

EKSTRAKSI CIRI DAN KLASIFIKASI DATA EKG MULTIVARIATE TIME SERIES

Apriyani Wulansari¹⁾, Edi Winarko²⁾

¹⁾ Program S2 Ilmu Komputer, FMIPA, UGM, Yogyakarta

²⁾ Jurusan Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta

*¹⁾apriyaniwulansari7@gmail.com, ²⁾ewinarko@ugm.ac.id

ABSTRAK

Penelitian ini melakukan ekstraksi ciri terhadap data EKG *multivariate time series*. Ekstraksi ciri merupakan langkah awal untuk melakukan klasifikasi data EKG. Tujuan dilakukan ekstraksi ciri untuk mengambil ciri-ciri yang terdapat pada objek di dalam EKG tersebut. Permasalahan utama dari penelitian ini adalah bagaimana menggunakan metode statistik yaitu, *mean*, *covariance*, dan *kurtosis* untuk ekstraksi ciri data EKG dari *dataset multivariate time series* sehingga nilai yang diperoleh bisa mewakili informasi *time series*. Tahapan berikutnya adalah tahapan klasifikasi yang menggunakan *Support Vector Machine* (SVM) yang dapat menghasilkan akurasi yang baik. Metode penelitian ini menggunakan metode SVM dimana metode ini menggunakan kernel polynomial untuk mengklasifikasikan data EKG yang telah diekstraksi ciri. Penelitian ini juga membandingkan penggunaan kernel SVM lainnya seperti RBF dan *linear*. Hasil dari data latih menunjukkan bahwa klasifikasi data EKG *multivariate time series* menggunakan *Support Vector Machine* (SVM) memberikan hasil akurasi 81,25% dengan kernel polynomial, penggunaan kernel RBF menghasilkan akurasi 95%, dan kernel *linear* menghasilkan 86,25%. Data uji mendapatkan akurasi 55%.

Kata kunci: Ekstraksi ciri, klasifikasi data EKG *multivariate time series*, metode statistik, SVM

PENDAHULUAN

Ekstraksi ciri-ciri ataupun dikenal ekstraksi fitur merupakan langkah awal yang dilakukan sebelum melakukan klasifikasi data EKG. Tujuan dilakukan ekstraksi ciri untuk mengambil ciri-ciri yang terdapat pada objek di dalam EKG tersebut. Objek EKG yang akan diekstraksi cirikan pada penelitian ini yaitu data temporal yang disajikan dalam bentuk *multivariate time series*. Metode statistik salah satu metode yang biasa digunakan untuk melakukan ekstraksi ciri terhadap objek tertentu. Kristomo (2014) menyatakan metode statistik adalah suatu metode yang efektif, karena dengan metode statistik

ekstraksi ciri dapat dilakukan pada ranah waktu, komputasi lebih sederhana karena tidak adanya proses transformasi.

Kutlu dan Kuntalp (2012) melakukan ekstraksi fitur menggunakan *Higher Order Statistics* (HOS) dari paket *wavelet packet dekomposisi* (WPD) nilai koefisiennya untuk pengenalan detak jantung. Hasil akurasi sistem yang diperoleh rata-rata sensitifitas 90%, rata-rata selektifitas 92%, dan rata-rata spesifitas 98%. Hasil dari penerepan pada HOS dari WPC pada fitur *highly discriminative* untuk klasifikasi untuk perbedaan aritmia EKG. Berbeda yang dilakukan oleh Banerjee dan Mitra (2014) membuat

klasifikasi yang sebelumnya melakukan ekstraksi fitur dan klasifikasi EKG tanpa membutuhkan konvensional fitur *time-plane* dan ekstraksi dari data *noisy*. Tujuan klasifikasi data EKG untuk mengklasifikasikan pola jantung normal dan abnormal. Algoritme yang diusulkan untuk analisis EKG menggunakan *cross-wavelet transform* (XWT) dan *wavelet coherence* (WCOH) dimana lokalitas similaritas diobservasi.

Beberapa penelitian dalam domain EKG telah banyak dilakukan diantaranya dilakukan oleh Thanapatay dan Suwansaroj (2010) bahwa pengklasifikasian menggunakan kertas EKG dari MIT-BIH dengan metode SVM dengan transform diskret wavelet dan PCA. Untuk mengekstraksi bentuk EKG menggunakan level 3 *dubechies* 1 sebagai filter digital untuk menghilangkan suara sinyal digital. Hasil yang diperoleh 99,6367% dengan LIBSVM. Sedangkan Castillo *et al.* (2012) mengklasifikasikan beberapa tipe aritmia. Sinyal pada aritmia disegmentasi dan ditransformasikan untuk meningkatkan hasil klasifikasi. Pendekatan yang digunakan dalam pengklasifikasian aritmia adalah menggunakan *hybrid intelligent system* yaitu *Fuzzy K-Nearest Neighbors*, *Multi Layer Perceptron* dengan *Gradient Descent* dan *Momentum Backpropagation*, *Multi Layer Perceptron* dengan *Scaled Conjugate Gradient Backpropagation*, dan Inferensi Fuzzy Mamdani.

Konsep *multivariate time series* adalah urutan terbatas dari *univariate time series* (Liu, *et al.*, 2016). Penelitian *multivariate time series* yang sudah diteliti adalah bidang keuangan, kedokteran, multimedia, *gesture recognition*, *video sequence matching* menggunakan *computer vision*. Penelitian untuk membedakan detak

jantung normal dan abnormal menggunakan EKG pernah dilakukan begitu juga penelitian analisis EEG.

Maharaj dan Alonso (2014) mengklasifikasikan data EKG *multivariate time series* tujuannya untuk membedakan pola kondisi jantung sehat dan memiliki spesifik kondisi jantung sesorang. Metode yang dilakukan untuk mengklasifikasikan sinyal EKG *multivariate time series* yaitu diskriminan dan analisis *wavelet*. Yu-Feng *et al.* (2015) membuat metodologi baru yang disebut REACT (*Reable EARly Classification*) untuk penggunaan *multivariate time series* dengan atribut *categorical* dan *attributes* disingkat dengan MTS-NC. Metode ini diterapkan pada teknik paralel GPU *computing* untuk mengembangkan mekanisme pembentukan diawal secara efisien. Penerepan metode ini dicobakan pada *datasets* MS70, robot, ECG, wafer, asthma. Hasilnya bahwa penelitian ini lebih awal dan efisien dalam pembentukan *classifier*.

Mehta dan Lingayat (2008) melakukan pengenalan gelombang QRS pada EKG. SVM digunakan sebagai klasifikasi untuk menggambarkan QRS dan area non-QRS. Evaluasi yang digunakan CSE ECG *database*. Hasil penelitian menunjukkan algoritme mencapai 99.3% dari tingkat deteksi. Berbeda penelitian dilakukan Ramirez *et al.* (...) membuat sistem klasifikasi otomatis untuk kematian jantung mendadak (SCD) dan kegagalan dalam pemompaan dari otonom (PFD) dan gagal jantung (CHF) repolarisasi EKG menggunakan metode SVM. Evaluasi terhadap sistem menggunakan *cross-validation* untuk mengetahui performa metode yang digunakan. Hasil yang diperoleh terbagi menjadi tiga kelompok, kombinasi optimal diperoleh persentase 79%.

Aziz *et al.* (2016) membuat sistem klasifikasi pada gen pada *dataset microarray*. Sebelum melakukan klasifikasi peneliti melakukan seleksi fitur menerapkan metode *independent component analysis* dan *fuzzy backward feature elimination* (FBFE) Metode yang diterapkan untuk klasifikasi adalah metode *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB) *classifier*. Pengujian dilakukan menggunakan lima *dataset* gen *microarray* menggunakan AUC. Hasil akurasi *mean* yang diperoleh 90.90% untuk 30 fitur yang terpilih dan 85.46% untuk 25 gen terpilih dengan menerapkan metode SVM dan NB *classifier*.

METODE PENELITIAN

Deskripsi Sistem

Sistem yang dirancang dan dibangun dalam penelitian ini klasifikasi *multivariate time series* yang sebelumnya dilakukan ekstraksi ciri pada EKG. Metode yang digunakan untuk melakukan ekstraksi ciri adalah metode statistik yaitu *mean*, *covariance*, dan *kurtosis* sedangkan untuk klasifikasi metodenya untuk penelitian ini adalah *support vector machine* (SVM). Tujuan dari penelitian ini untuk melakukan ekstraksi ciri dari *dataset EKG multivariate time series* dan mengklasifikasikannya sehingga dapat diketahui labelnya yaitu detak jantung normal dan abnormal.

Tahapan pada penelitian ini terbagi menjadi 3 tahapan seperti dalam Gambar 1 Tahapan penelitian meliputi pengambilan data EKG angka *multivariate time series*, ekstraksi ciri, dan klasifikasi menggunakan *Support Vector Machine* (SVM).



Gambar 1. Tahapan ekstraksi ciri dan klasifikasi data ekg multivariate time series

Data Penelitian

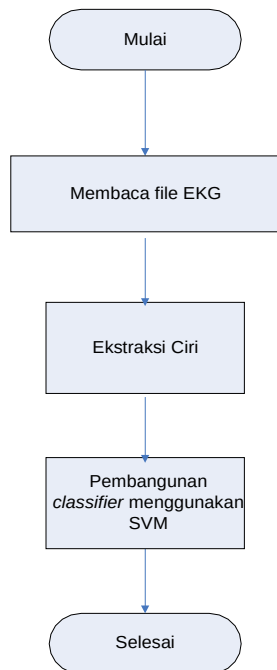
Data yang digunakan untuk pengklasifikasian pada penelitian ini adalah *dataset EKG multivariate time series* diambil dari *homepage* : www.mustafabaydogan.com/files/viewcategory/15-multivariate-time-series-classification.html.

Dataset EKG multivariate time series merupakan *dataset* berupa angka terdiri dari lima kolom yang terdiri dari kolom pertama yaitu *series* terdiri dari 100 *series*, *time index*, nilai-nilai dari *series* dan label ataupun kelas. Kelas yang terdapat pada *dataset* yaitu kelas normal dan abnormal. Detak jantung abnormal yang dalam *dataset* adalah *supraventricular premature beat*. EKG *database* terdiri dari kumpulan koleksi data *time series* dimana setiap *file* berisi urutan pengukuran satu elektrode selama satu detak jantung.

Rancangan Proses

Pada perancangan umum tersajikan dalam *flowchart* utama untuk menggambarkan klasifikasi data EKG *multivariate time series*. Penelitian ini langkah pertama yang dilakukan adalah melakukan ekstraksi ciri. *Dataset* ekg yang digunakan dalam format txt. *Dataset* ini terdiri dari lima kolom, pada kolom ke-4 dan kolom ke-5 yang merupakan nilai *series* untuk dilakukan ekstraksi ciri. Fitur yang baru merupakan hasil dari ekstraksi ciri digunakan sebagai input untuk melakukan klasifikasi. Metode untuk klasifikasi menggunakan *support vector machine* (SVM). Berikut Gambar 2

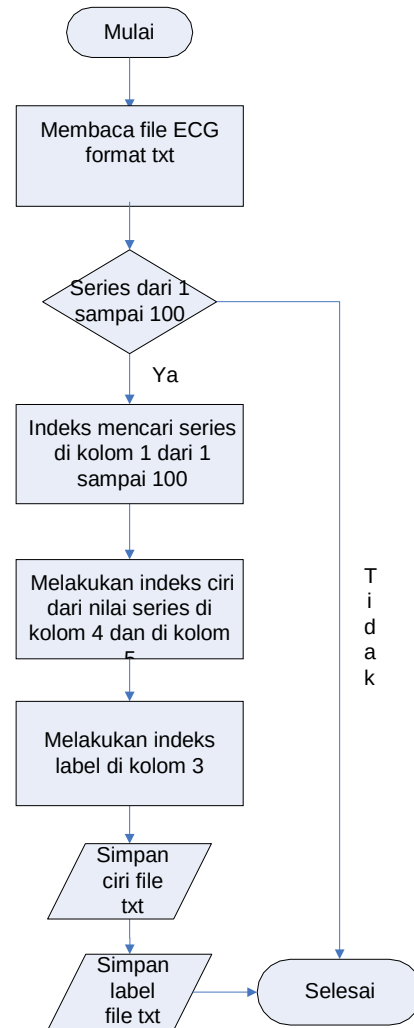
adalah *flowchart* utama utama klasifikasi data EKG.



Gambar 2. *Flowchart* utama klasifikasi data EKG multivariate time series

Ekstraksi Ciri

Ekstraksi ciri dilakukan dengan menggunakan menggunakan metode statistik yaitu *mean*, *covariance* dan *kurtosis*. Ekstraksi ciri dilakukan pada nilai *series* yang terdapat pada kolom ke-4 dan kolom ke-5 pada *dataset*. Tujuan dilakukan pengindeksan agar menemukan *series* dan lebel sehingga dapat menghitung ekstraksi ciri pada kolom ke-4 dan kolom ke-5. Pada Gambar 2 merupakan *flowchart* dalam melakukan ekstraksi ciri. Berikut Gambar 3 merupakan *flowchart* ekstraksi ciri.



Gambar 3. *Flowchart* membaca file dan pemisahan setiap series ciri dan label.

Mean

Menurut Larose (2005) *mean* dari suatu variabel adalah rata-rata (*average*) yang diambil oleh variabel. Nilai rata-rata dapat dicari dengan persamaan:

$$\bar{x} = \sum_{i=1}^N X_i / n$$

Keterangan:

$\bar{x}$ = Rata-rata / Mean

X_i = Nilai ke i

n

Keterangan:

μ = Rata-rata / Mean

X_i = Nilai ke i

N = Jumlah data

Covariance

Menurut Tan *et al.* (2006) data yang berupa variabel kontinu maka penyebaran datanya menggunakan *covariance* matriks S , dimana i dan j dimasukan menjadi S_{ij} merupakan *covariance* atribut i dan j dari data. Jika x_i dan x_j merupakan atribut dari i dan j , maka

$$S_{ij} = \text{covariance}(x_i, x_j)$$

Sehingga *covariance* (x_i, x_j) bentuk persamaan sebagai berikut yaitu persamaan 3.

$$\text{covariance}(x_i, x_j) = \frac{1}{m-1} \sum_{k=1}^m (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$$

Keterangan:

x_i dan x_j : Nilai atribut i dan j untuk objek k

$\bar{x}_i$ = Mean untuk i

$\bar{x}_j$ = Mean untuk j

M = Jumlah data

Kurtosis

Menurut Hasan (1999) Kurtosis adalah tingkat kepuncakan dari sebuah distribusi yang biasanya diambil secara relatif terhadap suatu distribusi normal. Kurtosis persamaan:

$$K = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right)^2}$$

Keterangan:

K = Kurtosis.

$\bar{X}$ = Rata-rata / Mean

N = Jumlah sampel

X_i = Nilai ke i

Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu metode *supervised learning*. support vector machine (SVM) bekerja pada ruang vektor, dimana tujuannya untuk memisahkan kelas dengan *hyperplane*. Menurut Pang-Ning *et al.* (2006) SVM memiliki beberapa perbedaan dalam penanganan sampel kecil, non-

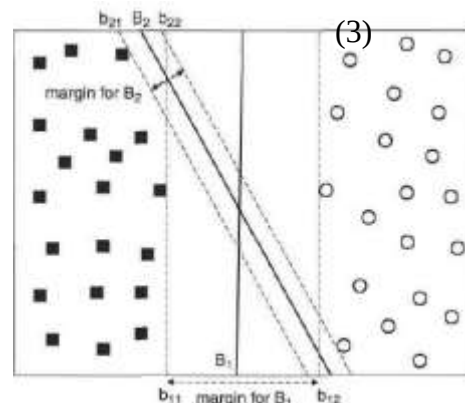
resistence, permasalahan pada *high dimensional* menjadikan SVM algoritme *machine learning* yang baik.

SVM diusulkan pertama kali oleh Vapnik yang telah dipelajari untuk klasifikasi, regresi dan estimasi kepadatan SVM pada intinya merupakan pengklasifikasian untuk memaksimalkan margin antara data *training* dan *decision boundary* atau memisahkan *hyperplane* secara optimal. Aspek lain pendekatan yang dilakukan adalah *decision boundary* menggunakan subset dari contoh *training*, disebut dengan *support vector*.

(3)

Maximum Margin Hyperplane

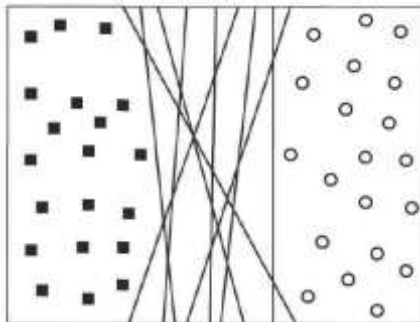
Dataset yang digunakan adalah *linear separable* contohnya bagaimana menemukan *hyperplane* yang kotak terletak pada satu sisi dari *hyperplane* dan semuanya lingkaran terletak disisi lain ditunjukkan pada Gambar 4. Pada Gambar 4 terlihat banyak kemungkinan adanya *hyperplane*. Classifier harus dipilih satu dari *hyperplane* untuk merespresentasikan *decision boundary*, berdasarkan seberapa baik diharapkan untuk performa dalam contoh tes.



Gambar 4. Margin decision boundary

Perbedaan jelas terlihat pada Gambar 5 bagaimana perbedaan pemilihan *hyperplane* yang merupakan efek generalisasi *error*, dengan pertimbangan dua *decision boundary*

yaitu B_1 dan B_2 . *Decision boundary* dalam hal ini kedua-duanya bisa dipisahkan dari contoh *training* kedalam masing-masing kelas dengan tidak melakukan error kesalahan klasifikasi. Setiap *decision boundary* B_i terasosiasi dengan *hyperplane* yang berpasangan ditunjukkan sebagai b_{i1} dan b_{i2} secara berturut-turut. b_{i1} diperoleh dengan memindahkan *hyperplane* dengan menjauhan dari *decision boundary* sampai bersentuhan dengan terdekat yaitu kotak, dimana b_{i2} diperoleh dengan memindahkan *hyperplane* sampai bersentuhan yang terdekat lingkaran. Jarak antara dua *hyperplane* ini diketahui sebagai margin klasifikasi. Margin dari B_1 lebih besar dari pada B_2 terlihat pada Gambar 5. B_1 beralih ke maximum margin *hyperplane* dari *training instance* pada contoh *training* ini.



Gambar 5. Kemungkinan *decision boundary* untuk dataset linier separable

SVM Linaer Separable

SVM linier merupakan klasifikasi yang melakukan pencarian untuk *hyperplane* dengan margin terbesar, diketahui sebagai *maximal margin classifier*.

Linear Decision Boundary

Permasalahan dari klasifikasi binary yang terdiri dari N contoh *training*. Setiap contoh ini ditunjukkan oleh baris (x_i, y_i) ($i = 1, 2, \dots, N$), dimana $x_i = (x_{i1}, x_{i2}, \dots, x_{id})^T$ berkoresponden dari kumpulan atribut. Suatu $y_i \in \{-1, 1\}$

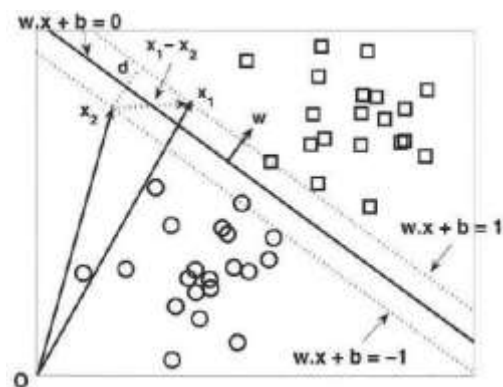
merupakan label. *Decision boundary* dapat dari suatu *linear classifier* dapat ditulis seperti bentuk berikut :

$$w \cdot x + b = 0,$$

dimana :

w, b = model dari parameter

Pada contoh di Gambar 6 kotak dan lingkaran merupakan dua dimensi dalam *training set*. *Decision boundary* yang membagi contoh pelatihan kedalam masing-masing kelas yang diilustrasikan dengan garis penuh.



Gambar 6. *Decision boundary* dan margin SVM

Berikut adalah contoh *decision boundary* pada persamaan (6).

$$w \cdot x_a + b = 0,$$

$$w \cdot x_b + b = 0.$$

Persamaan (6) menjelaskan jika x_a dan x_b ada dua titik dilokasi *decision boundary*.

Persamaan (7) menjelaskan apabila dikurangi, berikut persamaannya.

$$w \cdot (x_a - x_b) = 0,$$

dimana $x_a - x_b$ sama dengan vektor paralel untuk *decision boundary* dan langsung dari x_a ke x_b . Sejak dot product adalah zero, arah ke w harus tegak lurus pada *decision boundary*, terlihat pada gambar 7.

Pada gambar kotak x_s yang berada di atas *decision boundary*, diperlihatkan pada persamaan (7) berikut ini:

$$w \cdot x_s + b = k,$$

dimana $k > 0$. Begitu juga dengan lingkaran x_c berada di bawah *decision boundary*, terlihat pada persamaan (8):

$$w \cdot x_s + b = k''$$

dimana $k'' < 0$. Jika d labeli semua kotak dengan sebagai kelas +1 dan lingkaran sebagai kelas -1, maka yang dilakukan adalah memprediksi label y untuk setiap label y untuk setiap tes contoh z dalam persamaan (9):

$$y = \begin{cases} 1, & \text{if } w \cdot z + b > 0; \\ -1, & \text{if } w \cdot z + b < 0. \end{cases}$$

HASIL DAN PEMBAHASAN

Proses evaluasi pada penelitian ini dilakukan untuk mengetahui kinerja *classifier* terbaik dalam melakukan klasifikasi data ekg multivaraite time series dari fitur yang diperoleh dari ekstraksi ciri dengan metode statistik. Perbandingan kinerja antara *classifer* dilakukan dengan *cross validation* dari *classifer* yang dihasilkan dalam penelitian ini dengan *classifier* kNN (*Nearest Neighbor*) dan Naïve Bayes dari aplikasi Rapidminer 5.0. *Classifier* kNN sebagai pembanding pada proses evaluasi akan menggunakan nilai k yang berbeda-beda. Perbandingan kinerja antar *classifier* ini bertujuan untuk mengetahui seberapa baik kinerja *classifier* dalam penelitian ini dibandingkan dengan *classifier* lain dari aplikasi *data mining* yang umum digunakan.

Kinerja Classifier Support Vector Machine (SMV)

Data yang digunakan adalah 80 data uji untuk mengetahui performa *classifer* SVM. Parameter keberhasilan dibutuhkan berupa hasil akhir nilai klasifikasi yaitu akurasi, sensitivitas, dan spesifikasi. Nilai *confusion matrix* dan tabel klasifikasi tiap metode dapat dilihat pada Tabel 1. Pada Tabel 1

merupakan hasil data uji dengan dilatih menggunakan kernel polynomial.

Tabel 1. *Confusion Matrix* metode SVM dengan kernel Polynomial

		Kelas Hasil Prediksi	
		Positif	Negatif
Kelas	Positif	20	8
Asli	Negatif	7	45

Tabel 1 yaitu *confusion matrix* metode SVM dengan kernel polynomial memiliki arti bahwa:

TP	= 20
TN	= 45
FP	= 7
FN	= 8

Matriks kinerja yang digunakan adalah sensitivitas dan spesifisitas. Sensitivitas yaitu mengukur proporsi asli yang dikenali ataupun diklasifikasi secara benar sebagai positif (misalnya persentase detak jantung abnormal yang diidentifikasi abnormal). Spesifisitas yaitu mengukur proporsi negatif asli yang dikenali secara benar sebagai negatif (misalnya persentase detak jantung tidak abnormal yang diidentifikasi tidak abnormal yang diharapkan normal).

Hasil *confusion matrix* menunjukkan bahwa label 1 terklasifikasi sebagai *true positive* menyatakan sebagai kondisi detak abnormal yang positif terdeteksi sebagai detak abnormal sebanyak 20. *False positive* merupakan suatu kondisi detak normal yang dideteksi sebagai kondisi detak abnormal ada 7. Label 2 pada *confusion matrix* diklasifikasikan sebagai *false negative* yaitu suatu kondisi detak abnormal yang dideteksi sebagai detak normal ada 8 sedangkan *true negative*, yaitu kondisi detak normal dan positif dideteksi sebagai detak normal ada sebanyak 45. Pada

Tabel 2 menunjukkan kinerja data latih *classifier* SVM.

Tabel 2. Kinerja data latih *Classifier Support Vector Machine* (SMV)

Metode SVM	TP	TN	FP	FN	Akurasi	Sensitivitas	Spesifisitas
Kernel polynomial	20	45	7	8	81.25%	71.4%	86.5%
Kernel RBF	26	50	1	3	95%	89.65%	98.03%
Kernel Linear	6	11	2	1	86.25%	76.66%	92%

Kinerja *classifier support vector machine* (SVM) menggunakan kernel polynomial menghasilkan akurasi 81.25%, sensitifitas 71.4%, spesifisitas 86.5% hasil yang didapat kernel RBF 95% sedangkan kernel *linear* 86.25%.

Data yang akan dilatih menggunakan data uji yaitu 20 data untuk mengetahui performa *classifier* SVM pada penelitian ini. Pembangunan model klasifikasi menggunakan SVMStruct, menggunakan fungsi kernel polynomial. Parameter keberhasilan dibutuhkan yaitu hasil akhir nilai klasifikasi yaitu klasifikasi dan akurasi dari *correct rate*. Nilai *confusion matrix* diambil dari *counting matrix* dapat dilihat pada Tabel 3.

Tabel 3. *Confusion Matrix* metode SVM dengan kernel Polynomial

		Kelas Hasil Prediksi	
		Positif	Negatif
Kelas Asli	Positif	2	5
	Negatif	4	9

Tabel 3 yaitu *confusion matrix* metode SVM dengan kernel polynomial memiliki arti bahwa:

TP = 2

TN = 9

FP = 4

FN = 5

Hasil *confusion matrix* menunjukkan bahwa label 1

terklasifikasi sebagai *true positive* menyatakan sebagai kondisi detak abnormal yang positif terdeteksi sebagai detak abnormal sebanyak 2. *False positive* merupakan suatu kondisi detak normal yang dideteksi sebagai kondisi detak abnormal ada 4. Label 2 pada *confusion matrix* diklasifikasikan sebagai *false negative* yaitu suatu kondisi detak abnormal yang dideteksi sebagai detak normal ada 4 sedangkan *true negative*, yaitu kondisi detak normal dan positif dideteksi sebagai detak normal ada sebanyak 5.

Kinerja *classifier support vector machine* (SVM) menggunakan kernel polynomial menghasilkan akurasi 55%, sensitivitas 28.57%, spesifisitas 69.23%. Hasil akurasi yang tertinggi adalah kernel RBF yaitu 100%, sedangkan pada kernel *linear* adalah 85%.

Perbandingan Kinerja antara Classifier Hasil Penelitian dengan Classifier Lainnya Data latih

Perbandingan kinerja dilakukan antara *classifier* yang dibangun dalam penelitian ini, yaitu *support vector machine* (SVM) dengan *classifier* lain yang umum digunakan dalam permasalahan klasifikasi. *Classifier* pembandingan adalah *k-Nearest Neighbour* (kNN) dan Naïve Bayes. Model klasifikasi dari *classifier* kNN menggunakan nilai $k=1$, $k=3$, dan $k=5$,

serta naïve bayes yang digunakan dalam evaluasi ini dibangun dengan memanfaatkan aplikasi Rapidminer 5.0. Data yang digunakan adalah data latih yaitu 80 data. Pada *classifier* kNN dengan menggunakan $k=1$ menghasilkan akurasi 66%, penggunaan $k=3$ menghasilkan akurasi 77%, dan

penggunaan $k=5$ menghasilkan akurasi 80%. Hasil akurasi *classifier* naïve bayes diperoleh 86%, Pada Gambar 7 menunjukkan grafik perbandingan akurasi antara *classifier*, akurasi yang terbaik adalah metode SVM dengan kernel RBF.



Gambar 7. Grafik perbandingan antara *classifier*

Data uji

Perbandingan kinerja dilakukan antara *classifier* yang dibangun dalam penelitian ini, yaitu *support vector machine* (SVM) dengan *classifier* lain

yaitu kNN dengan $k=1$, $k=3$, $k=5$, dan Naïve Bayes menggunakan Rapidminer 5.0. Pada Gambar 8 merupakan perbandingan antara *classifier*.



Gambar 8. Grafik perbandingan antara *classifier*

Data yang digunakan adalah data uji yaitu 20 data. Pada *classifier* kNN dengan menggunakan $k=1$ menghasilkan akurasi 70%, penggunaan $k=3$ menghasilkan akurasi 85%, dan penggunaan $k=5$ menghasilkan akurasi 85%. Hasil akurasi *classifier* naïve bayes diperoleh 75%. Pada Gambar 8 menunjukkan grafik perbandingan akurasi antara *classifier*, akurasi yang terbaik adalah metode SVM dengan kernel RBF dan yang rendah adalah *classifier* SVM dengan kernel polynomial.

KESIMPULAN

Telah dibangun sistem klasifikasi data *ekg multivariate time series* yang fiturnya diperoleh dari ekstraksi ciri menggunakan metode statistik dan klasifikasinya menggunakan SVM dengan penggunaan aplikasi Matlab R2011a.

Ekstraksi ciri menggunakan metode statistik dan SVM menggunakan kernel polynomial dalam mengklasifikasikan detak jantung normal dan abnormal menggunakan data latih yaitu 80 data dengan $k=3$ *cross validation*, menghasilkan akurasi 81.25%, hasil lebih rendah dibandingkan dengan kernel RBF dan kernel *linear*. Data uji yaitu 20 data dengan menggunakan $k=3$ *cross validation* memperoleh hasil akurasi 55%.

DAFTAR PUSTAKA

- Aziz, R., C.K. Verma, dan Srivastava, N. 2016. A Fuzzy Based Feature Selection from Independent Component Subspace for Machine Learning Classification of Microarray Data. *Elsevier:Gnomics Data*, 4-15.
- Banerjee, S. Dan M. Mitra. 2014. A Cross Wavelet Transform Based Approach For ECG Feature Extraction and Classification Without Denosing, *IEEE-International Conference on Control, Instrumentation, Energy & Communication(CIEC14)*., 31 Januari-2 Februari 2014.
- Castillo, O., P. Melin, E. Ramírez. dan J. Soria. 2012. Hybrid intelligent system for cardiac arrhythmia classification with Fuzzy K-Nearest Neighbors and neural networks combined with a fuzzy system. *Expert Systems with Applications*, 39(3) : 2947–2955.
- Kristomo, D. 2014. Klasifikasi Suara Jantung Menggunakan Jaringan Neural dengan Ciri Statistis dan Spektral. *Tesis*. Jurusan Teknik Elektro dan Teknologi Informasi FT UGM. Yogyakarta.
- Kutlu, Y. dan D. Kuntalp. 2012. Feature Extraction for ECG Heartbeats Using Higher Order Statistics of WPD Coefficients. *Elsevier Computer Methods And Programs in Biomedicine* : 257-267.
- Yu-Feng, L., C. H-H, Tseng, V.S., dan P. Jian. 2015. Reliable Early Classification on Multivariate Time Series with Numerical Categorical Attributes, *Springer*, Switzerland.
- Maharaj, A.E. dan M.A. Alonso. 2014. Discriminant Analysis of Multivariate Time Series Using Wavelets, *Computational Statistics dan Data Analysis*, 70 : 67-87.
- Mehta, S.S. dan N.S. Lingayat. 2008. SVM-Based Algorithm for Recognition of QRS Complexes in Electrocardiogram, *Elsevier Masson SAS:IRBM*. 29 : 310-317.
- Ramirez, J., V. Monasterio, A. Mincholé, M. Llamedo, G. Lenis,

- I. Cygankiewicz, A. Bayés, M. Malik, J.P. Martínez, P.Laguna, dan E. Pueyo. 2015. *Journal of Electrocardiology*. 48 : 551-557.
- Pang-Ning, T., S. Michael, dan K. Vipin. 2006. *Introduction to Data Mining, 1st Ed.*, Addison-Wesley. Boston.
- Thanapatay, D. dan C. Suwansaroj. 2010. ECG beat classification method for ECG printout with Principle Components Analysis and Support Vector Machines., *1(Iceie)* : 72–75.
- Liu, W., W. Zhigang, dan L. Shan. 2016. An Effective Multivariate Time Series Classification Approach Using Echo State Network and Adaptive Differential evolution algorithm, *Elsevier* : 237-249.

