

PENERAPAN METODE *SUPPORT VECTOR MACHINE* (SVM) UNTUK KLASIFIKASI DATA EKSPRESI GEN *MICROARRAY*

Dwi Agustina¹⁾, Egicha Putri²⁾, Fatkhurokhman Fauzi³⁾, Silvie Noor Alawiyah⁴⁾, Rochdi Wasono⁵⁾
^{1,2,4)}Program Studi Statistika FMIPA, Universitas Muhammadiyah Semarang, ^{3,5)}Dosen Program Studi Statistika FMIPA, Universitas Muhammadiyah Semarang,

¹⁾dwiagstina@gmail.com ²⁾egichaptri@gmail.com ³⁾fatkhurokhmanfauzi@gmail.com
⁴⁾silvienooralawiyah@gmail.com ⁵⁾rochdi@unimus.ac.id

Abstrack

Pancreatic cancer is the most common malignancy of the pancreas, where the cells in the pancreas grow and multiply abnormally and the condition spreads rapidly to nearby organs and is difficult to detect because it does not cause immediate symptoms. Pancreatic cancer is currently the fourth leading cause of death in developed countries. As technology advances, a new field of science has now developed, namely bioinformatics. One of the implementations of bioinformatics is the use of computational, mathematical and statistical methods to help solve biological problems through the analysis of gene expression data. A computational technique for large-scale bioinformatics data to classify different types of gene expression samples from microarray data. Microarray technology is useful for identifying cancer genes and finding new biomarkers. In this study, classification analysis was carried out on microarray data resulting from the expression gene of tumor tissue and normal tissue in pancreatic cancer patients using the Support Vector Machine (SVM) method for analyzing data. Based on the analysis of the method, the results show that the best function classification with SVM method is kernel linier function with accuracy, recall, specificity, precision, AUC and error values of 100%, 100%, 100%, 100%, 1 and 0%.

Keywords : *Pancreatic Cancer, Bioinformatics, Microarray, Classification, SVM*

PENDAHULUAN

Bioinformatika adalah suatu ilmu atau metode mengumpulkan, dan menganalisa data biologi yang bersifat kompleks seperti data DNA, RNA, dan protein, menggunakan pendekatan komputasi. Bioinformatika melibatkan bidang ilmu komputer, matematika, dan statistik dalam memproses informasi yang masif dan sulit untuk dianalisa secara manual. Pada pertengahan tahun 1990 bioinformatika berkembang pesat terutama ketika peneliti berhasil mengumpulkan *genome* manusia, sebagian besar disebabkan oleh proyek genom manusia (*human genome project*) dan berkembangnya teknologi sekuen DNA.

Dengan menggunakan teknik komputasi untuk data bioinformatika berskala besar dan analisis data biologi yang rumit. Berbagai macam teknik komputasi, terutama teknik *Machine Learning* yang digunakan, contoh untuk memilih gen atau protein yang menghubungkan sifat yang terkait dan untuk mengklasifikasikan perbedaan tipe dari sampel ekspresi gen dari data *microarrays* (Yang, Yang, Zhou, & Zomaya, 2016).

Kanker merupakan suatu jaringan yang berbahaya, penyakit ini dapat berulang atau bertahan sepanjang masa hidup pasien dan mampu diturunkan ke anak cucu selanjutnya. Sebagian pasien kanker tidak pernah sepenuhnya bebas dari penyakit yang diderita. Pankreas adalah bagian dari sistem pencernaan yang membuat dan mengeluarkan enzim pencernaan kedalam usus, dan juga yang membuat dan

mengeluarkan hormon ke dalam darah untuk mengontrol metabolisme dan penyimpanan ke seluruh tubuh (Longnecker, 2014). Pankreas memberikan masalah kesehatan yang substansial, dengan sekitar 367.000 kasus baru didiagnosis di seluruh dunia pada 2015 dan terkait 359.000 kematian pada tahun yang sama.

Data Mining merupakan kumpulan dari kegiatan yang meliputi pengumpulan dan pemakaian data masa lalu untuk menemukan pola atau hubungan dalam data yang berukuran besar. Klasifikasi merupakan salah satu metode yang dapat menangani *big data*. Terdapat banyak metode yang ada dalam klasifikasi diantaranya adalah *Naïve Bayes Classifier*, *Support Vector Machine*, *Artificial Neural Network*, *Classification Tree*, *K-Nearest Neighbour*, dan lain-lain. Pada penelitian ini metode klasifikasi yang digunakan adalah metode perbandingan *Support Vector Machine* (SVM).

Metode *Support Vector Machine* (SVM) mempunyai kelebihan yaitu kemampuan dalam generalisasi atau mampu mengklasifikasikan data yang tidak termasuk data *training* secara tepat, tingkat generalisasi tidak dipengaruhi oleh dimensi peubah yang diamati dan proses komputasi yang lebih cepat karena hanya melibatkan beberapa observasi saja dalam pembentukan fungsi keputusannya (Siregar, 2017). Metode klasifikasi SVM dapat diaplikasikan diberbagai bidang, salah satunya adalah bidang kesehatan. Salah satu aplikasi dalam bidang kesehatan adalah dalam hal prakiraan ekspresi gen *microarray*.

Berdasarkan uraian latar belakang diatas, maka peneliti melakukan analisis dengan judul Penerapan Metode *Support Vector Machine* (SVM) Untuk Klasifikasi Data Ekspresi Gen *Microarray* yang mana harapannya dari hasil penelitian ini dapat diketahui pengklasifikasian data ekspresi gen jaringan tumor dan normal pada kanker pankreas.

Bioinformatika

Bioinformatika adalah ilmu yang mempelajari antar cabang ilmu pengetahuan yang menghubungkan antara ilmu komputer dan ilmu biologi. Menurut (Luscombe, Greenbaum, & Gerstein, 2001), bioinformatika melibatkan diri dengan menggunakan komputer untuk penyimpanan, pencarian keterangan, manipulasi, dan distribusi terkait informasi biologi macromolekul seperti DNA, RNA, dan Protein. Salah satu area bioinformatika adalah analisis ekspresi gen, yaitu sejumlah gen dapat ditentukan dengan mengukur levelnya dengan menggunakan berbagai teknik seperti *microarray* (Raza, 2012).

Microarray

Microarray adalah suatu chip atau slide mikroskop yang berisi rangkaian sampel jaringan, protein, RNA dan DNA. Jumlah sampel yang terdapat didalam *microarray* dapat berjumlah ratusan ribu. *Microarray* memanfaatkan sumber daya sekuen yang diciptakan oleh proyek genom dan upaya sekuensing lainnya yang memungkinkan untuk membandingkan ekspresi gen antara sel yang normal dan sel yang berpenyakit. *Microarray* mempunyai dua perbedaan yang mengarah pada jumlah gen yang dimasukkan pada setiap chip (Vicente, 2009):

1. One-channel *Microarray* adalah *microarray* yang berisi urutan gen dari suatu sampel untuk diproses hibridisasi, contoh dari jenis ini adalah Affymetrics GeneChip. Jenis ini membutuhkan dua chip untuk perbandingan sampel dan dilakukan dengan cara komputasi.
2. Two-channel *Microarray* adalah *microarray* yang berisi urutan gen dari dua sampel yang berbeda untuk diproses hibridisasi kompetitif. Jenis ini mempunyai warna neon yang berbeda yaitu hijau

atau merah. Untuk menampilkan sebuah sampel control vs sampel percobaan penyakit hanya membutuhkan satu chip.

Data Mining

Data mining atau sering disebut *knowledge discovery in database* (KDD) merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam suatu set data yang berukuran besar. Keluaran atau output dari data mining dapat digunakan sebagai perbaikan dalam pengambilan keputusan dimasa depan (santosa, 2007). Tahapan dalam proses *knowledge discovery* terdiri dari langkah-langkah sebagai berikut (Han & Kamber, 2006) yaitu : pembersihan data, integrasi data, seleksi data, transformasi data, proses mining, evaluasi pola dan presentasi pengetahuan.

Klasifikasi

Klasifikasi merupakan suatu proses untuk menentukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Pada Data mining kasus klasifikasi ada 2 yaitu bersifat linear dan tidak linear. Metode analisis data yang digunakan untuk membentuk model mendeskripsikan kelas data penting atau model mendeskripsikan trend data adalah klasifikasi. Fungsinya untuk memprediksi kelas data yang bersifat *categorical*, sedangkan prediksi untuk memodelkan fungsi yang bernilai *continuous* (Furey, 2000).

Data Training dan Data Testing

Data training digunakan oleh algoritma klasifikasi (misalnya *Decision Tree*, *Bayesian*, *Neural Network*, *SVM*) untuk membentuk sebuah model klasifikasi. Model ini merupakan representasi pengetahuan yang akan digunakan untuk mengukur sejauh mana klasifikasi berhasil melakukan klasifikasi dengan benar. Proporsi antara data training dan data testing tidak mengikat tetapi agar variasi dalam model tidak terlalu besar maka disarankan data training lebih besar dibandingkan data testing.

Pre-Processing

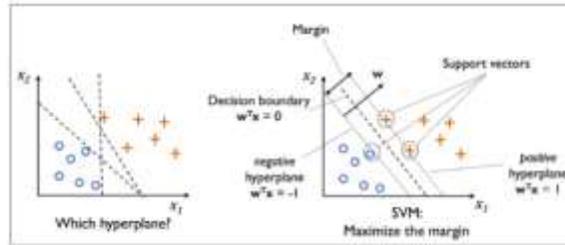
Preprocessing adalah proses yang dilakukan untuk membuang efek non biologis pada data yang berguna untuk memberikan hasil yang lebih baik (Serin, 2011). Preprocessing terbagi menjadi tiga tahapan yaitu, *background correction*, *normalization* dan *summerization*.

Filtering

Filtering adalah salah satu operasi dasar untuk menghapus sebagian data yang tidak diikutsertakan dalam analisis atau data yang tidak valid. Filtering dapat meningkatkan kualitas model, juga dapat membuat proses pemodelan lebih efisien. Fungsi *nsFilter* (*non specific filter*) menyediakan *one-stop shop* (serba ada) untuk berbagai pilihan filtering (penghapusan) fitur dari *expression set*.

Support Vector Machine (SVM)

SVM merupakan salah satu metode klasifikasi dalam data *mining*. SVM juga dapat melakukan prediksi baik pada klasifikasi maupun regresi (Santosa, 2007 dalam Pusphita, 2014). Cara kerja SVM pada masalah *non-linear* adalah dengan memasukkan konsep kernel pada ruang berdimensi tinggi. Pada ruang yang berdimensi ini, nantinya akan dicari pemisah atau yang sering disebut *hyperplane*.



Gambar 1. contoh Hyperplane

Fungsi klasifikasi pada SVM didefinisikan dengan persamaan: $f(x) = \text{sign}(w \cdot x + b)$

Apabila nilai dari $w \cdot x + b > 0$ maka diklasifikasikan kedalam kelas +1 dan jika $w \cdot x + b < 0$ data diklasifikasikan kedalam kelas -1.

SVM mempunyai 4 kernel yang dapat digunakan untuk menyelesaikan permasalahan linear dan non-linear diantaranya adalah:

1. Kernel Linear

$$K(x, x_k) = x_k^T x$$

2. Kernel Polynomial

$$K(x, x_k) = (x_k^T x + 1)^d$$

3. Kernel Gaussian Radial Basis Function (RBF)

$$K(x, x_k) = \exp\{-\|x - x_k\|_2^2 / \sigma^2\}$$

4. Sigmoid

$$K(x, x_k) = \tanh[\kappa x_k^T x + \theta]$$

Nilai akurasi dari model yang dihasilkan dengan menggunakan SVM sangat tergantung pada fungsi kernel dan nilai parameter yang digunakan. Parameter yang digunakan pada algoritma SVM adalah parameter Cost (C) dan Gamma (γ). Semakin besar nilai C akan menghasilkan penalty yang besar pula terhadap klasifikasinya. Pada fungsi kernel RBF parameter gamma (γ) digunakan untuk memtransformasikan data train ke ruang fitur yang kemudian dioptimasi menggunakan metode Lagrange Multipliers sehingga menghasilkan nilai α yang digunakan untuk menentukan support vector dan memperkirakan koefisien w (bobot) ataupun b (bias) pada model klasifikasi.

METODE PENELITIAN

Data Penelitian

Data yang peneliti gunakan adalah data sekunder yang peneliti dapatkan dari website resmi *National Center of Biotechnology Information (NCBI) Gene Expression Omnibus* dengan kode akses GSE16515_RAW yang berjudul “*Expression data from Mayo Clinic Pancreatic Tumor and Normal samples*” dengan url [/https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16515](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16515).

Metode Analisis Data

Metode analisis data yang digunakan pada penelitian ini adalah klasifikasi menggunakan metode klasifikasi yaitu *Support Vector Machine* (SVM). Analisis ini digunakan untuk pengklasifikasian kanker pankreas dengan dua kelas yaitu jaringan tumor dan normal. Perangkat lunak yang digunakan pada penelitian ini adalah *software* R studio version 4.0.0.

Variabel Penelitian dan Definisi Operasional Variabel

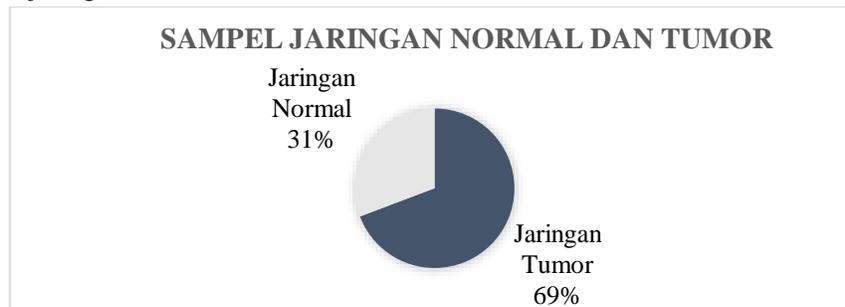
Adapun macam-macam variabel yang digunakan pada penelitian ini adalah sebagai berikut:

Nama Variabel	Definisi Operasional Variabel
Gen	Gen yang digunakan pada penelitian ini adalah hasil dari ekspresi gen jaringan tumor dan jaringan normal pada manusia penderita kanker pankreas.
Jaringan	Variabel yang menunjukkan sekumpulan sel yang menyusun setiap tubuh manusia. variabel ini memiliki 2 jaringan yang berbeda pada data yaitu jaringan tumor dan jaringan normal.

PEMBAHASAN

Analisis Deskriptif

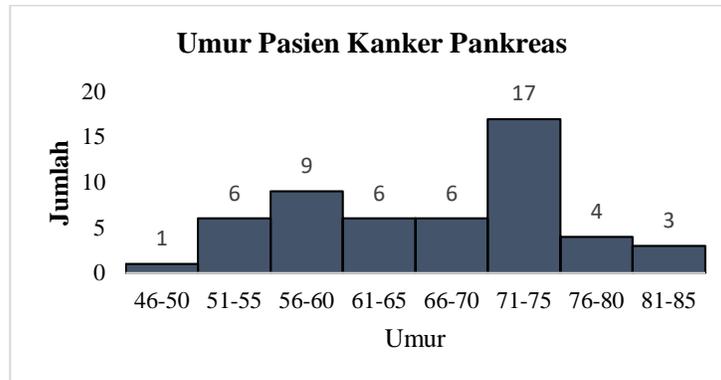
Data GSE16515 atau data set Kanker Pankreas merupakan kumpulan data ekspresi gen sebanyak 54675 nama gen dan 52 sampel yang menderita kanker pankreas yang dibedakan berdasarkan sampel jaringan tumor dan sampel jaringan normal.



Gambar 2. Pieplot Jaringan Tumor dan Normal

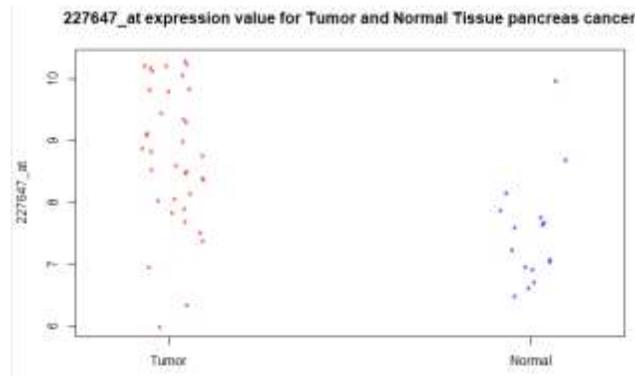
Berdasarkan data jenis kelamin pada pasien kanker pankreas lebih banyak dialami oleh pasien berjenis kelamin laki-laki, dimana jumlah pasien berjenis kelamin perempuan sebanyak 18 pasien dan untuk pasien berjenis kelamin laki-laki sebanyak 34 pasien. Dengan pasien normal sebanyak 16 pasien yang terdiri dari 4 pasien perempuan dan 12 pasien laki-laki. Pasien tumor sebanyak 36 pasien yang terdiri dari 14 pasien perempuan dan 22 pasien laki-laki.

Didapatkan pada **Gambar 3.** dibawah menunjukkan bahwa seseorang terkena penyakit kanker pancreas paling banyak di umur 71-75 tahun sebanyak 17 pasien. Pasien termuda kanker pankreas pada umur 49 tahun dan tertua pada umur 84 tahun. Dari hasil tersebut dapat diambil kesimpulan bahwa seseorang yang didiagnosa terkena kanker payudara dapat terjadi pada usia dewasa sampai manula.



Gambar 3. Barplot Umur Pasien

Pada Gambar 4. menunjukkan gambaran sebaran nilai ekspresi pada gen '227647_at' yang terletak di baris pertama pada kelas tumor dan normal, sumbu vertikal adalah ukuran nilai ekspresi gen dan sumbu horizontal adalah kelas kanker pankreas yang mempunyai jaringan tumor dan jaringan normal.



Gambar 4. Nilai ekspresi gen jaringan tumor dan normal

Pengolahan Data Bioinformatik

```
[1] "title"
[4] "submission_date"
[7] "channel_count"
[10] "characteristics_ch1"
[13] "molecule_ch1"
[16] "label_protocol_ch1"
[19] "scan_protocol"
[22] "platform_id"
[25] "contact_phone"
[28] "contact_institute"
[31] "contact_state"
[34] "contact_web_link"
[37] "relation"
[40] "sex:ch1"

"geo_accession"
"last_update_date"
"source_name_ch1"
"characteristics_ch1.1"
"extract_protocol_ch1"
"taxid_ch1"
"description"
"contact_name"
"contact_fax"
"contact_address"
"contact_zip/postal_code"
"supplementary_file"
"relation.1"
"tissue:ch1"

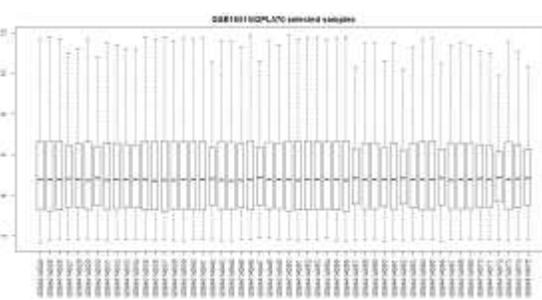
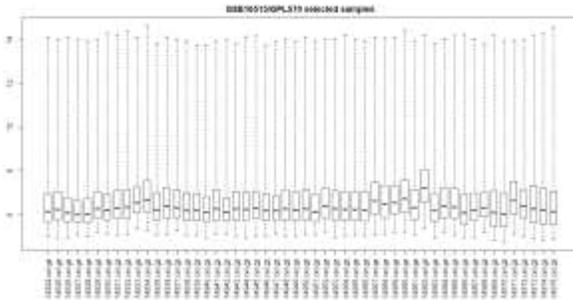
"status"
"type"
"organism_ch1"
"characteristics_ch1.2"
"label_ch1"
"hyb_protocol"
"data_processing"
"contact_email"
"contact_department"
"contact_city"
"contact_country"
"data_row_count"
"age:ch1"
```

Gambar 5. Pheno Data

Dari gambar diatas diperoleh informasi tentang pasien yang dijadikan sampel penelitian. Pada gambar tersebut diperoleh informasi bahwa setiap pasien diberikan kode GSM yang mencakup geo_accession, status, jenis kelamin, umur, sampai jaringan pasien. Secara keseluruhan, pheno data tersebut mempunyai sampel sebanyak 52 dan 41 variabel.

Pre-Pocessing

Fungsi preprocessing digunakan untuk menghilangkan nilai non biologis dan *noise* pada data. Untuk hasil dari tahapan *preprocessing* ini dapat dilihat dalam bentuk *boxplot*.



Gambar 6. *Boxplot* sebelum preprocessing

Gambar 7. *Boxplot* setelah preprocessing

Pada *boxplot* setelah dilakukan preprocessing terlihat data lebih baik daripada sebelum dilakukan preprocessing. Tahapan preprocessing pada data ekspresi gen digunakan untuk mengahapus nilai non biologis, sehingga yang terdapat pada *boxplot* setelah dilakukan preprocessing hanya data yang bersifat biologis berupa nilai ekspresi gen yang menyebabkan *boxplot* tersebut mempunyai nilai rata-rata sama.

Filtering

Filtering pada data *microarray* merupakan proses dalam pemilihan subset dari *probe* yang ada untuk pengecualian atau penyertaan dalam analisis. Tahapan ini berguna untuk penentuan atau pemilihan gen yang siap di analisis.

NsFilter	
Sample	Gen
52	2019

Tabel 1. Hasil filtering NsFilter

Gen	Sampel
417	52

Tabel 2. Hasil Feature Selection

Setelah mendapatkan pemfilteran data menggunakan nsFilter tahapan selanjutnya adalah *feature selection*. Pada tahapan multtest klasifikasi data pada penelitian diasumsikan berdistribusi normal, yang dilihat pada *qqnorm*. Hasil filtering pada **Tabel 2.** ini memberikan data dengan dimensi 417x52. Data inilah yang akan digunakan kedalam klasifikasi SVM.

Pembagian Data Training dan Testing

Data yang telah siap untuk di analisis dengan dimensi 417x52 dan selanjutnya data tersebut akan dibagi menjadi data *training* (data latih) dan data *testing* (data uji) untuk melakukan klasifikasi. Pembagian rasio data *training* dan data *testing* yaitu sebesar 80% : 20% dari total data yang siap dianalisis untuk masing-masing kelas.

Pembagian Data	
Data Training	Data Testing
41	11

Tabel 3. Pembagian data training dan data testing

Kernel Linier

Penggunaan metode SVM dengan kernel *linier* berdasarkan penyakit kanker pankreas pada jaringan tumor dan normal diperoleh hasil data *training* bahwa dari total data sebanyak 41, didapatkan bahwa terdapat 28 sampel yang terklasifikasi ke dalam kelas Jaringan Tumor dan terdapat 1 nilai kesalahan serta ada 12 sampel yang terklasifikasi kedalam kelas Jaringan Normal dan tidak ada nilai kesalahan. Proses *testing* dari sampel sebanyak 11 diperoleh kelas Jaringan Tumor sebanyak 7 sampel dengan tidak ada nilai kesalahan dan Jaringan Normal sebanyak 4 sampel dengan tidak ada nilai kesalahan pula.

$$\text{Error} = \frac{\text{Jumlah Prediksi Salah}}{\text{Jumlah Prediksi Keseluruhan}} = \frac{4}{11} = 0,364 = 36,4\%$$

Dari perhitungan nilai error diatas diperoleh tingkat kesalahan sebesar 36,4%, yang artinya bahwa nilai kesalahan pada metode SVM model kernel polynomial yang diperoleh dapat dikatakan cukup tinggi.

Kornel Polynomial

Penggunaan metode SVM dengan kernel *polynomial* berdasarkan penyakit kanker pankreas pada jaringan tumor dan normal diperoleh hasil data *training* bahwa dari total data sebanyak 41, didapatkan bahwa terdapat 29 sampel yang terklasifikasi ke dalam kelas Jaringan Tumor dan tidak ada nilai kesalahan serta ada 12 sampel yang terklasifikasi kedalam kelas Jaringan Normal dan tidak ada nilai kesalahan pula. Proses *testing* dari sampel sebanyak 11 diperoleh kelas Jaringan Tumor sebanyak 4 sampel dengan 1 nilai kesalahan dan Jaringan Normal sebanyak 3 dengan 3 nilai kesalahan pula.

$$\text{Error} = \frac{\text{Jumlah Prediksi Salah}}{\text{Jumlah Prediksi Keseluruhan}} = \frac{4}{11} = 0,364 = 36,4\%$$

Dari perhitungan nilai error diatas diperoleh tingkat kesalahan sebesar 36,4%, yang artinya bahwa nilai kesalahan pada metode SVM model kernel polynomial yang diperoleh dapat dikatakan cukup tinggi.

Kernel RBF

Penggunaan metode SVM dengan kernel *polynomial* berdasarkan penyakit kanker pankreas pada jaringan tumor dan normal diperoleh hasil data *training* bahwa dari total data sebanyak 41, didapatkan bahwa terdapat 29 sampel yang terklasifikasi ke dalam kelas Jaringan Tumor dan tidak ada nilai kesalahan serta ada tidak ada sampel yang terklasifikasi kedalam kelas Jaringan Normal dan terdapat 12 nilai kesalahan. Proses *testing* dari sampel sebanyak 11 diperoleh kelas Jaringan Tumor sebanyak 7 sampel dengan tidak ada nilai kesalahan dan tidak terdapat sampel pada Jaringan Normal dengan kesalahan 4.

$$\text{Error} = \frac{\text{Jumlah Prediksi Salah}}{\text{Jumlah Prediksi Keseluruhan}} = \frac{4}{11} = 0,364 = 36,4\%$$

Dari perhitungan nilai error diatas diperoleh tingkat kesalahan sebesar 36,4%, yang artinya bahwa nilai kesalahan pada metode SVM model kernel RBF yang diperoleh dapat dikatakan cukup tinggi.

Perbandingan Fungsi Kernel dan SVM

Perbandingan antara ketiga fungsi kernel yang telah dianalisis untuk menentukan fungsi terbaik pada data Kanker Pankreas. Penentuan fungsi terbaik dapat dilihat dari nilai *accuracy* dan *error* yang telah diperoleh dari data *testing*.

Fungsi	Kernel Linier	Kernel Polynomial	Kernel RBF
Accuracy	100%	63,6%	100%
Recall	100%	57,14%	63,6%

Precision	100%	80%	63,6%
Spesificity	100%	75%	0%
FPR	0	0,25	1
AUC	1	0,66	0,5
Error	0%	36,4%	36,4%

Tabel 4. Perbandingan Tiga Fungsi Kernel SVM

Dimana semakin kecil nilai *error* maka tingkat kesalahan dalam klasifikasi semakin baik. Berdasarkan hasil perbandingan dari ketiga fungsi kernel SVM tersebut, dapat disimpulkan bahwa fungsi terbaik dalam analisis klasifikasi dengan metode SVM adalah fungsi kernel *linier*.

Model Klasifikasi

Tabel 5. menunjukkan bahwa setiap *probe* yang ada menunjukkan peran dan *probe* yang paling berperan adalah X.205319 at, *probe* tersebut merupakan gen paling aktif yang menyebabkan kanker pankreas. Berikut merupakan pemaparan 10 bobot gen tertinggi klasifikasi metode SVM dengan fungsi kernel *linier*:

Probe Id	W
X.205319_at.	0,0435261
X.204351_at.	0,0396899
X.231646_at.	0,0348041
X.205009_at.	0,0335079
X.214385_s_at.	0,0280155
X.202504_at.	0,0270339
X.217546_at.	0,0263201
X.204914_s_at.	0,0261613
X.217109_at.	0,0228557
X.205064_at.	0,0225288

Tabel 5. Bobot Gen Tertinggi

KESIMPULAN

Berdasarkan hasil pembahasan yang telah dipaparkan, diperoleh beberapa kesimpulan. Kesimpulan yang diperoleh adalah sebagai berikut :

1. Pengolahan data ekspresi gen jaringan tumor dan normal penderita kanker pancreas pada data *microarray* adalah sebagai berikut :
Data GSE16515 merupakan kumpulan sampel data ekspresi dari mayo clinic sampel tumor dan sampel normal pada pasien kanker pancreas. dari data gen series GSE16515 didapatkan melalui matrix *AffyData*, dimana perintah tersebut hanya digunakan untuk membaca data yang berbentuk gen. Perintah *ReadAffy* adalah perintah untuk membaca definisi isi dari data, seperti *annotation*, jumlah sampel ekspresi gen jaringan tumor dan jaringan normal.
2. Hasil gambaran data ekspresi gen jaringan tumor dan normal penderita kanker pancreas pada data *microarray* adalah sebagai berikut :

- a. Berdasarkan data ekspresi gen jaringan tumor dan jaringan normal pada penderita kanker pankreas menunjukkan bahwa terdapat sebanyak 36 sampel mempunyai tumor pada kanker pankreas dan 16 sampel mempunyai jaringan normal pada kanker pankreas.
 - b. Gambaran data ekspresi gen jaringan tumor dan jaringan normal penderita kanker pankreas menjelaskan jika jaringan tumor lebih dominan menyebabkan kanker pankreas, dan penderita kanker pankreas lebih dominan berjenis kelamin laki-laki daripada perempuan dengan penderita terbanyak pada rentang umur 70 tahun.
3. Pengolahan data ekspresi gen jaringan tumor dan jaringan normal penderita kanker pankreas dengan menggunakan metode klasifikasi *Support Vector Machine* dengan fungsi kernel *linier*, kernel *polynomial* dan kernel RBF telah berhasil dilakukan. Hasil evaluasi klasifikasi dari fungsi kernel metode *Support Vector Machine* dilihat dari nilai *accuracy*, *recall*, *specificity*, *precision*, AUC dan *Error* adalah sebagai berikut :
- a. Hasil analisis klasifikasi menggunakan metode SVM fungsi kernel *linier* dapat memberikan nilai *accuracy*, *recall*, *precision*, *specificity*, AUC dan *error* berturut-turut adalah 100%, 100%, 100%, 100%, 1 dan 0%.
 - b. Hasil analisis klasifikasi menggunakan metode SVM fungsi kernel *polynomial* dapat memberikan nilai *accuracy*, *recall*, *precision*, *specificity*, AUC dan *error* berturut-turut adalah 63,6%, 57,14%, 80%, 75%, 0,66 dan 36,4%.
 - c. Hasil analisis klasifikasi menggunakan metode SVM fungsi kerne RBF dapat memberikan nilai *accuracy*, *recall*, *precision*, *specificity*, AUC dan *error* berturut-turut adalah 100%, 63,6%, 63,6%, 0%, 0,5 dan 36,4%.
 - d. Berdasarkan hasil klasifikasi di atas, diperoleh metode klasifikasi SVM dengan fungsi kernel terbaik dalam melakukan klasifikasi data ekspresi gen jaringan tumor dan jaringan normal pada penderita kanker pankreas adalah klasifikasi metode *Support Vector Machine* dengan fungsi kernel *linier*.

REFERENSI

- i. Ahmad, A. 2017. *Mengenal Artificial Intelligence, Machine learning, Neural Network, dan Deep Learning*. Yayasan Cahaya Islam Jurnal Teknologi Indonesia.
- ii. Ardeshir, B. 2002. *Science, Medicine, and the Future : Bioinformatics*. BMJ, 1018-1022.
- iii. Bramer, M. 2013. *Principle of Data Mining Second Edition*. London: Springer.
- iv. Brazma, A., et al. 2001. *A Quick Introduction To Elements Of Biology-Cell, Molecules, Genes, Functional Genomics, Microarrays*. EMBL-EBI.
- v. CMScript. 2017. *Pancreatic Cancer. Council for Medical Schemes*.
- vi. Diani, Rima., Wisesty, U, N., Aditsania, Annisa., 2017. Analisis Pengaruh Kernel *Support Vector Machine* (SVM) pada Klasifikasi Data *Microarray* untuk Deteksi Kanker. *Indonesian Journal Of Computing*, vol. 2, hal. 109-118.
- vii. Edi., 2017. Bioinformatika: Komputer, Statistika, Matematika, Biologi. *Jurnal TIMES Volume VI* No 1, hal. 23-25.

- viii. Hofmann, M. 2006. *Support Vector Machines — Kernels and the Kernel Trick*. http://www.cogsys.wiai.unibamberg.de/teaching/ss06/hs_svm/slides/SVM_Seminarbericht_Hofmann.pdf. [Diakses 20 Maret 2019].