

PENGUNAAN ALGORITMA *DETECT DEVIATING CELLS* (DDC) UNTUK MENGATASI PENCILAN SEL PADA REGRESI *PARTIAL LEAST SQUARES*

TANIA CHANDRA PRASTIWI



**DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

@Hak Cipta milik IPB University

Dengan ini saya menyatakan bahwa skripsi dengan judul “Penggunaan Algoritma *Detect Deviating Cells* (DDC) untuk Mengatasi Pencilan Sel pada Regresi *Partial Least Squares*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2024

Tania Chandra Prastiwi
G5401201003

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

ABSTRAK

TANIA CHANDRA PRASTIWI. Penggunaan Algoritma *Detect Deviating Cells* (DDC) untuk Mengatasi Pencilan Sel pada Regresi *Partial Least Squares*. Dibimbing oleh NGAKAN KOMANG KUTHA ARDANA dan HADI SUMARNO.

Data multivariat seringkali mengandung pencilan yang dapat mengganggu hasil analisis. Algoritma *Detect Deviating Cells* (DDC) digunakan untuk mengidentifikasi dan menangani pencilan pada tingkatan sel pada gugus data, menggantinya dengan nilai imputasi (X_{imp}) sehingga menciptakan gugus data yang disempurnakan untuk regresi *Partial Least Squares* (PLS) berikutnya. Akurasi model dievaluasi menggunakan *Root Mean Square Error* (RMSE) dan metrik *R-squared adjusted* (R^2_{adj}). Hasil penelitian menunjukkan bahwa algoritma DDC pada regresi PLS menghasilkan model yang lebih akurat dalam menduga nilai sebenarnya, terutama pada data yang mengandung pencilan dan multikolinearitas. Model ini juga mampu menjelaskan variabilitas data yang lebih baik dibandingkan dengan metode regresi *Ordinary Least Squares* (OLS).

Kata kunci: *detect deviating cells*, multikolinearitas, pencilan sel, *partial least squares*

ABSTRACT

TANIA CHANDRA PRASTIWI. Application of Detect Deviating Cells (DDC) Algorithm to Handle Cellwise Outliers in Partial Least Squares Regression. Supervised by NGAKAN KOMANG KUTHA ARDANA and HADI SUMARNO.

Multivariate data often contains outliers that can interfere the analysis results. The Detect Deviating Cells (DDC) algorithm was used to identify and handle cell-level outliers in the data clusters, replacing them with imputed values (X_{imp}) thus creating enhanced data clusters for subsequent Partial Least Squares (PLS) regression. Model accuracy was evaluated using Root Mean Square Error (RMSE) and adjusted R-squared metric (R^2_{adj}). The results show that the DDC algorithm in PLS regression produces a more accurate model in predicting the true value, especially in data containing outliers and multicollinearity. This model is also able to explain data variability better than the Ordinary Least Squares (OLS) regression method.

Kata kunci: cellwise outliers, detect deviating cells, multicollinearity, partial least squares



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2024
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PENGGUNAAN ALGORITMA *DETECT DEVIATING CELLS* (DDC) UNTUK MENGATASI PENCILAN SEL PADA REGRESI *PARTIAL LEAST SQUARES*

TANIA CHANDRA PRASTIWI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana Matematika pada
Program Studi Matematika

**DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



Penguji pada Ujian Skripsi:
Prof. Dr. Ir. I Wayan Mangku, M.Sc.



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Skripsi : Penggunaan Algoritma *Detect Deviating Cells* (DDC) untuk Mengatasi Pencilan Sel pada Regresi *Partial Least Squares*

Nama : Tania Chandra Prastiwi
NIM : G5401201003

Disetujui oleh

Pembimbing 1:
Ir. Ngakan Komang Kutha Ardana, M.Sc.



Pembimbing 2:
Dr. Ir. Hadi Sumarno, M.S.



Diketahui oleh

Ketua Departemen Matematika:
Dr. Ir. Endar Hasafah Nugrahani, M.S.
NIP 196312281989032001



Tanggal Ujian: 02 Agustus 2024

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Oktober 2023 sampai bulan Juni 2024 ini ialah pencilan, dengan judul “Penggunaan Algoritma *Detect Deviating Cells* (DDC) untuk Mengatasi Pencilan Sel pada Regresi *Partial Least Squares*”. Penyusunan karya ilmiah ini tidak lepas dari bantuan banyak pihak, untuk itu penulis mengucapkan terima kasih kepada:

1. kedua orang tua penulis Bapak Sigit dan Ibu Kirana serta kakak penulis Prasna Fikri Oktavian yang selalu memberikan doa dan dukungan kepada penulis,
2. Bapak Ir. Ngakan Komang Kutha Ardana, M.Sc. dan Dr. Ir. Hadi Sumarno, M.S. selaku dosen pembimbing tugas akhir yang telah meluangkan waktu dalam membimbing dan memberi saran,
3. Bapak Prof. Dr. Ir. I Wayan Mangku, M.Sc. selaku dosen penguji tugas akhir yang telah meluangkan waktu dalam memberi saran,
4. dosen pembimbing akademik, serta seluruh dosen dan staf Departemen Matematika IPB yang telah memberikan ilmu dan bantuan selama masa studi di Departemen Matematika IPB,
5. keluarga besar Matematika 57 yang sudah menemani perjalanan empat tahun di IPB,
6. Dias Riyan Utomo yang telah menemani, memberikan dukungan, tekanan dan semangat kepada penulis,
7. pihak-pihak lainnya yang telah banyak membantu penulis yang tidak dapat disebutkan satu per satu.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2024

Tania Chandra Prastiwi



DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	x
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Regresi Linear Berganda	3
2.2 Pencilan Sel	3
2.3 <i>Ordinary Least Squares</i> (OLS)	4
2.4 <i>Partial Least Squares</i> (PLS)	5
2.5 <i>Detect Deviating Cells</i> (DDC)	7
2.6 <i>Root Mean Square Error</i> dan <i>R-Squared</i>	9
III METODE	11
3.1 Data	11
3.2 Prosedur Analisis Data	11
IV HASIL DAN PEMBAHASAN	12
4.1 Eksplorasi Data	12
4.2 Regresi <i>Ordinary Least Squares</i>	14
4.3 Regresi <i>Partial Least Squares</i>	16
4.4 Regresi OLS dengan DDC	18
4.5 Regresi PLS dengan DDC	22
4.6 Evaluasi Model	24
V SIMPULAN DAN SARAN	27
5.1 Simpulan	27
5.2 Saran	27
DAFTAR PUSTAKA	28
LAMPIRAN	30
RIWAYAT HIDUP	36

Hak Cipta Dilindungi Undang-undang
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



DAFTAR TABEL

1	Dugaan parameter model regresi dengan metode OLS data tanpa pencilan	15
2	Dugaan parameter model regresi dengan metode OLS data proporsi pencilan 5%	16
3	Dugaan parameter model regresi PLS data tanpa pencilan	17
4	Dugaan parameter model regresi PLS data proporsi pencilan 5%	18
5	Dugaan parameter model regresi OLS dengan DDC pada data tanpa pencilan	20
6	Dugaan parameter model regresi OLS dengan DDC pada data proporsi pencilan 5%	22
7	Dugaan parameter model regresi PLS dengan DDC data tanpa pencilan	23
8	Dugaan parameter model regresi PLS dengan DDC data proporsi pencilan 5%	24
9	Nilai akurasi model regresi linear berganda	25

DAFTAR GAMBAR

1	Visualisasi paradigma pencilan baris (kiri) dan pencilan sel (kanan)	4
2	Algoritma DDC	9
3	Matriks korelasi antar variabel data tanpa pencilan	12
4	Diagram kotak data tanpa pencilan	13
5	Matriks korelasi antar variabel data proporsi pencilan 5%	13
6	Diagram kotak data proporsi pencilan 5%	14
7	Plot DDC data tanpa pencilan	19
8	Plot DDC data proporsi pencilan 5%	21
9	Grafik perbandingan nilai RMSE	25
10	Grafik perbandingan nilai R^2_{adj}	26

DAFTAR LAMPIRAN

1	Data tanpa pencilan	31
2	Data proporsi pencilan 5%	33
3	Nilai sel pencilan dan X imputasi data proporsi pencilan 5% dari metode DDC	35