

REDUKSI DIMENSI SET DATA DENGAN DRC PADA METODE KLASIFIKASI SVM DENGAN UPAYA PENAMBAHAN KOMPONEN KETIGA

Eko Prasetyo

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Bhayangkara Surabaya
Jl. A. Yani 114 Surabaya 60231, Indonesia
Email: eko1979@yahoo.com

Abstrak

Set data yang diolah dalam sistem seperti data mining, information retrieval, computer vision, atau sistem-sistem lain yang menggunakan set data sebagai basis data utama dalam menyelesaikan kasus yang ditangani, bisa memiliki ukuran yang sangat besar dalam hal jumlah fitur yang digunakan. Banyak keuntungan yang didapat jika dilakukan reduksi dimensi. Kunci keuntungannya adalah banyak algoritma data mining yang bekerja dengan baik jika dimensi lebih rendah. Penelitian ini mengembangkan metode Dimension Reduction Technique for K-Means Clustering Algorithm (DRC) dengan menambahkan komponen ketiga yaitu z. Hasilnya, kinerja akurasi metode yang diusulkan (DRC 3 DIM) dalam mereduksi dimensi pada metode klasifikasi SVM mampu memberikan akurasi yang tetap relatif baik ketika jumlah dimensi awal masih tidak banyak. Sedangkan waktu komputasi yang dibutuhkan, baik untuk training maupun prediksi masih dapat ditoleransi untuk dapat digunakan, setelah mempertimbangkan bahwa waktu training dan prediksi berada pada level pertengahan ketika dibandingkan dengan metode pembanding.

Kata kunci: reduksi, dimensi, drc, klasifikasi.

1. PENDAHULUAN

Set data yang diolah dalam sistem seperti *data mining, information retrieval, computer vision*, atau sistem-sistem lain yang menggunakan set data sebagai basis data utama dalam menyelesaikan kasus yang ditangani, bisa memiliki ukuran yang sangat besar dalam hal jumlah fitur yang digunakan. Contoh yang paling tampak adalah sistem *information retrieval* yang melakukan pemrosesan dokumen dalam melakukan sistem temu kembali informasi, sistem ini menggunakan frekuensi kemunculan kata dalam dokumen sebagai basis fitur untuk setiap dokumen. Dalam kasus seperti ini (Tan et al, 2006), biasanya jumlah fitur bisa ribuan atau puluhan ribu kata (komponen) yang digunakan sebagai fitur, satu fitur untuk setiap kata dalam kamus. Contoh kasus yang lain adalah set data runut waktu (*time series*) yang berisi data transaksi harian setiap stok barang selama periode 30 tahun. Dalam kasus ini, atribut adalah harga yang ditetapkan pada hari-hari tersebut, dan jumlahnya tentu saja ribuan.

Jumlah fitur yang besar memang diharapkan bahwa akan semakin banyak karakter data yang dapat di-*capture* menjadi model sistem. Tetapi jangan lupa bahwa jumlah fitur yang semakin besar juga berpengaruh pada waktu komputasi yang juga lama. Jumlah fitur yang besar bisa mengakibatkan terjadinya *curse of dimensionality* (Tan et al, 2006), dimana dalam masalah ini mengakibatkan banyak jenis analisis data menjadi secara signifikan lebih berat/susah dengan semakin meningkatnya jumlah dimensi. Secara khusus, dengan semakin meningkatnya jumlah dimensi, maka data menjadi semakin meningkat sebaran tempat yang dibutuhkan. Untuk klasifikasi, hal ini berarti bahwa tidak ada obyek data yang cukup untuk dapat membuat model yang handal dalam memberikan label kelas pada semua obyek yang ada. Untuk masalah pengelompokan (*clustering*), definisi kepadatan dan jarak antar data, yang merupakan hal yang kritis dalam clustering, menjadi kurang berarti. Hasilnya, banyak algoritma klasifikasi dan clustering (dan algoritma analisis data yang lain) menjadi bermasalah pada data dengan dimensi tinggi – berupa menurunkan akurasi klasifikasi dan kualitas cluster yang jelek (Tan et al, 2006).

Banyak keuntungan yang didapat pada pekerjaan reduksi dimensi. Kunci keuntungannya adalah banyak algoritma data mining yang bekerja dengan baik jika dimensi (jumlah atribut/fitur pada data) lebih rendah. Alasannya adalah reduksi dimensi dapat menghilangkan fitur yang tidak relevan, dan mengurangi *noise*, selain itu juga mengurangi *curse of dimensionality*. Alasan lain adalah reduksi dimensi dapat menjadikan model lebih dapat dipahami karena dibangun

menggunakan fitur yang lebih sedikit. Selain itu, reduksi dimensi dapat mempermudah dalam visualisasi data. Bahkan jika reduksi dimensi tidak mengurangi data menjadi dua atau tiga dimensi, data sering divisualisasikan dengan mengamati pasangan dua atau tiga atribut tetapi jumlah atributnya telah banyak berkurang. Keuntungan terakhir adalah jumlah waktu dan memori yang dibutuhkan oleh algoritma data mining berkurang dengan adanya reduksi dimensi (Tan et al, 2006).

Sejumlah teknik reduksi dimensi telah banyak dikembangkan oleh para peneliti. Dalam Tan et al (2006) dan Prasetyo (2012), reduksi dimensi secara linear seperti Principal Component Analysis (PCA) dan Singular Value Decomposition (SVD) menjadi dua metode reduksi dimensi yang sangat luas penggunaannya. Metode yang lain seperti Factor Analysis, Locally Linear Embedding (LLE) (Roweis dan Saul, 2000), Multidimensional Scaling (MDS) (Cox dan Cox, 1994), FastMap (Faloutsos dan Lin, 1995), ISOMAP (Tenenbaum, 1998), dan sejumlah metode lain juga menjadi metode alternatif dalam reduksi dimensi untuk kasus-kasus tertentu.

Penelitian yang dilakukan oleh Bishnu dan Bhattacharjee (2012) melakukan reduksi dimensi dalam kerangka kerja Dimension Reduction Technique for K-Means Clustering Algorithm (DRC). Metode ini bekerja dengan cara mempartisi setiap fitur menjadi dua komponen yaitu komponen x (cosinus) dan y (sinus), kemudian dihitung rata-rata semua fitur pada masing-masing komponen x dan y . Hasilnya berupa data dengan dua fitur (dua dimensi). Set data hasilnya menjadi mudah divisualisasikan. Tingkat error yang didapatkan relatif lebih rendah daripada ketika fitur tidak direduksi atau dikurangi dengan metode decision tree.

Makalah ini memaparkan hasil penelitian ketika DRC diterapkan sebagai reduksi dimensi pada metode klasifikasi. Data yang akan dilakukan proses klasifikasi dilewatkan proses reduksi dimensi dengan DRC, kemudian diproses menggunakan algoritma klasifikasi. Pemaparan makalah ini dibagi menjadi 5 bagian. Bagian 1 menyajikan pendahuluan yang melatarbelakangi penulis melakukan penelitian. Bagian 2 menyajikan penelitian-penelitian terkait yang menjadi dasar bagi penulis untuk melakukan penelitian. Bagian 3 menyajikan kerangka kerja penelitian, dan pengembangan metode DRC. Bagian 4 menyajikan pengujian dan analisis yang dilakukan untuk mengukur kinerja DRC. Dan bagian 5 menyajikan simpulan dari hasil penelitian dan saran untuk penelitian berikutnya.

2. PENELITIAN TERKAIT

Metode reduksi dimensi Dimension Reduction Technique for K-Means Clustering Algorithm (DRC) yang diusulkan oleh Bishnu dan Bhattacharjee (2012) melakukan pemartisian setiap fitur menjadi komponen x (cosinus) dan y (sinus). Kemudian dihitung rata-rata semua fitur pada komponen x dan y .

Komponen x_{ij} adalah komponen x dari data ke- i fitur ke- j , dirumuskan oleh persamaan (1).

$$x_{ij} = r_{ij} \cos \theta_j \quad (1)$$

Untuk komponen y_{ij} adalah komponen y dari data ke- i fitur ke- j , dirumuskan oleh persamaan (2).

$$y_{ij} = r_{ij} \sin \theta_j \quad (2)$$

Untuk r_{ij} adalah nilai data ke- i fitur ke- j . Untuk $i = 1, \dots, N$, dan $j = 1, \dots, M$. Untuk N adalah jumlah data, sedangkan M adalah jumlah fitur (dimensi). Sedangkan θ untuk masing fitur ditentukan oleh persamaan (3).

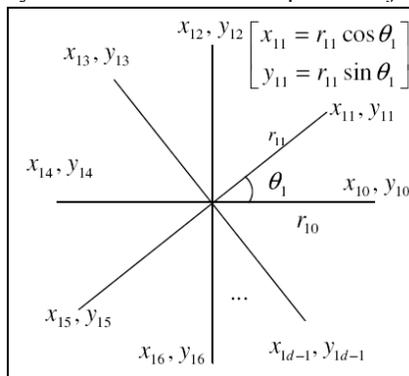
$$\theta_j = \theta_{j-1} + \frac{360^\circ}{M}, \theta_1 = 0^\circ \quad (3)$$

Selanjutnya, dua fitur baru untuk data ke- i (X_i dan Y_i) didapatkan dengan menghitung rata-rata masing-masing komponen x dan y , seperti pada persamaan (4) dan (5).

$$X_i = \frac{1}{M} \sum_{j=1}^M x_{ij} \quad (4)$$

$$Y_i = \frac{1}{M} \sum_{j=1}^M y_{ij} \quad (5)$$

Ilustrasi komponen x dan y dalam metode DRC seperti disajikan pada gambar 1.



Gambar 1. Plot data komponen x dan y (Bishnu dan Bhattacharjee, 2012)

Percobaan yang dilakukan Bishnu dan Bhattacharjee (2012) adalah dengan membandingkan penggunaan DRC sebagai reduksi dimensi pada set data sebelum dilakukan clustering. Perbandingan dilakukan terhadap metode Single K-Means tanpa reduksi dimensi, dan Single Decision Tree K-Means (SDTK). Pengujian dilakukan pada 3 set data publik yang diunduh dari UCI Machine Learning Repository yaitu Iris, Wine, dan Glass. Hasilnya menunjukkan bahwa jumlah iterasi proses clustering yang dilakukan menjadi berkurang dibandingkan dengan 2 metode yang lain, error (ketidak sesuaian hasil cluster dengan label kelas yang sebenarnya) yang terjadi relatif lebih kecil dibanding yang lain, yaitu mencapai 6.66% hingga 39.71%. Jumlah dimensi yang dihasilkan juga paling kecil yaitu selalu dua fitur, *X* dan *Y*.

3. KERANGKA KERJA PENELITIAN DAN PENGEMBANGAN METODE

Penelitian ini melakukan pengamatan kinerja metode DRC ketika diterapkan pada kasus klasifikasi, dimana fungsi DRC adalah untuk melakukan reduksi dimensi sebelum menerapkan metode klasifikasi. Dalam penelitian ini, dilakukan uji coba penambahan fitur hasil reduksi dalam DRC yaitu komponen *z* (tangen). Uji coba juga dilakukan terhadap DRC tanpa komponen *z* (DRC 2 DIM) untuk mengetahui pengaruh penggunaan komponen *z* dalam mempertahankan keberadaan karakteristik data.

Kerangka kerja penelitian yang dilakukan seperti yang disajikan dalam makalah ini sebagai berikut:

1. Melakukan studi literatur metode-metode reduksi dimensi.
 Pada tahap ini penulis melakukan studi terhadap sejumlah metode-metode reduksi dimensi seperti, PCA, Singular Value Decomposition (SVD), FastMap, ISOMAP, dan DRC.
2. Melakukan pengamatan kinerja DRC murni ketika diterapkan sebagai reduksi dimensi dalam kasus klasifikasi.
 Pada tahap ini penulis melakukan pengamatan kinerja DRC ketika digunakan sebagai reduksi dimensi pada data yang akan dilakukan proses klasifikasi. Pengamatan yang dilakukan penulis menunjukkan bahwa DRC memberikan pengaruh dalam mengurangi waktu yang dibutuhkan untuk proses pelatihan model, tetapi akurasi prediksi yang didapatkan ternyata relatif lebih kecil dibandingkan tanpa reduksi dimensi.
3. Melakukan eksperimen penggunaan fitur komponen *z* (tangen) untuk mempertahankan karakteristik data karena proses reduksi dimensi yang dilakukan DRC.
 Pada penelitian ini, penulis juga mencoba menambahkan satu komponen sebagai komponen ketiga, yaitu *z* (tangen). Komponen ini didapatkan dengan menghitung nilai fitur ke-*j* dari data ke-*i* yang dikalikan dengan tangen dari sudut θ yang digunakan fitur ke-*j*. Selanjutnya disebut dengan DRC 3 DIM.
4. Melakukan uji perbandingan DRC murni (DRC 2 DIM), DRC dengan tambahan komponen *z* (DRC 3 DIM), dan PCA ketika diterapkan sebagai reduksi dimensi pada klasifikasi dengan metode Support Vector Machine (SVM).

Pada tahap ini, penulis melakukan uji banding kinerja antara metode klasifikasi dengan tiga tipe: tanpa reduksi dimensi, menggunakan reduksi dimensi dengan PCA, menggunakan reduksi dimensi dengan DRC murni, dan menggunakan DRC dengan tambahan komponen z (tangen).

- Menarik kesimpulan kinerja DRC dengan komponen z (tangen) dalam mempengaruhi kinerja klasifikasi.

Dalam penelitian ini, penulis melakukan penambahan komponen z (tangen) sebagai salah satu fitur hasil reduksi. Nilainya didapatkan dari hasil logaritma 10 dari hasil perkalian nilai fitur terhadap tangen dari sudut θ . Formula yang digunakan seperti pada persamaan (6)

$$z_{ij} = \log_{10}(r_{ij} \tan \theta_j) \quad (6)$$

Sedangkan untuk fitur ketiga sebagai fitur hasil reduksi didapatkan dengan menghitung rata-rata dari semua komponen z pada data ke- I , seperti pada persamaan (7).

$$Z_i = \frac{1}{M} \sum_{j=1}^M z_{ij} \quad (7)$$

Dengan penambahan fitur Z tersebut diharapkan semakin bertambah karakter data yang muncul pada hasil reduksi dimensi.

4. HASIL DAN ANALISIS

4.1. Set data dan parameter pengujian

Metode DRC yang sudah ditambah fitur Z tersebut kemudian dilakukan pengujian kinerja, baik akurasi prediksi, maupun waktu komputasi saat pelatihan dan prediksi. Perbandingan dilakukan terhadap DRC murni, DRC dengan tambahan komponen z (tangen), dan PCA ketika diterapkan sebagai reduksi dimensi pada klasifikasi dengan metode Support Vector Machine (SVM). Parameter fungsi kernel yang digunakan oleh SVM adalah rbf.

Pengujian dilakukan terhadap 6 set data publik yang diunduh dari UCI Machine Learning Repository yaitu: Iris (150 record, 4 fitur), Vertebral Column (310 record, 6 fitur), Wine (178 record, 13 fitur), Glass (214 record, 9 fitur), Blood (748 record, 4 fitur), dan Ionosphere (351 record, 33 fitur). Sistem pengujian menggunakan 4-fold Cross Validation, dimana 75% digunakan sebagai data latih dan 25% digunakan sebagai data uji. Kinerja diukur dari nilai rata-rata dari 4 sesi Cross Validation.

Karena data-data pada setiap fitur mempunyai jangkauan nilai yang berbeda, maka dilakukan pra-pemrosesan yaitu normalisasi. Sebelum dilakukan proses pengujian, semua data pada setiap fitur dilakukan normalisasi. Normalisasi yang digunakan adalah zero-mean, sesuai syarat yang diminta oleh PCA.

Hasil pengujian disajikan pada tabel 1, 2, 3, 4, 5, dan 6. Baris metode SVM merupakan pengujian kinerja dengan menerapkan metode SVM tanpa reduksi dimensi, baris metode PCA-SVM merupakan pengujian dimana set data dilakukan reduksi dimensi terlebih dahulu dengan PCA kemudian data hasil reduksi dimensi dimasukkan ke metode SVM untuk menguji kinerja. Baris DRC 2 DIM – SVM serupa dengan PCA – SVM tetapi metode reduksi dimensi menggunakan 2 komponen seperti yang diusulkan oleh Bishnu dan Bhattacharjee (2012). Sedangkan baris DRC 3 DIM – SVM serupa dengan PCA – SVM tetapi metode reduksi dimensi menggunakan 3 komponen seperti yang diusulkan dalam makalah ini.

Kolom Dimensi merupakan data jumlah dimensi yang diolah oleh SVM baik tanpa reduksi dimensi maupun setelah melalui reduksi dimensi. Nilai untuk SVM selalu sama dengan jumlah dimensi awal, nilai untuk PCA – SVM didapat dari jumlah komponen utama PCA (dari besar ke kecil) yang secara akumulasi > 95% variasi data hasil transformasi (dilihat dari nilai latent / eigen value setiap komponen utama). Sedangkan DRC 2 DIM – SVM dan DRC 3 DIM – SVM masing-masing jumlah dimensinya selalu tetap 2 dan 3.

Kolom Akurasi merupakan data akurasi kinerja ketika melakukan prediksi dengan teknik partisi data 4-fold Cross Validation, nilainya didapatkan dari rata-rata prediksi 4 sesi partisi data. Kolom Waktu Training merupakan waktu yang dibutuhkan oleh SVM untuk melakukan pelatihan data latih sehingga mendapatkan support vector-nya, nilai juga rata-rata waktu pelatihan 4 sesi partisi data, satuan waktu yang disajikan adalah mili detik. Kolom Waktu Prediksi merupakan

waktu yang dibutuhkan oleh SVM untuk melakukan prediksi pada data uji, nilai juga rata-rata waktu prediksi da uji 4 sesi partisi data, satuan waktu yang disajikan adalah mili detik.

Baris paling bawah menyatakan poin hasil perbandingan antara DRC 3 DIM – SVM terhadap 3 metode pembanding, dinyatakan dengan tanda “*”, tanda “***” menyatakan DRC 3 DIM – SVM unggul dibanding 3 metode yang lain (tertinggi/tercepat), tanda “**” menyatakan DRC 3 DIM – SVM unggul dibanding 2 metode yang lain (tinggi/cepat), tanda “*” menyatakan DRC 3 DIM – SVM unggul dibanding 1 metode yang lain (pertengahan), tanda “-” menyatakan DRC 3 DIM – SVM kalah dibanding semua metode yang lain (rendah/lama).

4.2. Hasil pengujian dan analisis

Hasil pengujian untuk set data Iris disajikan pada tabel 1. Dengan menurunnya jumlah dimensi dalam metode DRC 3 DIM – SVM akurasi masih dapat dipertahankan jika dibandingkan dengan SVM murni (95.32% vs 94.01%) dan unggul terhadap PCA – SVM dan DRC 2 DIM – SVM (dua bintang dari skala tiga bintang). Sedangkan waktu pelatihan termasuk level pertengahan (dua bintang dari skala tiga bintang), dan prediksi termasuk lebih lama dibanding yang lain, bahkan oleh SVM murni yang menggunakan seluruh fitur baik untuk pelatihan maupun prediksi.

Tabel 1. Hasil pengujian kinerja pada set data Iris

Metode	Dimensi	Akurasi (%)	Waktu Training (mili detik)	Waktu Prediksi (mili detik)
SVM	4	95.32	95.95	0.93
PCA – SVM	2	93.99	132.47	0.93
DRC 2 DIM - SVM	2	88.02	70.68	0.92
DRC 3 DIM - SVM	3	94.01	109.91	0.94
<i>Unggul</i>		**	*	-

Hasil pengujian untuk set data Vertebral Column disajikan pada tabel 2. Dengan menurunnya jumlah dimensi dalam metode DRC 3 DIM – SVM akurasi masih dapat dipertahankan jika dibandingkan dengan terhadap PCA – SVM dan DRC 2 DIM – SVM , tetapi nilai akurasi 81.6% masih kalah jika dibandingkan dengan SVM murni 86.11%, tetapi unggul dibanding 2 metode lainnya (dua bintang). Untuk Waktu Training dan Waktu Prediksi, DRC 3 DIM – SVM berada pada level pertengahan diantara metode pembanding.

Tabel 2. Hasil pengujian kinerja pada set data Vertebral Column

Metode	Dimensi	Akurasi (%)	Waktu Training (mili detik)	Waktu Prediksi (mili detik)
SVM	6	86.11	442.10	1.51
PCA - SVM	4	80.95	228.86	1.59
DRC 2 DIM - SVM	2	75.81	51.08	1.56
DRC 3 DIM - SVM	3	81.60	376.72	1.56
<i>Unggul</i>		**	*	*

Hasil pengujian untuk set data Wine disajikan pada tabel 3. Hasil yang berbeda dari set data sebelumnya ada di set data Wine. Akurasi tertinggi terdapat di SVM dengan PCA, sedangkan akurasi DRC 3 DIM – SVM lebih tinggi dibanding SVM murni dan SVM dan DRC 2 DIM – SVM (dua bintang). Untuk Waktu Training pada DRC 3 DIM – SVM tetap pada level pertengahan, sedangkan Waktu Prediksi masuk kategori cepat.

Tabel 3. Hasil pengujian kinerja pada set data Wine

Metode	Dimensi	Akurasi (%)	Waktu Training (mili detik)	Waktu Prediksi (mili detik)
SVM	13	81.36	12.17	1.46
PCA – SVM	5	93.80	119.34	1.04
DRC 2 DIM - SVM	2	78.65	37.27	1.10
DRC 3 DIM - SVM	3	82.08	55.69	1.08
<i>Unggul</i>		**	*	**

Hasil pengujian untuk set data Glass disajikan pada tabel 4. Hasil berbeda juga dapat dilihat pada hasil pengujian pada set data Glass. Akurasi DRC 3 DIM – SVM adalah yang tertinggi (tiga bintang) dari metode pembandingan (93.94%). Sedangkan Waktu Training berada pada level paling lama. Sedangkan Waktu Prediksi berada pada level tertinggi.

Tabel 4. Hasil pengujian kinerja pada set data Glass

Metode	Dimensi	Akurasi (%)	Waktu Training (mili detik)	Waktu Prediksi (mili detik)
SVM	9	90.21	244.87	1.31
PCA – SVM	5	92.53	281.17	1.06
DRC 2 DIM - SVM	2	91.15	164.36	1.10
DRC 3 DIM - SVM	3	93.94	291.78	1.03
<i>Unggul</i>		***	-	***

Hasil pengujian untuk set data Blood disajikan pada tabel 5. Pada hasil pengujian di set data ini, akurasi DRC 3 DIM – SVM berada pada level pertengahan. Sedangkan Waktu Training paling cepat dibanding metode yang lain, sedangkan Waktu Prediksi berada pada level pertengahan.

Tabel 5. Hasil pengujian kinerja pada set data Blood

Metode	Dimensi	Akurasi (%)	Waktu Training (mili detik)	Waktu Prediksi (mili detik)
SVM	4	79.68	1,006.56	7.21
PCA – SVM	3	79.68	1,079.07	6.89
DRC 2 DIM - SVM	2	78.21	680.19	6.82
DRC 3 DIM - SVM	3	79.28	667.73	6.93
<i>Unggul</i>		*	***	*

Hasil pengujian untuk set data Ionosphere disajikan pada tabel 6. Set data Ionosphere merupakan set data dengan jumlah fitur paling besar. Disini, kinerja DRC baik 2 dimensi maupun 3 dimensi jatuh jauh dari harapan (79.21%), hal ini diduga adalah akibat hilangnya variasi fitur akibat reduksi yang dipaksakan menjadi 2 dan 3 (dari 33 menjadi 2 dan 3). Sehingga untuk set data berukuran besar, kiranya metode DRC baik 2 atau 3 dimensi perlu dikaji lebih lanjut. Untuk Waktu Training, DRC 3 DIM – SVM berada pada level pertengahan (2 bintang), tetapi beda tipis terhadap DRC 2 DIM. Demikian pula untuk Waktu Prediksi, DRC 3 DIM unggul dibanding 3 metode lainnya tetapi beda tipis terhadap DRC 2 DIM.

Tabel 6. Hasil pengujian kinerja pada set data Ionosphere

Metode	Dimensi	Akurasi (%)	Waktu Training (mili detik)	Waktu Prediksi (mili detik)
SVM	33	92.02	439.50	1.56
PCA – SVM	5	92.31	622.32	1.59
DRC 2 DIM – SVM	2	79.21	148.74	1.43
DRC 3 DIM – SVM	3	79.21	149.73	1.42
<i>Unggul</i>		-	**	***

Dari hasil pengujian secara empiris pada 6 set data tersebut, maka dapat diringkas khusus untuk perilaku DRC 3 DIM dalam kinerjanya mereduksi dimensi terkait dengan kinerja Akurasi, Waktu Training dan Waktu Prediksi, seperti disajikan pada tabel 7. Dari data yang disajikan pada tabel 7, dapat diamati bahwa kinerja akurasi DRC 3 DIM dalam mereduksi dimensi pada metode klasifikasi SVM mampu memberikan akurasi yang tetap relatif baik (Pertengahan hingga Tertinggi) ketika jumlah dimensi awal masih tidak banyak, tetapi ketika digunakan untuk mereduksi data dengan dimensi tinggi seperti Ionosphere, kinerja akurasi metode klasifikasi SVM menjadi rendah. Hal ini diduga adalah jumlah variasi fitur yang banyak telah hilang akibat reduksi yang selalu tetap menjadi 3 dimensi pada semua fitur set data. Hal ini menjadikan hasil rekomendasi yang diberikan apakah metode DRC 3 DIM dapat digunakan untuk reduksi dimensi hanya diberikan pada 5 set data, sedangkan untuk Ionosphere tidak direkomendasikan, meskipun Waktu Training dan Waktu Prediksi tergolong cepat, 2 parameter kinerja tersebut dapat dinisbikan karena pada akhirnya yang akan tampak dari luar dalam penilaian kinerja adalah ketepatan dalam prediksi (akurasi), bukan waktu komputasi.

Tabel 7. Perilaku kinerja DRC 3 DIM

Set data	Dimensi Awal	Akurasi	Waktu Training	Waktu Prediksi	Rekomendasi Penggunaan
Iris	4	Tinggi	Pertengahan	Lama	Ya
Vertebral Colum	6	Tinggi	Pertengahan	Pertengahan	Ya
Wine	13	Tinggi	Pertengahan	Cepat	Ya
Glass	9	Tertinggi	Lama	Tercepat	Ya
Blood	4	Pertengahan	Tercepat	Pertengahan	Ya
Ionosphere	33	Rendah	Cepat	Cepat	Tidak

Dari penyajian yang diberikan pada bagian 4 ini dapat diringkas bahwa metode DRC 3 DIM yang diusulkan dalam makalah ini dapat digunakan untuk reduksi dimensi set data dengan dimensi yang tidak tinggi. Untuk set data dengan dimensi yang tinggi masih perlu kajian penelitian lebih lanjut dapat memberikan hasil reduksi dimensi.

5. KESIMPULAN

Secara umum, DRC 3 DIM memberikan kinerja reduksi yang relatif baik dalam mereduksi dimensi sekaligus tetap mempertahankan akurasi prediksi oleh metode klasifikasi ketika menggunakan data hasil reduksi, meskipun masih ada sejumlah kekurangan. Dari penelitian yang dilakukan, maka dapat disimpulkan sebagai berikut:

1. Metode DRC 3 DIM relatif dapat digunakan untuk melakukan reduksi dimensi pada data dengan jumlah dimensi yang tidak besar.
2. Akurasi kinerja yang diberikan oleh metode klasifikasi ketika menggunakan data hasil reduksi dimensi dengan DRC 3 DIM masih dapat dinilai pada level pertengahan hingga tertinggi dibandingkan dengan tanpa reduksi dimensi atau metode reduksi dimensi yang lain.

3. Waktu komputasi yang dibutuhkan baik untuk training maupun prediksi masih dapat ditoleransi untuk dapat digunakan mengingat waktu training dan prediksi berada pada mayoritas level pertengahan ketika dibandingkan dengan metode pembandingan.

Penelitian yang sudah dilakukan dan disajikan dalam makalah ini ada kelebihan dan kekurangan, maka saran yang dapat penulis berikan pada penelitian berikutnya untuk memperbaiki kekurangan yang ditemukan adalah sebagai berikut:

1. Perlu kajian lebih lanjut untuk mengetahui perilaku kinerja yang diberikan pada metode ini ketika bekerja pada set data dengan dimensi yang tinggi.
2. Perlu uji banding terhadap metode reduksi dimensi yang lain seperti LLE, Fastmap, ISOMAP, SVD, dan metode-metode yang lain.
3. Kajian uji kinerja menggunakan hasil transformasi DRC 3 DIM pada metode klasifikasi yang lain masih perlu dilakukan, seperti pada metode Artificial Neural Network, Decision Tree, dan sebagainya.

DAFTAR PUSTAKA

- Bishnu, P.S., Bhattacharjee, V. (2012), *A Dimension Reduction Technique for K-Means Clustering Algorithm*, 1st International Conference on Recent Advances in Information Technology.
- Cox, T. dan Cox, M. (1994), *Multidimensional scaling*, Chapman & Hall, London, UK.
- Faloutsos, C. dan Lin, K.I. (1995), FastMap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets, *In Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data*, pages 163–174, New York, NY, USA, ACM Press.
- Prasetyo, E. (2012), *Data Mining – Konsep dan Aplikasi Menggunakan Matlab*, edisi 1, Andi Offset: Yogyakarta.
- Roweis, S.T. dan Saul, L.K. (2000), *Nonlinear dimensionality reduction by Locally Linear Embedding*, Science, 290(5500):2323–2326.
- Tan, P.N., Steinbach, M., Kumar, V. (2006), *Introduction to Data Mining*, 1st Ed, Pearson Education: Boston San Fransisco New York.
- Tenenbaum, J.B. (1998), *Mapping a manifold of perceptual observations*, In *Advances in Neural Information Processing Systems*, volume 10, pages 682–688, Cambridge, MA, USA.
- UCI Machine Learning Repository , 1 Juni 2014, <http://archive.ics.uci.edu/ml/datasets.html>