

## Outlier Detection Using Minimum Vector Variance Algorithm with Depth Function and Mahalanobis Distance

### Pendeteksian Pencilan Menggunakan Algoritma *Minimum Vector Variance* Dengan Fungsi *Depth* dan Jarak Mahalanobis

Puji Puspa Sari<sup>1\*</sup>, Erna Tri Herdiani<sup>2\*</sup>, Nurtiti Sunusi<sup>3\*</sup>

#### Abstract

Outliers are observations where the point of observation deviates from the data pattern. The existence of outliers in the data can cause irregularities in the results of data analysis. One of the solution to solve this problem is to use statistical approach. The aim of this research is to detect outlier using the Minimum Vector Variance (MVV) algorithm. In this algorithm, the data sorting criteria is changed using the Robust Depth Mahalanobis. The results obtained in this study indicate that the outliers distance value and the location of the outliers in the plot are different for each ordering modification using the three methods namely RDMMVV (Robust Depth Mahalanobis MVV), DMMVV (Depth Mahalanobis MVV), and MMVV (Mahalanobis MVV) methods.

**Keywords:** Depth, Mahalanobis, MVV, outliers

#### Abstrak

Pencilan merupakan suatu pengamatan yang titik pengamatannya menyimpang dari pola data. Keberadaan pencilan pada data dapat mengakibatkan penyimpangan terhadap hasil analisis data. Salah satu solusi untuk permasalahan tersebut adalah dengan pendeteksian pencilan menggunakan pendekatan statistik. Tujuan dalam penelitian ini adalah mendeteksi pencilan menggunakan algoritma *Minimum Vector Variance* (MVV). Dalam algoritma ini, dilakukan perubahan kriteria pengurutan data menggunakan *Robust Depth* Mahalanobis. Hasil yang diperoleh dalam penelitian ini menunjukkan bahwa nilai jarak pencilan serta letak pencilan dalam plot berbeda untuk setiap modifikasi pengurutan menggunakan ketiga metode yaitu metode RDMMVV (*Robust Depth* Mahalanobis MVV), DMMVV (*Depth* Mahalanobis MVV), dan MMVV (Mahalanobis MVV).

**Kata kunci:** *Depth*, Mahalanobis, MVV, Pencilan.

\* *Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin*

*Email:* [pujipuspasari04@gmail.com](mailto:pujipuspasari04@gmail.com)<sup>1</sup>; [herdiani.erna@gmail.com](mailto:herdiani.erna@gmail.com)<sup>2</sup>; [ntitisanusi@gmail.com](mailto:ntitisanusi@gmail.com)<sup>3</sup>



## 1. PENDAHULUAN

Pencilan merupakan suatu pengamatan yang jarak titik pengamatannya relatif jauh dari pusat data dan titik pengamatannya menyimpang dari pola data [2]. Menurut [15] mengidentifikasi pencilan pada data multivariat lebih kompleks dibanding pada data univariat, karena *outlier* pada multivariat berada dalam ruang  $n$ -dimensi. Pada umumnya, *outlier* dapat terjadi karena kesalahan manusia, kesalahan instrumen, perilaku curang, perubahan perilaku sistem atau kesalahan sistem, dan penyimpangan alami di dalam populasi. Keberadaan pencilan pada data dapat mengakibatkan penyimpangan terhadap hasil analisis data dan menyebabkan ketidakakuratan hasil analisis dalam penelitian, oleh karena itu salah satu solusinya adalah dengan mendeteksi adanya pencilan pada data [7]. Konsep dasar dalam pendeteksian *outlier* pada data multivariat adalah dengan mengukur jarak setiap titik ke pusat datanya [8]. Titik sampel dikatakan pencilan jika memiliki nilai jarak yang relatif lebih besar dibanding mayoritas sampel yang lain.

Beberapa metode yang digunakan untuk mendeteksi pencilan diantaranya adalah menggunakan Jarak Mahalanobis dan metode Minimum Covariance Determinan (MCD). Metode Jarak Mahalanobis adalah suatu metode yang menggunakan vektor rata-rata dan matriks varians kovarians [17]. Metode *Minimum Covariance Determinan* (MCD) digunakan untuk mendeteksi pencilan berdasarkan nilai determinan matriks varians-covarians. Prinsip metode MCD adalah mendapatkan subhimpunan dari keseluruhan pengamatan yang matriks varians-kovariansnya memiliki determinan terkecil diantara semua kombinasi kemungkinan data. Pendeteksian pencilan dengan metode MCD dilakukan berdasarkan jarak robust dan nilai *cut-offnya* [4], algoritma *Fast Minimum Covariance Determinant* (FMCD) merupakan pengembangan dari metode MCD, ide dasarnya menggunakan teknik MCD dengan menggunakan iterasi yang selektif (*selective iteration*). Untuk data yang kecil MCD dapat ditentukan dengan eksak oleh FMCD. Sedangkan, untuk data yang besar FMCD akan memberikan kesimpulan yang sangat akurat pada data yang memuat  $h$  atau lebih observasi [16], dan metode *Minimum Vector Variance* (MVV) merupakan modifikasi algoritma FMCD dengan menggunakan ukuran *Vector Variance* (VV) yang minimum [10].

MVV merupakan metode *robust* terhadap pencilan dibandingkan dengan metode lainnya yang telah disebutkan sebelumnya, karena memiliki beberapa keunggulan yaitu nilai *breakdown point* yang tinggi, bersifat *affine equivariance*, dan efisiensi komputasi dengan tingkat kompleksitas yang rendah. Kriteria pengurutan metode MVV menggunakan jarak Mahalanobis menggunakan vektor mean sampel dan matriks variansi kovariansi sampel [10]. Salah satu keunggulan MVV dibanding FMCD adalah metode MVV dapat digunakan pada data dengan determinan sama dengan nol karena tidak terpengaruh oleh nilai varians/kovarians yang sama dengan nol [11]. Beberapa peneliti yang menggunakan MVV diantaranya adalah [1] menggunakan MVV dengan pembobot *reweighted*, menghasilkan skema *reweighted* yang mampu mempertahankan *breakdown point* 0,5 dan mencapai efisiensi yang lebih tinggi pada distribusi normal, [12] menerapkan *Robust Principal Component Analysis* untuk mereduksi data yang telah dianalisis menggunakan metode MVV, [20] menggunakan  $T^2$  Hotelling pada MVV dan menghasilkan penelitian yang menunjukkan bahwa peta kendali MVV memiliki kinerja kompetitif relatif terhadap diagram kontrol, [7] menggunakan metode MVV dengan mengkontaminasi data dengan pencilan untuk melihat keakuratan metode MVV.

Kriteria jarak Mahalanobis merupakan salah satu metode untuk mengukur atau menghitung jarak tiap observasi terhadap pusat datanya [14]. Sedangkan, kriteria *depth* adalah gagasan tentang kedalaman data yang dikemukakan oleh Tukey pada tahun 1975 sebagai alat grafis untuk memvisualisasikan kumpulan data bivariat dan sejak itu diperluas ke kasus multivariat [18]. Kedalaman titik relatif terhadap kumpulan data mengukur seberapa dalam titik itu berada di dalam sebaran data. Konsep *Depth* dalam data merupakan kriteria baru dalam mengurutkan data multivariat. Peneliti [5] dalam tulisannya telah menyatakan bahwa kuadrat jarak mahalanobis dapat dinyatakan bukan dalam bentuk invers matriks varians kovarians, tetapi dalam bentuk kriteria baru menggunakan *Depth*. Menurut [21] perhitungan komputasi menggunakan jarak Mahalanobis akan mengalami kendala jika jumlah sampel mencapai ratusan atau ribuan. Penelitian lainnya mengatakan bahwa jarak Mahalanobis dapat dimodifikasi menggunakan *statistical depth function* menjadi *robust depth* Mahalanobis yang dapat memberikan hasil *affine invariant* lebih baik daripada jarak Mahalanobis [6]. Selain itu, penelitian terdahulu yang menerapkan fungsi *depth* pada jarak Mahalanobis diantaranya adalah [22] menghasilkan sifat-sifat yang lebih baik dibanding Mahalanobis, sedangkan [19] menerapkan fungsi *depth* pada bagan kendali dengan menggunakan kriteria *Vector Variansi*.

Berdasarkan beberapa penelitian tersebut, peneliti tertarik untuk mengubah kriteria pengurutan dalam algoritma *Minimum Vector Variance* dengan menggunakan fungsi *Depth* dan Jarak Mahalanobis untuk pelabelan pencilan pada data multivariat.

## 2. METODE PENELITIAN

### 2.1 Jarak Mahalanobis

Jarak Mahalanobis merupakan generalisasi dari Jarak Euclid yang distandarisasi, selain itu jarak ini dapat mengatasi masalah perbedaan skala dalam data dan mempertimbangkan korelasi antar peubah. Pada metode ini juga didefinisikan sebagai jarak dua titik yang melibatkan kovarians atau korelasi antar peubah [3].

Jarak mahalanobis merupakan metode untuk mendeteksi pencilan pada data multivariat. Jarak mahalanobis diperoleh dengan menghitung jarak tiap observasi terhadap pusat datanya. Kuadrat jarak mahalanobis dihitung dengan rumusan [14] yang didefinisikan dengan persamaan berikut:

$$d_{MD}^2 = (x_i - \bar{x})^t S^{-1} (x_i - \bar{x}) > \chi_{p,(1-\alpha)}^2 \quad (1)$$

Dimana

$d_{MD}^2$  = kuadrat jarak pengamatan ke-  $i$

$x_i$  = nilai pengamatan ke-  $i$

$\bar{x}$  = vektor rata-rata dari pengamatan

$S^{-1}$  = invers matriks variansi kovariansi sampel

### 2.2 Fungsi *Depth*

Fungsi *Depth* merupakan gagasan baru yang dikembangkan secara intensif dalam dekade terakhir di bidang statistik non parametrik, geometri komputasi, aljabar, dan ilmu komputer. Fungsi *Depth* terkait erat dengan pengurutan data multivariat, estimasi *robust*, dan deteksi outlier. Salah satu yang paling banyak digunakan dalam statistik dan bidang terkait adalah Mahalanobis

*Depth*. Secara teoritis, fungsi *depth* tersebut dibangun agar memiliki fungsi yang lebih baik dan efisien secara komputasi. Praktis, peran fungsi kedalaman dalam aplikasi semakin luas dan luas. Dalam penelitian ini, versi sampel kedalaman mahalnobis didefinisikan sebagai berikut [22]: Misal  $X_1, X_2, \dots, X_n$  adalah sampel acak dari distribusi normal multivariat. Persamaan untuk vektor rata-rata sampel ( $\bar{X}$ ) dan matriks variansi kovarians sampel ( $S$ ) adalah sebagai berikut:

$$\bar{X} = \frac{1}{h} \sum_{i \in H} X_i \quad (2)$$

$$S = \frac{1}{h-1} \sum_{i \in H} (X_i - \bar{X})(X_i - \bar{X})^t \quad (3)$$

Versi sampel untuk Mahalanobis *Depth* dari  $X_i$  dapat ditulis sebagai berikut [6]:

$$MD_i = \frac{1}{1 + (X_i - \bar{X})^t S^{-1} (X_i - \bar{X})} \quad (4)$$

Persamaan tersebut digunakan untuk mengukur seberapa dalam  $X$  terhadap sampel acak  $X_1, X_2, \dots, X_n$ . Semakin besar nilai  $MD_i$  semakin dekat titik  $X$  ke pusat  $\bar{X}$ .

### 2.3 Robust Depth Mahalanobis

*Robust Depth* Mahalanobis merupakan pengembangan dari *Depth* Mahalanobis. Dalam literatur, misalnya [9], [13], [16], dan [10] persamaan jarak dihitung berdasarkan definisinya. Jadi, untuk jarak mahalnobis dan *depth* mahalnobis dibutuhkan inversi matriks kovarian sampel  $S$  untuk menghitungnya. Ini adalah perhitungan yang memakan waktu terutama untuk kumpulan data berdimensi tinggi. Kompleksitas komputasinya dalam hal jumlah operasi dalam algoritmanya termasuk tinggi. [6] mendefinisikan ulang *depth* Mahalanobis dengan memperkenalkan fungsi *depth* yang baru sebagai berikut:

1. Persamaan ini setara dengan *depth* Mahalanobis dalam artian bahwa dapat memberikan urutan multivariat yang sama, yaitu urutan titik pusat-*output* yang sama yang dijelaskan oleh poin kedua dan ketiga dalam Definisi. 1 [13];
2. Perhitungannya tidak terlalu rumit dibandingkan dengan *depth* Mahalanobis.

Definisi baru *depth* Mahalanobis atau *robust depth* Mahalanobis akan dirumuskan menggunakan determinan dari matriks yang dipartisi, sebagai berikut:

Misal  $X_1, X_2, \dots, X_n$  adalah sampel acak dari distribusi  $p$ -variate, dengan kriteria pengurutan data sebagai berikut:

$$M_i = \begin{bmatrix} 1 & (X_i - \bar{X})^t \\ (X_i - \bar{X}) & S \end{bmatrix} \quad (5)$$

$M_i$  merupakan matriks dengan ukuran  $(p + 1) \times (p + 1)$  dengan  $X_i; i = 1, 2, \dots, n$  dan  $S$  merupakan matriks variansi kovariansi sampel pengamatan.

### 2.4 Robust Depth Mahalanobis dengan Minimum Vector Variance

Pada penelitian ini, pelabelan pencilan pada data multivariat menggunakan algoritma *Minimum Vector Variance* (MVV) dengan mengubah kriteria pengurutan jarak Mahalanobis

menjadi *Robust Depth* Mahalanobis. Adapun langkah-langkah yang dilakukan adalah sebagai berikut:

- 1) Mengasumsikan himpunan data pengamatan berdistribusi normal multivariate  $\{X_1, X_2, \dots, X_n\}$
- 2) Mengambil himpunan data yang terdiri dari  $h = \frac{(n+p+1)}{2}$  data, sebutlah himpunan data ini dengan  $H_{old}$ .
- 3) Menghitung vektor rata-rata  $\bar{X}_{H_{old}}$  dan matriks variansi kovariansi  $S_{H_{old}}$  dari data  $H_{old}$
- 4) Menghitung  $M_i$  dengan  $i = 1, 2, \dots, n$ , berikut matriks  $M_i$

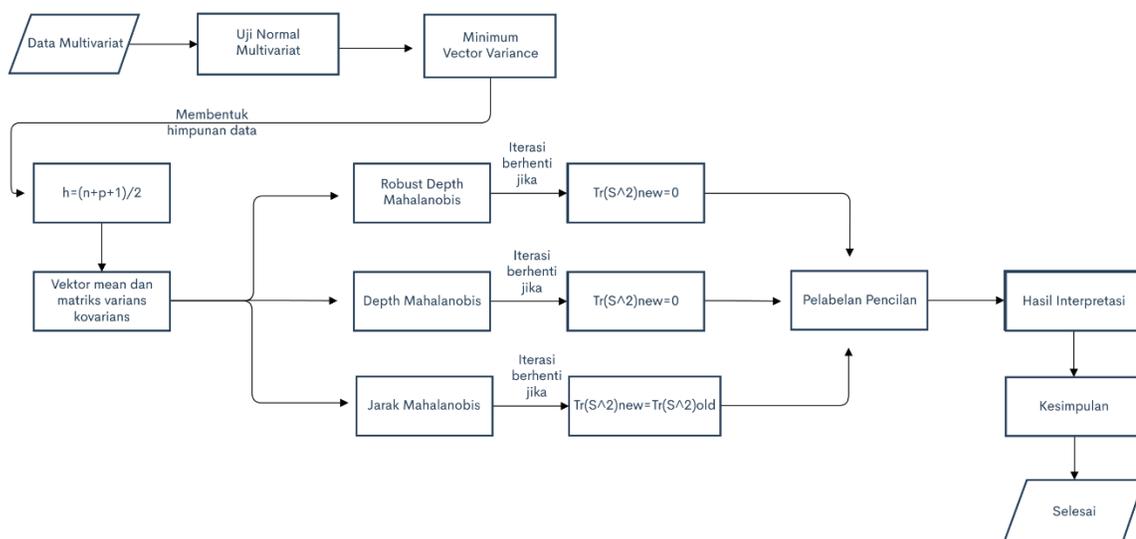
$$M_i = \begin{bmatrix} 1 & (X_i - \bar{X})^t \\ (X_i - \bar{X}) & S \end{bmatrix}$$

- 5) Mengurutkan hasil perhitungan dari yang terbesar ke yang terkecil. Urutan ini akan memberikan permutasi indeks observasi  $\pi$ . Misalnya hasil pengurutan data tersebut adalah

$$M_{\pi(1)} \geq M_{\pi(2)} \geq \dots \geq M_{\pi(n)}$$

- 6) Membentuk suatu himpunan baru yang terdiri dari  $h$  observasi dengan indeks  $\pi(1), \pi(2), \dots, \pi(h)$  dan berilah nama  $H_{new}$  dengan himpunan data  $\{X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(h)}\}$
- 7) Menghitung vektor mean  $\bar{X}_{H_{new}}$  dan matriks kovariansi  $S_{H_{new}}$  dari data  $H_{new}$
- 8) Jika  $Tr(S_{H_{new}}^2) = 0$  maka proses berhenti. Jika  $Tr(S_{H_{new}}^2) \neq Tr(S_{H_{old}}^2)$  maka ulangi langkah ke 2-8. Proses dilanjutkan sampai iterasi ke- $k$  atau jika  $Tr(S_{H_k}^2) - Tr(S_{H_{k+1}}^2) \leq \epsilon$ .  $\epsilon$  merupakan bilangan konstan terkecil.
- 9) Misal  $T_{M_{VV}}$  dan  $S_{M_{VV}}$  adalah vektor mean dan matriks kovariansi dari proses yang telah dijalankan. Maka subsampel dari *Robust Depth* Mahalanobis dalam metode *Minimum Vektor Variance* didefinisikan sebagai subset  $H_{new} = \{X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(h)}\}$  dari  $T_{M_{VV}}$  dan  $S_{M_{VV}}$ .

## 2.5 Diagram Alir Penelitian



**Gambar 1.** Diagram Alir Penelitian

**3. HASIL**

Data yang digunakan dalam penelitian ini merupakan data simulasi yang terdiri dari 500 variabel, 1000 observasi, dan 50 data kontaminasi pencilan. Data ini diaplikasikan pada metode MVV dengan menggunakan 3 estimator pengurutan data multivariat yakni Jarak Mahalanobis, *Depth* Mahalanobis, dan *Robust Depth* Mahalanobis. Asumsi yang harus terpenuhi jika menggunakan data riil adalah data harus berdistribusi normal multivariat dan setiap variabel data saling berkorelasi.

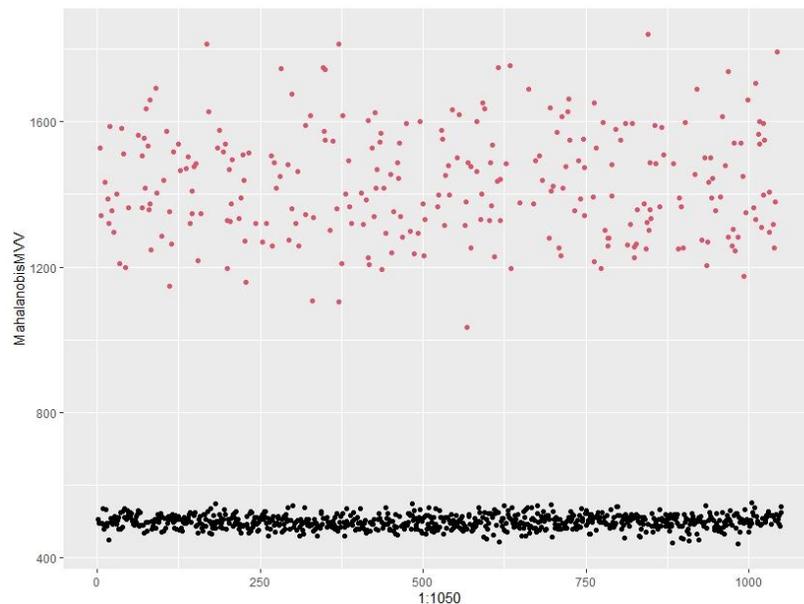
Berdasarkan prosedur analisis MVV pada data multivariat jika diperoleh hasil  $Tr(S_{H_{new}}^2) = Tr(S_{H_{old}}^2)$  maka proses dihentikan. Kemudian, data dengan  $Tr(S^2)$  yang paling minimum yang akan digunakan untuk menghitung Jarak *Robust* MVV. Tetapi, karena dalam penelitian ini menggunakan fungsi *Depth* maka syarat nilai minimum trace akan ditentukan jika  $Tr(S_{H_{new}}^2) = 0$ . Data dengan nilai trace yang memenuhi syarat dari ketiga metode akan digunakan dalam pelabelan pencilan. Metode tersebut terdiri dari Mahalanobis *Minimum Vector Variance* (MMVV), *Depth* Mahalanobis *Minimum Vector Variance* (DMMVV), dan *Robust Depth* Mahalanobis *Minimum Vector Variance* (RDMMVV). Berikut adalah tabel yang menunjukkan kemampuan dari ketiga metode:

**Tabel 1.** Kemampuan Metode berdasarkan Nilai dan Letak Pencilan

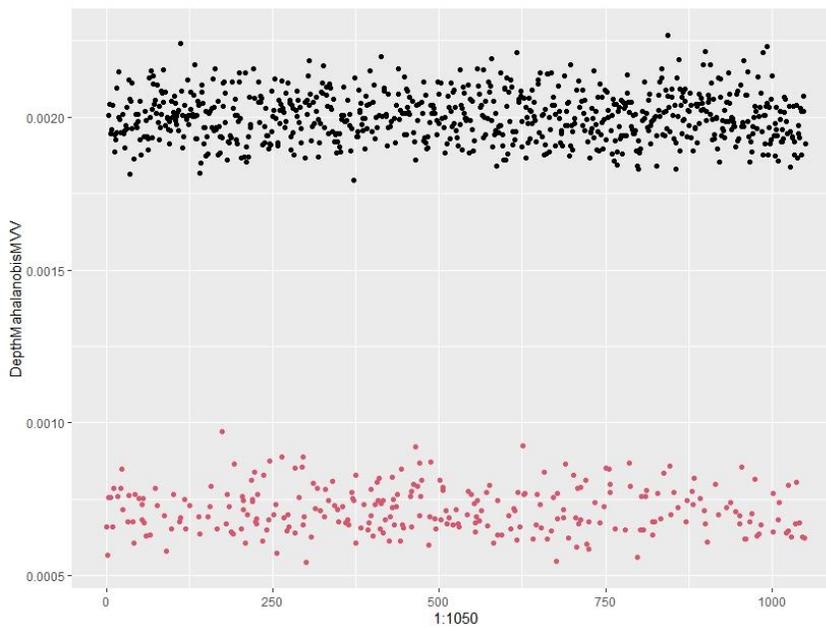
No.	Metode	Interval Nilai Jarak Pencilan	Letak Pencilan pada Plot
1.	MMVV	(1300) - (1750)	Di atas
2.	DMMVV	(0.0005) - (0.001)	Di bawah
3.	RDMVV	(-3.008e-88) - (-1.811e-88)	Di bawah

Tabel 1 memberikan informasi terkait metode yang digunakan dalam penelitian yakni MMVV, DMMVV, dan RDMMVV. Pada interval pencilan diperoleh perbedaan yang signifikan berdasarkan interval nilai jarak pencilan diantara ketiga metode. Perbedaan interval nilai pencilan pada setiap metode disebabkan oleh adanya modifikasi yang dilakukan pada pengurutan data yang berdampak pada jarak setiap pengamatan.

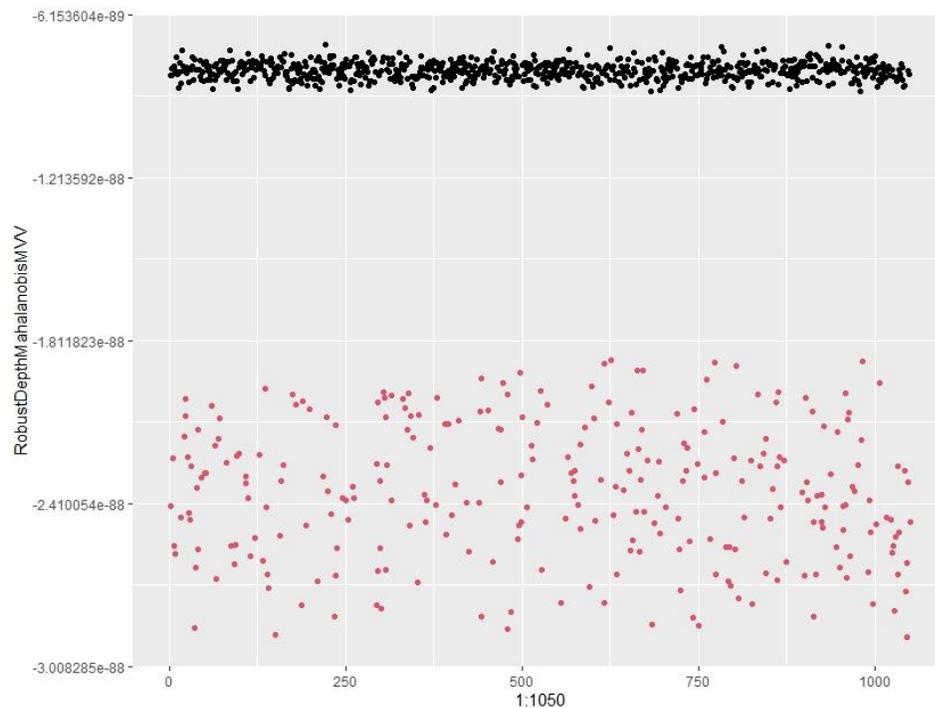
Selain itu, perbedaan yang disebabkan oleh modifikasi dari pengurutan data pada algoritma MVV adalah letak pencilan pada *scatterplot*. Pada metode MMVV letak pencilan berada pada bagian atas data yang bukan pencilan karena pengurutan data yang digunakan hanya jarak Mahalanobis. Berbeda halnya dengan metode DMMVV dan RDMMVV, pada *scatterplot* dapat dilihat bahwa letak pencilan berada pada bagian bawah dari data yang bukan pencilan. Hal ini disebabkan karena kedua metode ini mengandung kriteria *Depth* yang menggunakan kedalaman untuk mengukur jarak pengamatan. Berikut adalah hasil pendeteksian pencilan dalam penelitian ini dapat dilihat pada gambar berikut:



**Gambar 2.** Pendeteksian Pencilan menggunakan Mahalanobis *Minimum Vector Variance*



**Gambar 3.** Pendeteksian Pencilan menggunakan *Depth Mahalanobis Minimum Vector Variance*



**Gambar 4.** Pendeteksian Pencilan menggunakan *Robust Depth Mahalanobis Minimum Vector Variance*

Berdasarkan gambar 2,3, dan 4 dapat dilihat bahwa secara visual dari plot yang dihasilkan terdapat perbedaan sebaran titik-titik observasi antara pencilan dan bukan pencilan. Untuk gambar 2 sebaran data pencilan terletak pada bagian atas dengan titik-titik pengamatan berwarna merah. Sedangkan, untuk data yang bukan pencilan terletak pada bagian bawah diberi warna hitam untuk menandai titik-titik pengamatan. Letak pencilan pada MMVV terletak pada bagian atas karena kriteria pengurutan data multivariat hanya menggunakan jarak Mahalanobis. Untuk gambar 3 dan 4 sebaran data pencilan terletak pada bagian bawah data yang bukan pencilan dengan titik-titik pengamatan berwarna merah. Sedangkan, untuk data yang bukan pencilan terletak pada bagian atas data pencilan dengan diberi warna hitam untuk menandai titik-titik pengamatan. Letak pencilan berada di bagian bawah karena menggunakan fungsi depth dan jarak Mahalanobis. Fungsi depth merupakan salah satu metode dalam mengukur jarak berdasarkan kedalaman data sehingga pada gambar 3 dan 4 sebaran pencilan berada dibawah. Hal ini menunjukkan bahwa observasi dinyatakan sebagai pencilan jika semakin dalam titik suatu observasi dari titik pusat datanya.

Selain itu, perbedaan diantara ketiga metode pengurutan data dalam MVV dapat dilihat pada hasil pengurutan jarak tiap gambar. Gambar 4 atau pendeteksian pencilan yang menggunakan *Robust Depth Mahalanobis MVV* memiliki nilai jarak yang cenderung lebih rendah dibanding gambar 2 dan 3 yang menggunakan MMVV dan DMMVV, sehingga metode ini dapat dengan jelas memisahkan antara pencilan dan bukan dengan melihat nilai observasi yang semakin jauh dari pusat datanya. Pencilan pada data multivariat pada algoritma MVV ditentukan apabila jarak yang diperoleh dari proses algoritma lebih besar atau sama dengan  $\chi^2_{(p;1-\alpha)}$ .

#### 4. KESIMPULAN

Metode *Robust Depth* Mahalanobis MVV atau RDMMVV dalam penelitian ini menunjukkan nilai jarak yang lebih rendah dibanding DMMVV dan MMVV. Artinya pencilan dalam data dengan sangat mudah diidentifikasi berdasarkan nilai jarak yang sangat jauh. Selain itu, Fungsi *Depth* dalam RDMMVV berfungsi sangat baik dengan menunjukkan kedalaman data dalam plot data yang digunakan.

#### ACKNOWLEDGEMENT

Penulis mengucapkan terima kasih kepada pembimbing penelitian yang senantiasa memberikan dukungan dan arahan untuk menyelesaikan penelitian ini.

#### DAFTAR PUSTAKA

- [1] Ali, Hazlina, 2013. On Robust Mahalanobis Distance issued from minimum vector variance. Universiti Utara Malaysia. *Far East Journal of Mathematical Sciences*
- [2] Barnett, V. and Lewis, T., 1984. Outliers in Statistical Data. 2nd Edition, *John Wiley & Sons*, Chichester.
- [3] Boni, Melda Putri., 2018. Mengelompokkan Subjek Menggunakan Mahalanobis Distance dan PCA. Sumatra Utara. Tesis Magister Jurusan Matematika Universitas Sumatra Utara
- [4] Butler RW, Davies PL, Jhun M., 1993. Asymptotics for the Minimum Covariance Determinant estimator. *Ann Stat* 1993, 21:1385–1400
- [5] Djauhari, M.A., 2005. Improved Monitoring of Multivariate Process Variability, *Journal of Quality Technology*, 37, 32-39.
- [6] Djauhari, M.A., Umbara, R. F., 2006. On Mahalanobis Depth Function, paper ini telah dipresentasikan di *International Conference on Mathematics and Natural Sciences (ICMNS)*, Bandung, Indonesia
- [7] E T Herdiani, 2017. Modifikasi Penaksir Robust dalam Pelabelan Outlier Multivariat. *Jurnal Matematika, Statistika, dan Komputasi* Vol. 14, No.1, 46-53
- [8] E T Herdiani, P P Sari, and N Sunusi, 2019. Detection of Outliers in Multivariate Data using Minimum Vector Variance Method. *Journal of Physics: Conference Series*, Volume 1341 Issue 9
- [9] Hadi AS., 1992. Identifying Multivariate Outlier in Multivariate Data. *Journal of Royal Statistical Society*. 3(2):761-771.
- [10] Herwindiati, D.E., Djauhari, M.A., and Mashuri, M., 2006. Robust multivariate outlier labeling. *Communication in Statistics*.
- [11] Herwindiati, D. E., 2006. A New Criterion in Robust Estimation for Location and Covariance Matrix and Its Application for Outlier Labeling, Disertasi, Institut Teknologi Bandung.
- [12] Herwindiati D.E., and Sani M., 2009. The Robust Principal Component Using Minimum Vector Variance, *Proceedings of the World Congress on Engineering 2009*. Vol I, 1 – 3 July 2009. London, U.K.

- [13] Liu, R. Y., 1990. On a Notion of Data Depth Based on Random Simplices. *The Annals of Statistics*, 18(1), 405–414. doi:10.1214/aos/1176347507
- [14] Mahalanobis P.C., 1936. On the generalised distance in statistics. *Proceedings of the National Institute of Science India*, 2: 4955.
- [15] Manoj, K., Senthamarai Kannan, K., 2013. Comparison of Methods for detecting Outliers, *International Journal of Scientific and Engineering Research*, 4(9), 709-714
- [16] Rousseeuw, P. J., & Driessen, K. V., 1999. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, 41(3), 212–223. doi:10.1080/00401706.1999.10485670
- [17] Rousseeuw, P.J. & Van Zomeren, B.C., 1990. Unmasking Multivariate Outlier and Leverage Points. *Journal of the American Statistical Association*. 85: 633-639.
- [18] Singh, K., Parelius, J. M., & Liu, R. Y., 1999. Multivariate analysis by data depth: descriptive statistics, graphics and inference *The Annals of Statistics*, 27(3), 783–858. doi:10.1214/aos/1018031260
- [19] Suwanda Idris, Lisnur Wachidah, Teti Sofiyayanti, Erwin Harahap, 2019. The Control Chart of Data Depth Based on Influence Function of Variance Vector. *J. Phys.: Conf. Ser.* 1366 012125
- [20] Syed Yahaya, Sharipah Soaad and Ali, Hazlina and Omar, Zurni, 2011. *An alternative hotelling  $T^2$  control chart based on Minimum Vector Variance (MVV)*. *Modern Applied Science*, 5 (4). pp. 132-151. ISSN 1913-1844
- [21] Ye, N., Chen, Q., 2001. An Anomaly Detection Technique Based on A Chi Square Statistic for Detecting Intrusion into Information Systems, *Quality and Reliability Engineering International*, *Qual. Realb. Engng. Int.*, 17, 105-112.
- [22] Zuo, Y., & Serfling, R., 2000. General notions of statistical depth function. *The Annals of Statistics*, 28(2), 461–482. doi:10.1214/aos/1016218226