



Plagiarism Checker X Originality Report

Similarity Found: 13%

Date: Saturday, June 13, 2020

Statistics: 302 words Plagiarized / 2345 Total words

Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

Estimator Nadaraya-Watson dengan pendekatan Cross Validation dan Generalized Cross Validation untuk Mengestimasi Produksi Jagung Febriolah Lamusu¹, Tedy Machmud², Resmawan³ 1,2,3 Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Gorontalo febriolarol@gmail.com Abstract. This paper is to compare the methods of Cross Validation and Generalized Cross Validation using kernel regression with the Nadaraya-Watson estimator on corn production data in Gorontalo Province.

Variables in this study are corn production as the dependent variable and planting area, harvested area, and fertilizer as independent variables. The result of this study indicates that the kernel regression with the Nadaraya-Watson estimator using the Generalized cross validation method is better for corn production estimation with optimal bandwidth value 742392.2 with $h_1 = 400$, $h_2 = 400$ and $h_3 = 400$. Value of $R^2 = 99.99\%$ with MSE 202583.9.

Keywords: Kernel; Estimator Nadaraya-Watson; Cross Validation; Generalized Cross Validation Abstrak. Artikel ini membandingkan metode Cross Validation dan Generalized Cross Validation menggunakan regresi kernel dengan estimator Nadaraya Watson pada data produksi jagung di provinsi Gorontalo. Variabel dalam penelitian ini adalah produksi jagung sebagai variabel dependen dan luas tanam, luas panen, pupuk sebagai variabel independen.

Hasil penelitian ini menunjukkan bahwa regresi kernel dengan estimator Nadaraya-Watson menggunakan metode Generalized cross validation lebih baik untuk mengestimasi produksi jagung dengan nilai bandwidth optimal 742392,2 dengan $h_1 = 400$, $h_2 = 400$ dan $h_3 = 400$. Nilai $R^2 = 99,99\%$ dengan MSE 202583,9. Kata Kunci:

Kernel; Estimator Nadaraya-Watson; Cross Validation; Generalized Cross Validation
Pendahuluan Regresi nonparametrik merupakan metode pendugaan model yang dilakukan berdasarkan pendekatan yang tidak terikat asumsi bentuk kurva regresi tertentu dimana kurva regresi hanya diasumsikan mulus [1].

Beberapa metode pendekatan yang paling populer digunakan dalam regresi nonparametrik adalah spline [2], fourier [3], wafalet [4] dan kernel. Dalam penelitian ini menggunakan kernel Karena memiliki kemampuan yang baik dalam memodelkan data yang tidak mempunyai pola tertentu [5]. Selain itu regresi kernel lebih fleksibel, bentuk matematisnya mudah, dan dapat mencapai tingkat konvergen yang relative cepat [6].

Regresi kernel memiliki beberapa fungsi, diantaranya fungsi Gaussian, fungsi Epanechnikov, fungsi Triangle, fungsi Uniform, fungsi Cosinus, dan fungsi Kuadratik [7]. Fungsi kernel yang biasa digunakan adalah fungsi Gaussian karena fungsi Gaussian lebih umum digunakan dan lebih smooth dibandingkan dengan fungsi kernel yang lain [8]. Dalam regresi kernel salah satu estimator yang sering digunakan yaitu estimator Nadaraya-Watson.

Beberapa penelitian sebelumnya tentang estimator Nadaraya-Watson dapat dilihat pada [9,10,11] Estimator Nadaraya-Watson dengan pendekatan kernel tergantung pada dua parameter yaitu fungsi kernel dan pemilihan bandwidth yang digunakan [7]. Namun diantara keduanya pemilihan bandwidth memiliki pengaruh yang paling kuat terhadap estimasi yang dihasilkan [6].

Metode yang digunakan untuk mendapatkan bandwidth yang optimal adalah dengan menggunakan metode Cross Validation (CV) dan Generalized Cross Validation (GCV) [12]. Provinsi Gorontalo dikenal sebagai salah satu sentra produksi jagung nasional yang mampu memenuhi kebutuhan komoditas di pasar dalam negeri, maupun untuk melakukan ekspor. Menurut dinas pertanian provinsi Gorontalo Produksi jagung di provinsi Gorontalo terus tajam dari 692.000 ton di 2016, menjadi 1,5 juta ton di 2018.

Bahkan dari total ekspor jagung 380.000 ton di 2018, sebanyak 113.000 tonnya adalah hasil produksi petani provinsi Gorontalo. Namun pada tahun 2019 terjadi penurunan produksi jagung di provinsi Gorontalo [13]. Ada sekitar 4.405 hektar lahan pertanian padi dan jagung yang mengalami puso [14]. Untuk itu, perlu dilakukan upaya agar produksi jagung di provinsi Gorontalo tetap terjaga.

Salah satu upaya yang dapat dilakukan adalah dengan melakukan pemodelan untuk memprediksi dan mengetahui produksi jagung di provinsi Gorontalo. Data produksi jagung adalah data yang berfluktuatif dan tidak membentuk suatu pola hubungan

tertentu yang tidak diketahui bentuk fungsinya, sehingga pendekatan nonparametrik adalah pendekatan yang paling tepat untuk digunakan.

Hasil pemodelan ini diharapkan dapat membantu pihak terkait dalam melakukan langkah-langkah strategis agar tidak mengalami kerugian yang signifikan. Metode Penelitian Perbandingan metode yang digunakan untuk penelitian ini yaitu metode Cross Validation (CV) dan Generalized Cross Validation (GCV) menggunakan estimator Nadaraya-Watson. Metode Cross Validation (CV). Salah satu metode untuk menentukan nilai bandwidth adalah Cross Validation (CV).

Metode Cross Validation atau sering disebut CV adalah metode pendugaan data untuk menunjuk kanapa yang harus dilakukan jika pengulangan observasi tersedia. Pada pemilihan bandwidth yang optimum didasarkan nilai Cross Validation (CV) yang minimum [9]. $h = 1 / \sqrt{n} [\sum_{i=1}^n (y_i - \bar{y})^2]^{1/2}$ Dengan h : Cross Validation h : Bandwidth n : Variabel dependent y_i : Nilai estimasi Metode Generalized Cross Validation (GCV).

Metode Generalized Cross Validation (GCV) dalam regresi kernel adalah satu metode untuk memilih bandwidth optimal dengan meminimalkan fungsi GCV. Optimasi GCV adalah memilih h optimal yang meminimalkan nilai GCV [7]. $h = 1 / \sqrt{n} [\sum_{i=1}^n (y_i - \bar{y})^2]^{1/2}$ Dengan h : nilai h pada bandwidth h : banya ksubjek n : data actual subjek ke i : hasil estimasi subjek ke i : jumlah dari elemen diagonal utama matriks penghalus $n \times n$ Estimator Nadaraya-Watson.

Estimator Nadaraya-Watson diperkenalkan Pada tahun 1964 oleh Nadaraya dan Watson. Dimana estimator ini untuk memperkirakan m sebagai rata-rata tertimbang secara lokal dengan menggunakan kernel sebagai fungsi pembobotan [15]. $1/n \sum_{j=1}^n K((x - x_j)/h) y_j$ h : nilai Bandwidth n : nilai amatan variabel prediktor ke- i : nilai amatan variabel respon ke- i : nilai random variabel X atau dapat dengan nilai tertentu dari variabel X Hasil dan Pembahasan Scatter Plot Dalam menentukan model regresi nonparametrik kernel langkah awal yang dilakukan adalah membuat scatter plot antara variabel dependen dan varibel independen.

/// Gambar 1. Plot produksijagungdenganluastanam, luaspanen dan pupuk. berdasarkan Gambar 1, diketahui hubungan X_1 dan Y membentuk garis lurus. Akan tetapi data banyak terkumpul di titik awal, sedangkan data lainnya nilainya cukup jauh berbeda

sehingga ada beberapa data pencilan yang terlihat pada Gambar. Begitupun dengan hubungan X2 dan Y membentuk garis lurus namun data banya kterkumpul dititik awal dan data lainnya berada cukup jauh sehinga ada beberapa data pencilan.

Pada gambar hubungan X3 dan Y diketahui tidak membentuk pola apapun. Data menyebar secara acak sehingga tidak jelas hubungan X3 dan Y. Ada juga beberapa data pencilan yang terlihat pada Gambar. Sehingga pola data seperti ini sangat cocok untuk didekati dengan pendekatan regresi nonparametrik. Analisis Korelasi Uji analisis korelasi dalam regresi nonparametrik kernel dilakukan untuk mengetahui variabel-variabel independent yang berpengaruh terhadap variabel dependent. Dibawah ini merupakan hasil uji korelasi yang disajikan dalam tabel 1. Tabel 1.

Analisis Korelasi Variabel _Nilai Sig _Nilai Korelasi _Luas Tanam _0,00 _0,907 _Luas Panen _0,00 _0,947 _Pupuk _0,00 _0,734 _ Berdasarkan tabel 1 dapat dilihat nilai signifikan dari ketiga variabel $(0,00) > ?? (0,05)$ artinya variabel luas tanam, luas panen dan pupuk berpengaruh signifikan terhadap variabel produksi jagung. Pemilihan Bandwidth optimum dengan Cross Validation (CV).

Pemilihan bandwidth optimum sangat penting dalam regresi kernel, karena nilai bandwith memiliki pengaruh yang paling kuat terhadap hasil estimasi. Berikut adalah nilai bandwidth optimum menggunakan metode Cross Validation (CV) dengan software R diperoleh. Dapat dilihat pada tabel 2. Tabel 2. Nilai Bandwidth dengan CV h1_h2_h3 _CV _MSE _1500 _1500 _1500 _65176849 _8931740 _ Berdasarkan Tabel 2 diketahui bandwidth optimum berdasarkan kriteria Cross Validation (CV) adalah bandwidth yang memiliki nilai Cross Validation (CV) terkecil yaitu 65176849 dengan nilai MSE sebesar 8931740 dan bandwidth optimum adalah $h1 = 1500, h2 = 1500, h3 = 1500$. Berikut adalah plot nilai bandwidth h1, h2 dan h3 dengan CV. Dapat dilihat pada gambar 2. /// Gambar 2.

Nilai Bandwidth dengan CV Berdasarkan Gambar 2 dapat dilihat nilai bandwidth optimum CV h1 berada pada titik 1500, bandwidth optimum CV h2 berada pada titik 1500 dan bandwidth optimum CV h3 berada pada titik 1500. Pemilihan Bandwidth optimum dengan Generalized Cross Validation (GCV). Berikut adalah nilai bandwidth optimum menggunakan metode Generalized Cross Validation (GCV) dengan software R diperoleh. Dapat dilihat pada Tabel 3. Tabel 3.

Nilai Bandwidth dengan GCV h1_h2_h3 _GCV _MSE _400 _400 _400 _742392,2 _202583,9 _ Berdasarkan Tabel 3 diketahui bandwidth optimum berdasarkan kriteria Generalized Cross Validation (GCV) adalah yang memiliki nilai Generalized Cross Validation (GCV) terkecil yaitu 742392.2 dengan MSE sebesar 202583.9 dan bandwidth

optimum adalah $h_1 = 400$, $h_2 = 400$, $h_3 = 400$. Berikut adalah plot nilai bandwidth h_1 , h_2 dan h_3 dengan GCV. Dapat dilihat pada gambar 3.

/// Gambar 3. Plot Bandwidth GCV Berdasarkan Gambar 3 dapat dilihat nilai bandwidth optimum GCV h_1 berada pada titik 400, bandwidth optimum GCV h_2 berada pada titik 400 dan bandwidth optimum GCV h_3 berada pada titik 400. Estimasi Nadaraya-Watson dengan metode Cross Validation (CV).

Berdasarkan perhitungan bandwidth optimum Cross Validation model regresi kernel Gaussian dengan estimator Nadaraya-Watson menghasilkan persamaan untuk memprediksi produksi jagung sebagai berikut. $Y = 1500$, $X_1 = 1$, $X_2 = 2$, $X_3 = 3$

$$\hat{Y} = \frac{1}{n} \sum_{j=1}^n Y_j \left(\frac{1}{h_1} - \frac{1}{h_1^2} X_{1j} \right) \left(\frac{1}{h_2} - \frac{1}{h_2^2} X_{2j} \right) \left(\frac{1}{h_3} - \frac{1}{h_3^2} X_{3j} \right)$$
Dengan $\frac{1}{n} = \frac{1}{44}$, $\frac{1}{h_1} = 1$, $\frac{1}{h_2} = 2$, $\frac{1}{h_3} = 3$

$$\hat{Y} = \frac{1}{44} \sum_{j=1}^{44} Y_j (1 - X_{1j}) (2 - X_{2j}) (3 - X_{3j})$$
Estimasi Nadaraya-Watson dengan metode Generalized Cross Validation (GCV).

Berdasarkan perhitungan bandwidth optimum Generalized Cross Validation model regresi kernel Gaussian dengan estimator Nadaraya-Watson menghasilkan persamaan untuk memprediksi produksi jagung sebagai berikut. $Y = 400$, $X_1 = 1$, $X_2 = 2$, $X_3 = 3$

$$\hat{Y} = \frac{1}{n} \sum_{j=1}^n Y_j \left(\frac{1}{h_1} - \frac{1}{h_1^2} X_{1j} \right) \left(\frac{1}{h_2} - \frac{1}{h_2^2} X_{2j} \right) \left(\frac{1}{h_3} - \frac{1}{h_3^2} X_{3j} \right)$$
Dengan $\frac{1}{n} = \frac{1}{44}$, $\frac{1}{h_1} = 1$, $\frac{1}{h_2} = 2$, $\frac{1}{h_3} = 3$

$$\hat{Y} = \frac{1}{44} \sum_{j=1}^{44} Y_j (1 - X_{1j}) (2 - X_{2j}) (3 - X_{3j})$$
Berikut adalah kurva data actual produksi jagung dan data estimasi menggunakan metode pendekatan Cross Validation (CV) dan Generalized Cross Validation (GCV) di provinsi Gorontalo.

/ Gambar 4. Plot Estimasi data actual dengan CV dan GCV Berdasarkan gambar diatas terlihat jelas bahwa kurva data aktual dan prediksi data menggunakan estimator Nadaraya-Watson dengan metode Cross Validation dan Generalized Cross Validation memiliki pola yang sama dan saling berhimpit.

Namun diantara kedua metode, Generalized Cross Validation lebih mendekati data aktual. Perbandingan Model. Untuk menentukan metode terbaik dalam melakukan estimasi produksi jagung di provinsi Gorontalo dapat membandingkan nilai koefisien determinasi dan MSE yang dihasilkan. Berikut adalah tabel hasil perhitungan nilai koefisien dan MSE dari metode Cross Validation (CV) dan Generalized Cross Validation (GCV). Tabel 4.

Nilai R-Square dan MSE Metode
 _Nilai R-Square _MSE _Cross Validation _0,995349
 _8931740 _Generalized Cross Validation _0,9998944 _202583,9 _
 Berdasarkan tabel 4 didapatkan nilai R-Square Cross Validation (CV) sebesar 0.995349 dan nilai R-Square

dari Generalized Cross Validation (GCV) sebesar 0.9998944. Dari perhitungan nilai koefisien kedua metode menunjukkan bahwa metode Generalized Cross Validation (GCV) menghasilkan nilai koefisien determinasi yang lebih baik dibandingkan dengan metode Cross Validation (CV).

Sedangkan nilai MSE didapatkan nilai MSE Cross Validation (CV) = 8931740 sedangkan Generalized Cross Validation (GCV) = 202583,9. Dari perhitungannya nilai MSE kedua metode menghasilkan nilai MSE Generalized Cross Validation (GCV) lebih kecil dibandingkan dengan nilai MSE Cross Validation (CV). Sehingga metode terbaik untuk mengestimasi produksi jagung di provinsi Gorontalo adalah metode Generalized Cross Validation (GCV).

Kesimpulan Metode Generalized Cross Validation (GCV) lebih baik untuk mengestimasi data produksi jagung di provinsi Gorontalo dengan nilai bandwidth optimum = 742392,2 nilai koefisien determinasi = 0,9998944 dan MSE = 202583,9 dibandingkan dengan metode Cross Validation (CV). Daftar Pustaka Eubank, R. L. Nonparametric Regression and Spline Smoothing. Second Edition. Marcel Dekker Inc. New York. 1999. Budi, L., Nur C., dan Toha S.

Estimasi fungsi regresi dalam model regresi nonparametric biresponse spline menggunakan estimator smoothing spline dan estimator kernel, Vol. 15 No. 2. 20-24.2019. Laila, R. K., Alan, P., dan Rukun S. Pemodelan regresi semiparametrik dengan pendekatan deret fourier. Jurnal Gaussian, Vol. 9 No 1. 50-63. 2020. Suparti., Rezzy, E. C., Budi W., dan Hasbi Y. The shift invariant discrete wavelet transform (SIDWT) with inflation time series application. Jurnal of mathematics, 8 (4).

2016. W. Hardle. Applied nonparametric regression, Cambridge University Press, New York. 1994. W. Hardle. Applied nonparametric regression, Cambridge University Press, New York. 1990. Suparti., R. Santoso., A. Prahutama., dan A. R. Devi. Regresi Nonparametrik. Wade Group. Ponorogo. 2018. Komang, G. dan Gusti, A. estimator kernel dalam model regresi nonparametrik. Jurnal matematika, Vol. 2 No. 1. 19-30. 2012. Nurul, A., Noami, N.D.,

dan Shantika, M. Estimasi model regresi nonparametrik kernel menggunakan estimator Nadaraya-Watson. Jurnal matematika, statistika dan terapannya, Vol. 8 No. 4, 633-638. 2019. Ida, Purwanti. Regresi nonparametrik kernel menggunakan estimator Nadaraya-Watson dalam data time series. Jurnal penelitian, Vol. 1 No. 1. 49-56. 2019. Tiani, W. U., Martyana, P., dan Vega, Z. V. Kernel nonparametric regression for the modeling of the productivity wetland paddy.

jurnal international seminar education and development of Asia. 2018. R.T. Ogden.
Essential wavelets for statistical applications and data analysis. Boston. Birkhauser. 1997.
Badan Pusat Statistik (BPS) Provinsi Gorontalo. Provinsi Gorontalo Dalam Angka 2019.
Gorontalo: BPS Provinsi Gorontalo. 2019. Badan Pusat Statistik (BPS) Provinsi Gorontalo.
Statistik Hortikultura Provinsi Gorontalo 2018. Gorontalo: BPS Provinsi Gorontalo. 2018
Saputra, J. A.,

dan Listyani, E. Pemilihan Bandwidth pada Estimator Nadaraya-Watson dengan Tipe
Kernel Gaussian pada Data Time Series. Jurnal Pendidikan Matematika Dan Sains, Vol. 1
No. 1. 1–7. 2016. Bontemps, C., Robin, J., dan Van den Berg, G. Equilibrium search with
continuous productivity dispersion: Theory and non parametric estimation. International
Economic Review, 41, 305-358. 2000.

INTERNET SOURCES:

<1% - <https://arxiv.org/pdf/1906.10221>
<1% -
https://www.researchgate.net/publication/267140899_Introduction_to_Nonparametric_Regression
<1% - <https://www.scribd.com/document/396696071/2>
<1% -
<http://digilib.unila.ac.id/24975/2/SKRIPSI%20TANPA%20BAB%20PEMBAHASAN.pdf>
1% - http://eprints.undip.ac.id/42514/1/Sulton_Syafii.pdf
<1% -
<https://id.scribd.com/doc/167852556/Jurnal-Matematika-Vol-1-No-1-Januari-2013>
1% - <http://repository.its.ac.id/3070/7/1315201011-Master-Theses.pdf>
<1% - <http://digilib.unila.ac.id/11208/15/BAB%20II.pdf>
<1% - http://ejurnal.its.ac.id/index.php/sains_seni/article/download/22477/3635
2% -
<http://www.neraca.co.id/article/113802/sektor-primer-pemerintah-sebut-perlu-adanya-ekspor-untuk-jaga-harga-jagung>
<1% -
<https://id.scribd.com/doc/272997946/RPJM-Bidang-Pangan-Dan-Pertanian-2015-2019>
<1% - <https://laporanpraktikumbersama.blogspot.com/feeds/posts/default>
1% -
<https://dataquest.co.id/2015/07/13/cara-memilih-strategi-manajemen-pengetahuan-yang-tepat/>
<1% - <https://www.scribd.com/document/401655910/Thesis-ARIP-RAMADAN-pdf>

<1% -

<http://sigitnugroho.id/e-Skripsi/2015/08/Pemodelan%20Jumlah%20Kasus%20Angka%20Kematian%20Bayi%20dengan%20GWPR.pdf>

<1% -

https://www.researchgate.net/publication/285592575_PERBANDINGAN_ESTIMATOR_KERNEL_DAN_ESTIMATOR_SPLINE_DALAM_MODEL_REGRESI_NONPARAMETRIK_I_KOMANG_GDE_SUKARSA_1_I_GUSTI_AYU_MADE_SRINADI_2_NI_LUH_AYU_PUSPA_LESTARI

1% - <https://ejournal3.undip.ac.id/index.php/gaussian/article/download/14693/14217>

<1% - <http://digilib.unila.ac.id/7457/124/BAB%20II.pdf>

1% -

<https://id.123dok.com/document/6qmj535q-fungsi-kernel-pada-metode-regresi-non-parametrik.html>

<1% - http://eprints.undip.ac.id/48059/4/BAB_IV.pdf

<1% - <https://jimfeb.ub.ac.id/index.php/jimfeb/article/view/2430/2212>

<1% - <http://repository.ub.ac.id/2510/>

<1% -

https://caridokumen.com/download/model-konsumsi-rumah-tangga-di-indonesia-_5a44d921b7d7bc7b7a8cb5da_pdf

<1% - http://eprints.undip.ac.id/78469/1/C33-2018_Prosiding_VARIANSI_UNM.pdf

<1% - <https://idoc.pub/documents/termodinamika-kimia-pon22o591yn0>

<1% - <https://vaskoedo.wordpress.com/2008/page/16/>

1% -

<https://tatangkostaman.blogspot.com/2010/08/pasca-panen-jagung-pendahuluan-pasca.html>

<1% - <https://id.scribd.com/doc/57102219/STIS-SE-064962-2010>

<1% - http://ejurnal.its.ac.id/index.php/sains_seni/article/download/4512/1358

<1% -

https://www.researchgate.net/publication/323409962_Optimasi_Pemodelan_Regresi_Linier_Berganda_Pada_Prediksi_Jumlah_Kecelakaan_Sepeda_Motor_Dengan_Algoritme_Genetika

<1% -

https://www.researchgate.net/publication/322765511_Spline_Truncated_Multivariabel_pada_Permodelan_Nilai_Ujian_Nasional_di_Kabupaten_Lombok_Barat

1% - https://ejournal.undip.ac.id/index.php/media_statistika/article/view/25892

<1% -

<https://www.scribd.com/document/322617070/The-Shift-Invariant-Discrete-Wavelet-Transform-SIDWT-with-Inflation-Time-Series-Application>

<1% - <http://digilib.unila.ac.id/7457/20/DAFTAR%20%20PUSTAKA.pdf>

1% - <https://gorontalo.bps.go.id/publication.html>

<1% -

<https://www.bps.go.id/statictable/2012/04/19/1606/persentase-rumah-tangga-menurut-provinsi-jenis-kelamin-krt-yang-bekerja-dan-daerah-tempat-tinggal-2009-2018.html>

1% - <https://ideas.repec.org/p/ctl/louvir/2010026.html>

<1% -

<https://nyuad.nyu.edu/en/academics/divisions/social-science/faculty/christian-bontemps/publications.html>