahid Suyitno H, 2015].

3. Hasil dan Pembahasan

Pada tahap ini data yang dibutuhkan adalah data luas lahan dan banyak produksi dari dataset yang telah di *cleaning* agar dapat di olah dengan analisis data mining menggunakan algoritma linear regresi. Data yang digunakan adalah dataset yang telah di *cleaning* yaitu data lahan dan produksi pertanian di india tahun 1997 sampai dengan 2014 yang dapat dilihat pada tabel 1.

Tabel 1. Dataset Lahan Dan Produksi Pertanian

Tahun	Total Lahan (M2)	Total Produksi
1997	837230	1.448.852
1998	1224571	2.212.772
1999	1077146	1.577.894
2000	3138947	5.239.518
2001	672453	1.426.007
2002	1146475	1.957.602
2003	1139938	2.074.196
2004	1467182	2.351.862
2005	1406229	2.584.169
2006	1333460	2.564.217
2007	1402557	2.638.643
2008	1127142	2.436.884
2009	1131428	2.231.482
2010	1002654	1.988.851
2011	1162427	2.526.731
2012	1130034	2.841.672
2013	1119449	2.563.817
2014	1126026	2.695.480

Sumber: Dataset kaggle 1997-2014 (2019)

Perhitungan

Berdasarkan dari data pada tabel 1 perkembangan lahan pertanian setiap tahun tidak stabil sehingga berpengaruh terhadap hasil produksi yang semakin tahun semakin sedikit. Setelah dihitung rata-rata lahan pertanian beras di India berkurang 14,78% dan hasil produksi berkurang pada angka 8,88% dimana rata-rata perkembangan lahan dan hasil produksi ini dapat diperoleh dengan membandingkan pendapatan setiap tahunnya dengan hasil pada tabel 2.

Tabel 2. Rata-rata perkembangan lahan dan produksi 1997-2014

Tahun	Lahan M ²	Produksi	Rata-Rata hasil Produksi	Rata-Rata Perkembangan Lahan
1997	837.230	1.448.852	0.00%	
1998	1.224.571	2.212.772	34.52%	31,63%
1999	1.077.146	1.577.894	-40.24%	-13,69%
2000	3.138.947	5.239.518	69.88%	65,68%
2001	672.453	1.426.007	-267.43%	-366,79%
2002	1.146.475	1.957.602	27.16%	41,35%
2003	1.139.938	2.074.196	5.62%	-0,57%
2004	1.467.182	2.351.862	11.81%	22,30%
2005	1.406.229	2.584.169	8.99%	-4,33%
2006	1.333.460	2.564.217	-0.78%	-5,46%
2007	1.402.557	2.638.643	2.82%	4,93%
2008	1.127.142	2.436.884	-8.28%	-24,43%
2009	1.131.428	2.231.482	-9.20%	0,38%

Tahun	Lahan M ²	Produksi	Rata-Rata hasil Produksi	Rata-Rata Perkembangan Lahan
2010	1.002.654	1.988.851	-12.20%	-12,84%
2011	1.162.427	2.526.731	21.29%	13,74%
2012	1.130.034	2.841.672	11.08%	-2,87%
2013	1.119.449	2.563.817	-10.84%	-0,95%
2014	1.126.026	2.695.480	4.88%	0,58%
Total	22645348,55	43360648,18	-150,90%	-251,33%
rata-rata	1258074,919	2408924,899	-8,88%	-14,78%

Sumber: Dataset kaggle 1997-2014 (2019)

Dapat diketahui bahwa rata-rata lahan dari tahun ke tahun berkurang sebesar 14,78%. Maka dapat dijadikan acuan untuk prediksi lahan pada tahun 2020 mendatang. Dimana lahan akan berkurang hingga mencapai 367.440 m², hal ini dapat hitung berdasarkan hasil perkembangan lahan pada tabel 2.

Tabel 3. Prediksi Lahan dan Produksi

Tahun	Prediksi	Prediksi
2015	959.551	2.456.209
2016	817.688	2.238.177
2017	696.799	2.039.500
2018	593.782	1.858.458
2019	505.996	1.693.488
2020	431.188	1.543.161

Sumber: Dataset kaggle 1997-2014 (2019)

Lahan akan berkurang sampai 431.188 M², maka dapat dilakukan estimasi perubahan lahan dan hasil produksi sampai tahun 2020 dengan persamaan, hasil prediksi tersebut ditunjukkan pada tabel 3.

Tabel 4. Tahapan Perhitungan

Tahun	X (Lahan)	Y (Produksi)	X ²	Y ²	XY
1997	837.230	1.448.852	700.954.072.900	2.099.172.117.904	1.213.022.359.960
1998	1.224.571	2.212.772	1.499.574.134.041	4.896.359.923.984	2.709.696.420.812
1999	1.077.146	1.577.894	1.160.243.505.316	2.489.749.475.236	1.699.622.210.524
2000	3.138.947	5.239.518	9.852.988.268.809	27.452.548.872.324	16.446.569.307.546
2001	672.453	1.426.007	452.193.037.209	2.033.495.964.049	958.922.685.171
2002	1.146.475	1.957.602	1.314.405.957.453	3.832.205.277.188	2.244.342.542.153
2003	1.139.938	2.074.196	1.299.457.937.083	4.302.287.304.092	2.364.453.718.073
2004	1.467.182	2.351.862	2.152.621.818.035	5.531.254.678.895	3.450.608.569.933
2005	1.406.229	2.584.169	1.977.480.731.680	6.677.928.852.044	3.633.933.905.900
2006	1.333.460	2.564.217	1.778.116.051.646	6.575.207.079.422	3.419.280.809.003
2007	1.402.557	2.638.643	1.967.167.512.755	6.962.438.622.953	3.700.848.965.930
2008	1.127.142	2.436.884	1.270.449.088.164	5.938.403.629.456	2.746.714.305.528

Tahun	X (Lahan)	Y (Produksi)	X ²	Y ²	XY
2009	1.131.428	2.231.482	1.280.129.319.184	4.979.511.916.324	2.524.761.216.296
2010	1.002.654	1.988.851	1.005.314.502.283	3.955.526.629.566	1.994.128.452.450
2011	1.162.427	2.526.731	1.351.236.902.306	6.384.370.860.261	2.937.141.042.645
2012	1.130.034	2.841.672	1.276.976.841.156	8.075.099.755.584	3.211.185.976.848
2013	1.119.449	2.563.817	1.253.166.063.601	6.573.157.609.489	2.870.062.376.833
2014	1.126.026	2.695.480	1.267.934.552.676	7.265.612.430.400	3.035.180.562.480
Total	22645349	43360648	32860410296296	116024330999171	61160475428085

Sumber: Dataset kaggle 1997-2014 (2019)

x = Luas Lahan

y = Hasil Produksi

Model Persamaan Linear Regresi sebagai berikut:

$$Y = a + bX \quad \dots\dots\dots(1)$$

dimana:

Y = Variabel Response

X = Variabel Predictor

a = Konstanta

b = Koefisien Regresi. besaran Response yang ditimbulkan oleh predictor

$$x = 22.645.349$$

$$y = 43.360.648$$

$$x.y = 61.160.475.428.085$$

$$x^2 = 32.860.410.296.296$$

$$a = \frac{(\sum y)(\sum x^2) - (\sum y)(\sum x^2)}{n(\sum x^2) - (\sum x)^2} \quad \dots\dots\dots(2)$$

$$a = \frac{(43.360.648)(32.860.410.296.296) - (43.360.648)(32.860.410.296.296)}{18(32.860.410.296.296) - (22.645.349)^2}$$

$$a = 506.490$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad \dots\dots\dots(3)$$

$$b = \frac{18(61.160.475.428.085) - (22.645.349)(43.360.648)}{18(32.860.410.296.296) - (22.645.349)^2}$$

$$b = 1,5122$$

Untuk menghitung lahan yang dibutuhkan pada tahun 2020 maka dapat diprediksi hasil produksi pada tahun 2020 dengan menggunakan rumus dibawah ini:

$$Y = a + bX$$

$$= 506.490 + 1,5122(1543161)$$

$$= 1.158.524 \text{ Ton}$$

Jumlah produksi di tahun yang diprediksikan ditahun 2020 pada tabel 3 dapat disimpulkan bahwa hasil produksi pada tahun 2020 adalah 1.453.161 Ton. Sehingga untuk mengetahui estmasi lahan pertanian yang diperlukan pada tahun 2020 dapat menggunakan rumus (1) :

$$Y = a + bX$$

$$1.543.160,87 = 506.490 + 1,5122(X)$$

$$1,5122(X) = 1.543.160,87 - 506.490$$

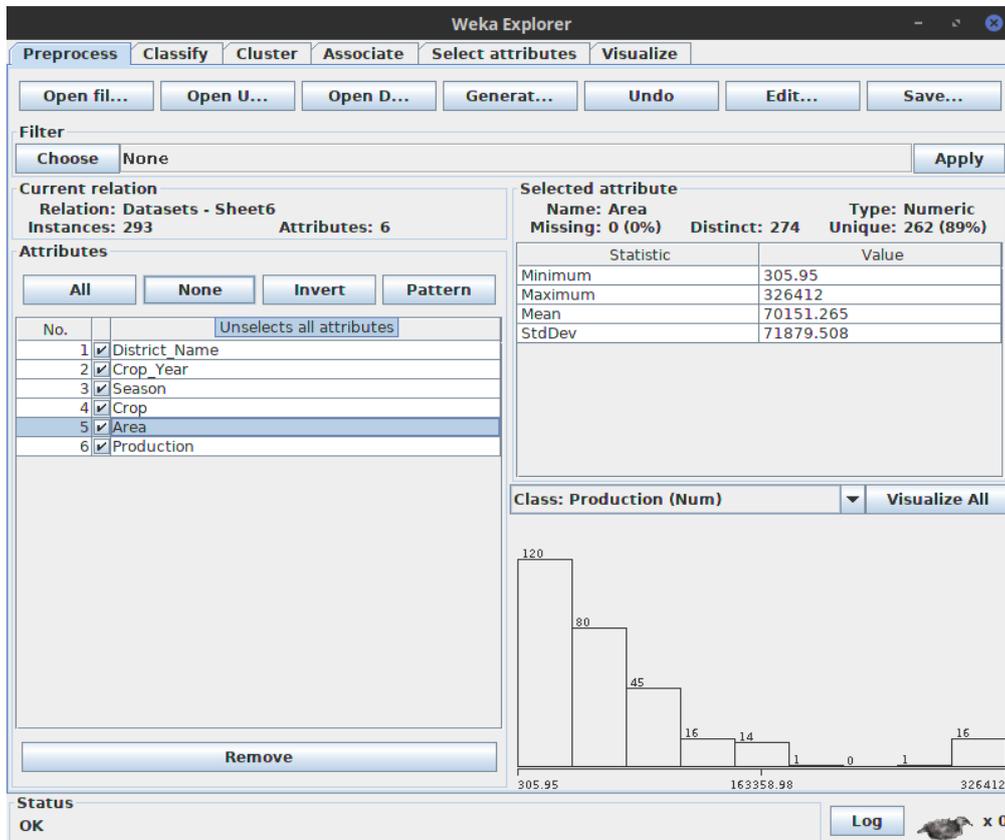
$$X = \frac{1036670,6791}{(1.543.160,87)}$$

$$X = 685.547,51M^2$$

Berdasarkan proses perhitungan regresi linear maka dapat diketahui estimasi ketersediaan lahan pertanian di india pada tahun 2020 adalah 685.547,87 M². Dari hasil itu dapat ditarik kesimpulan bahwa dengan lahan yang berkurang akan membuat hasil produksi juga berkurang sampai pada titik 8,88%.

Pengujian hasil prediksi

Pengujian data dilakukan menggunakan aplikasi pengujian *weka* terhadap hasil perhitungan prediksi hasil produksi pertanian sampai dengan tahun 2020 menggunakan algoritma regresi linier pada gambar 2. Pastikan semua variabel yang ada telah sesuai dan tidak ada *missing value*.



Sumber: Hasil Penelitian (2019)

Gambar 2. Input Dataset Oleh Aplikasi Weka

Pastikan data yang dimasukkan benar dan sesuai dengan apa yang akan kita hitung. Peneliti menggunakan dua variable yaitu x untuk lahan, sedangkan y untuk hasil produksi pertanian, seperti ditunjukkan pada gambar 3.

Relation: Datasets - Sheet6						
No.	District_Name Nominal	Crop_Year Numeric	Season Nominal	Crop Nominal	Area Numeric	Production Numeric
1	BARPETA	1997.0	Winter	Rice	103015.0	76589.0
2	BONGAIGAON	1997.0	Winter	Rice	56841.0	38596.0
3	DHUBRI	1997.0	Winter	Rice	54249.0	34696.0
4	DIBRUGARH	1997.0	Winter	Rice	77456.0	88908.0
5	HAILAKANDI	1997.0	Winter	Rice	35500.0	35504.0
6	BEGUSARAI	1997.0	Winter	Rice	12907.0	12172.0
7	BHAGALPUR	1997.0	Winter	Rice	51405.0	81057.0
8	BHOJPUR	1997.0	Winter	Rice	106982.0	205530.0
9	BIRBHUM	1997.0	Winter	Rice	309991.0	840840.0
10	DARJEELING	1997.0	Winter	Rice	28884.0	34960.0
11	BARPETA	1998.0	Winter	Rice	88835.0	93412.0
12	BONGAIGAON	1998.0	Winter	Rice	56133.0	54267.0
13	DHUBRI	1998.0	Winter	Rice	41378.0	34890.0
14	DIBRUGARH	1998.0	Winter	Rice	75380.0	113898.0
15	HAILAKANDI	1998.0	Winter	Rice	36500.0	54613.0
16	BEGUSARAI	1998.0	Winter	Rice	14762.0	8186.0
17	BHAGALPUR	1998.0	Winter	Rice	49104.0	60757.0
18	BHOJPUR	1998.0	Winter	Rice	102606.0	220745.0
19	ERNAKULAM	1998.0	Winter	Rice	20462.0	36613.0
20	BISHNUPUR	1998.0	Winter	Rice	16140.0	42510.0
21	IMPHAL EAST	1998.0	Winter	Rice	6390.0	13520.0
22	BHADRAK	1998.0	Winter	Rice	169693.0	239827.0
23	BOUDH	1998.0	Winter	Rice	48450.0	50511.0
24	CUTTACK	1998.0	Winter	Rice	127524.0	225009.0
25	JHARSUGUDA	1998.0	Winter	Rice	28454.0	41204.0
26	BIRBHUM	1998.0	Winter	Rice	312904.0	890320.0

Sumber: Hasil Penelitian (2019)

Gambar 3 Pengecekan Data Yang Akan Dihitung.

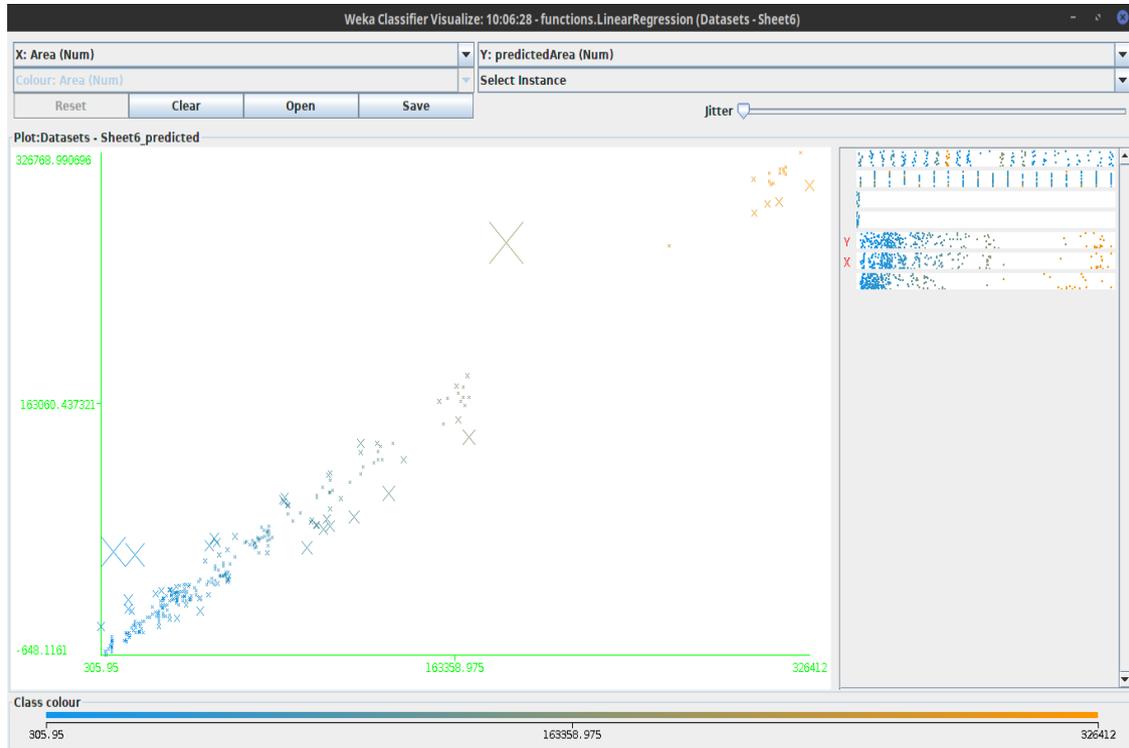
Data ini langsung diproses menggunakan algoritma linear regresi dan menggunakan pilihan test 10 kali *cross validation* dimana program ini akan dihitung 10 kali sehingga akan menghasilkan *error* yang berbeda. Dibawah ini merupakan hasil yang diterima setelah proses perhitungan menggunakan algoritma regresi linier.

Weka Explorer	
Preprocess	Classify
Classifier	LinearRegression -S 0 -R 1.0E-8
Test options	<input type="radio"/> Use training set <input type="radio"/> Supplied test set <input checked="" type="radio"/> Cross-validation Folds 10 <input type="radio"/> Percentage split % 66
(Num) Production	
Result list (right-click for options)	07:16:49 - functions.LinearRegression 07:17:50 - functions.LinearRegression 10:05:33 - functions.LinearRegression 10:06:28 - functions.LinearRegression 10:06:40 - functions.LinearRegression
Classifier output	Linear Regression Model Area = 10796.0889 * District_Name=BEGUSARAI, BOKARO, DARJEELING, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + 7203.5468 * District_Name=BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + 6182.1763 * District_Name=ERNAKULAM, BISHNUPUR, IMPHAL EAST, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + 4520.7511 * District_Name=ERNAKULAM, BISHNUPUR, IMPHAL EAST, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + 7406.9665 * District_Name=ERNAKULAM, BISHNUPUR, IMPHAL EAST, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + 18699.3105 * District_Name=ERNAKULAM, BISHNUPUR, IMPHAL EAST, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + -7285.9732 * District_Name=ERNAKULAM, BISHNUPUR, IMPHAL EAST, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + 7493.1673 * District_Name=ERNAKULAM, BISHNUPUR, IMPHAL EAST, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + 15533.4719 * District_Name=ERNAKULAM, BISHNUPUR, IMPHAL EAST, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + 12136.3771 * District_Name=ERNAKULAM, BISHNUPUR, IMPHAL EAST, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + 27312.8650 * District_Name=ERNAKULAM, BISHNUPUR, IMPHAL EAST, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + 31031.3616 * District_Name=ERNAKULAM, BISHNUPUR, IMPHAL EAST, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + 41600.0822 * District_Name=ERNAKULAM, BISHNUPUR, IMPHAL EAST, JHARSUGUDA, GODDA, HAILAKANDI, BHAGALPUR, DHUBRI, HAZARIBAGH, BONGAIGAON, BOUDH, GUMLA, DIBRUGARH, BARPETA, EAST SINGHBUM, BHOJPUR, CUTTACK, BHADRAK, BIRBHUM + -1024.9458 * Crop_Year + 0.156 * Production + 2063108.1199
	Time taken to build model: 0.03 seconds
	=== Cross-validation ===
	=== Summary ===
	Correlation coefficient 0.99
	Mean absolute error 6158.4048
	Root mean squared error 10130.6575
	Relative absolute error 12.4448 %
	Root relative squared error 14.0582 %
	Total Number of Instances 293

Sumber: Hasil Penelitian (2019)

Gambar 4. Hasil Perhitungan Menggunakan Weka

Gambar 4 adalah hasil perhitungan menggunakan Weka menggunakan dataset belum di transformasi. Hasil pada gambar 3 menggunakan regresi linear dapat dilihat bahwa korelasi koefisiennya adalah 0,9172 ini menunjukkan bahwa hubungan linear kedua variable ini mempunyai hubungan yang kuat dan sesuai. Pada gambar 5 yang menunjukkan persamaan garis dari lahan dan hasil produksi.



Sumber: Hasil Penelitian (2019)

Gambar 5 Grafik Perkembangan Lahan Dan Produksi

4. Kesimpulan

Hasil analisa dan pembahasan dengan teknik data mining menggunakan algoritma regresi linier dan pengujian menggunakan aplikasi weka dapat disimpulkan bahwa: 1) Estimasi dari ketersediaan lahan pada tahun 2020 adalah 685.547,87 M² dengan hasil produksi mencapai 1.543.161 Ton. Sehingga dapat diketahui bahwa semakin tahun lahan pertanian semakin berkurang. 2) Dengan koefisien korelasi yang telah ditemukan adalah 0.99 dapat disimpulkan bahwa hasil perhitungan antara kedua variable “kuat”. Kedua variable mempunyai bobot yang sama kuat dengan hasil yang ditemukan. Hasil ini di lakukan dengan 10 kali test sampling data. 3) Perhitungan dengan metode linear regresi telah mendapat *relative absolute error* sebesar 12.4448%. Dibanding dengan penelitian sebelumnya yang tidak mencantumkan hasil error yang diterima.

Referensi

- Fikri A. 2013. Penerapan Data Mining Untuk Mengetahui Tingkat Kekuatan Beton Yang Dihasilkan Dengan Metode Estimasi Menggunakan Linear Regression. Fak. Ilmu Komput. UDINUS: 1–12.
- Nafi'iyah N. 2016. Perbandingan Regresi Linear , Backpropagation Dan Fuzzy Mamdani Dalam Prediksi Harga Emas. Semin. Nas. Inov. dan Apl. Teknol. di Ind.: 291–296.
- Suyitno PPW, Herlawati. 2015. Member Supermall Terhadap Peningkatan Jumlah Pengunjung Pada Supermall Karawang Metode Regresi Linier Berganda Kualitas Super. J. ICT 2: 101–116.

- Warih, Andana EI; Rahayu Y. 2014. Penerapan Data Mining Untuk Menentukan Estimasi. 1–5.
- Yanto R. 2018. Implementasi Data Mining Estimasi Ketersediaan Lahan Pembuangan Sampah menggunakan Algoritma Simple Linear Regression. J. RESTI (Rekayasa Sist. dan Teknol. Informasi) 2: 361–366.