

PERBANDINGAN ANALISIS KLASIFIKASI ANTARA DECISION TREE DAN SUPPORT VECTOR MACHINE MULTICLASS UNTUK PENENTUAN JURUSAN PADA SISWA SMA

Rizky Ade Putranto¹, Triastuti Wuryandari², Sudarno³

¹Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

^{2,3}Staff Pengajar Jurusan Statstika FSM Universitas Diponegoro

ABSTRACT

Data mining is a process that employs one or more of Machine Learning techniques to analyze and extract knowledge automatically. Analysis of data mining is to determine the classification of a new data record into one of several categories that have been defined previously, also known as Supervised Learning. Classification Decision Tree is one of the well-known technique in data mining and is one of the popular methods in the decision making process of a case in which the method is obtained entropy criteria, information gain and gain ratio. Classification Support Vector Machine Multiclass (SVMM) is known as the most advanced machine learning techniques to handle multi-class case where the output of the data set has more than two classes or categories. This final project aims to compare the level of accuracy and error rate of Decision Tree classification and prediction majors SVMM for high school students at SMAN 1 Jepara. The total accuracy of 88,57% and 11,43% error rate for the classification decision tree and the total accuracy of 87,14% and the error rate for the classification SVMM 12,86%.

Keywords : Data Mining, Machine Learning, Supervised Learning, Decision Tree, Support Vector Machine Multiclass

1. PENDAHULUAN

Setiap lembaga mempunyai sistem operasional yang setiap transaksi kegiatan operasinya selalu dicatat dan didokumentasikan. Pendokumentasian setiap transaksi sangat berguna bagi lembaga tersebut untuk segala keperluan. Data-data tersebut tersimpan dalam sebuah basis data berkapasitas besar [6].

Munculnya data mining didasarkan pada kenyataan bahwa jumlah data yang tersimpan dalam basis data semakin besar. Data mining sendiri berisi pencarian trend atau pola tertentu yang diinginkan dalam basis data yang besar untuk membantu pengambilan keputusan di waktu yang akan datang [3]. Data mining berhubungan dengan sub-area statistik yang disebut *Exploratory Data Analysis* (Analisis Data Eksplorasi) yang mempunyai tujuan sama dan bersandar pada ukuran statistik. Data mining berpotensi tinggi jika data yang tepat dikumpulkan dan disimpan dalam sebuah gudang data (*data warehouse*) [3].

Analisis klasifikasi data mining adalah menentukan sebuah record data baru ke salah satu dari beberapa kategori yang telah didefinisikan sebelumnya, disebut juga dengan *supervised learning* [10]. Metode-metode yang telah dikembangkan oleh periset untuk menyelesaikan kasus klasifikasi, antara lain: Pohon keputusan (*Decision Tree*), *Naïve Bayes*, Jaringan Syaraf Tiruan, Analisis Statistik, Algoritma Genetik, *Rough Sets*, *k-Nearest Neighbour*, Metode Berbasis Aturan, *Memory Based Reasoning*, *Support Vector Machine* [10].

Berdasarkan beberapa metode klasifikasi tersebut, ingin dilakukan perbandingan analisis klasifikasi antara metode *Decision Tree* dan *Support Vector Machine Multiclass* untuk penentuan jurusan pada siswa SMA. Tingkat penguasaan ilmu pengetahuan dan teknologi erat kaitannya dengan kesejahteraan dan perekonomian suatu negara. Masyarakat yang berpendidikan, berwawasan, berbudi, dan terampil dapat membawa bangsanya

menjadi negara yang maju dan disegani bangsa lain. Oleh karena itu, pendidikan sebagai salah satu landasan utama meraih impian tersebut. Generasi berpendidikan turut mempengaruhi angkatan kerja di sebuah negara, tanpa terkecuali Indonesia.

Tujuan penjurusan ini adalah agar minat dan bakat seseorang dapat terarah dengan spesifik pada bidang ilmu tersebut yang selanjutnya dapat melanjutkan ke jenjang perguruan tinggi dan setelah itu memperoleh suatu pekerjaan yang sesuai minat dan bakat seseorang. Akhirnya tercapai suatu tujuan peningkatan penguasaan ilmu pengetahuan dan teknologi yang erat kaitannya dengan kesejahteraan dan perekonomian suatu negara.

2. TINJAUAN PUSTAKA

2.1 Konsep Data Mining

Data mining adalah proses yang melakukan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Pembelajaran data mining berbasis induksi (*induction-based learning*) merupakan proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep yang akan dipelajari [3].

2.2 Operasi Data Mining

1. Prediksi (*prediction driven*)

Untuk menjawab pertanyaan apa dan sesuatu yang bersifat remang-remang atau transparan. Operasi prediksi digunakan untuk validasi hipotesis, querying dan pelaporan, analisis multidimensi, OLAP (*Online Analytic Processing*), serta analisis statistik.

2. Penemuan (*discovery driven*)

Bersifat transparan dan untuk menjawab pertanyaan “mengapa?”, operasi penemuan digunakan untuk analisis data eksplorasi, pemodelan prediktif, segmentasi database, analisis keterkaitan (*link analysis*) dan deteksi deviasi [3].

2.3 Konsep Decision Tree

Klasifikasi *decision tree* merupakan salah satu teknik terkenal dalam data mining dan merupakan salah satu metode yang popular dalam menentukan keputusan suatu kasus. Metode ini tidak memerlukan proses pengelolaan pengetahuan terlebih dahulu dan dapat menyelesaikan kasus-kasus yang memiliki dimensi yang besar[13].

2.4 Algoritma C4.5

Dalam analisis *decision tree* ini sebenarnya jenis algoritma yang bisa digunakan ada beberapa macam, seperti algoritma induksi CART (*Classification and Regression Tree*), algoritma induksi ID3 yang dibuat tahun 1970 sampai awal tahun 1980 oleh J. Ross Quinlan, seorang peneliti di bidang *machine learning* dan yang ketiga adalah algoritma induksi C4.5. Dari ketiga algoritma induksi tersebut peneliti memilih algoritma induksi C4.5. Algoritma C4.5 ini merupakan pengembangan dari algoritma ID3 yang dibuat oleh Quinlan, dan algoritma ini memiliki kelebihan yaitu mudah dimengerti, fleksibel, dan menarik karena dapat divisualkan dalam bentuk gambar *decision tree* [2].

2.5 Memilih Atribut Decision Tree

Atribut *Decision Tree* adalah atribut yang memungkinkan untuk mendapatkan *decision tree* yang paling kecil ukurannya atau atribut yang bisa memisahkan obyek menurut kelasnya. Atribut yang dipilih adalah atribut yang “*purest*” (paling murni). Dalam satu cabang anggotanya berasal dari satu kelas maka cabang ini disebut *pure*. Semakin *pure* suatu cabang maka semakin baik. Ukuran *purity* dinyatakan dengan tingkat *impurity*. Salah satu kriteria *impurity* adalah *information gain* [7].

$$\text{entropy}(S) = \sum_{i=1}^m -p(w_i|S) \cdot \log_2 p(w_i|S)$$

dengan:

S = Himpunan kasus
 m = Banyak kelas data

$p(v_i|S)$ = Proporsi kelas ke- i dalam semua data latih yang diproses di node S

Information Gain

Information gain dari output data atau variabel dependent S yang dikelompokkan berdasarkan atribut J , dinotasikan dengan $gain(S, J)$. Information gain, gain dari atribut J relatif terhadap output data S [7]. Dirumuskan sebagai berikut

$$Gain(S, J) = Entropy(S) - \sum_{i=j}^n p(v_i|S) * E(S_i)$$

dengan:

S = Himpunan kasus
 J = Fitur

n = Banyak kelas dalam *node* (akar)

$p(v_i|S)$ = Proporsi nilai v muncul pada kelas dalam *node* (akar)

$E(S_i)$ = Entropi komposisi nilai v dari kelas ke- j dalam data ke- i node tersebut

Split Information dan Gain Ratio

Untuk menghitung *gain ratio* perlu diketahui *split information*. *Split information* dihitung dengan formula sebagai berikut [7].

$$SplitInformation(S, J) = - \sum_{i=1}^n p(v_i|S) \log_2 p(v_i|S)$$

Selanjutnya *gain ratio* dihitung dengan cara [7].

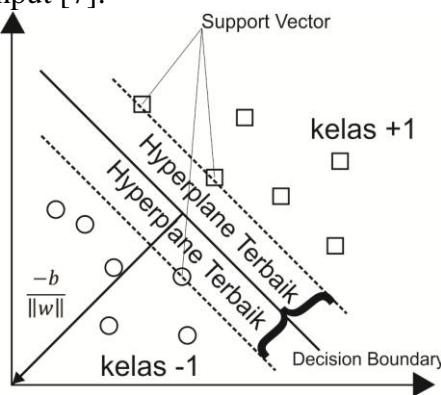
$$Gainratio(S, J) = \frac{Gain(S, J)}{Split Information (S, J)}$$

2.7 Support Vector Machine

Support Vector Machine (SVM) dikenal sebagai teknik pembelajaran mesin (*Machine Learning*) paling mutakhir setelah pembelajaran mesin sebelumnya yang dikenal sebagai *Neural Networks* (NN). Baik SVM maupun NN tersebut telah berhasil digunakan dalam pengenalan pola. Pembelajaran dilakukan dengan menggunakan pasangan data input dan data output berupa sasaran yang diinginkan. Pembelajaran dengan cara ini disebut dengan pembelajaran terarah (*supervised learning*) [5].

2.7.1 Hyperplane

Klasifikasi SVM dijelaskan secara sederhana untuk mendapatkan atau mencari garis terbaik sebagai hyperplane terbaik yang dapat memisahkan dua kelas berbeda antara kelas (+1) dan kelas (-1) pada ruang input [7].



Gambar 1. Hyperplane pada SVM

dengan:

$$\begin{aligned}\text{Hyperplane} &= w \cdot x_i + b = 0 \\ \text{Kelas -1} &= w \cdot x_i + b = -1 \\ \text{Kelas +1} &= w \cdot x_i + b = +1\end{aligned}$$

2.7.2 Klasifikasi *Linear Separable*

Hyperplane klasifikasi linier SVM dinotasikan sebagai berikut [9].

$$f(x) = w^T x_i + b$$

Untuk mencari *hyperplane* terbaik dapat digunakan metode *Quadratic Programming* (QP) yaitu dengan cara meminimalkan [9].

$$\frac{1}{2} \|w\|^2$$

Solusi untuk optimalisasi ini dapat diselesaikan dengan *Lagrange Multiplier* sebagai berikut [6].

$$Lp = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{\ell} \alpha_i y_i (w^T \cdot x_i + b) - 1$$

dengan syarat:

$$\begin{aligned}\alpha_i &= \text{Lagrange multiplier} \text{ berkorespondensi dengan } x_i \\ i &= 1, 2, \dots, \ell\end{aligned}$$

Setelah permasalahan QP ditemukan nilai α_i , maka kelas dari data yang akan diprediksi atau data *testing* dapat ditentukan berdasarkan persamaan berikut [7].

$$f(z) = \left(\sum_{i=1}^{\ell} \alpha_i y_i x_i \cdot z \right) + b = 0$$

dengan:

$$\begin{aligned}\ell &= \text{jumlah data yang menjadi support vector} \\ x_i &= \text{support vector} \\ z &= \text{data uji yang akan diprediksi kelasnya (data testing)} \\ x_i \cdot z &= \text{inner-product antara } x_i \text{ dan } z\end{aligned}$$

2.7.3 Klasifikasi *Linear Non Separable*

Pembatas yang sudah dimodifikasi untuk kasus non-separable [7].

$$y_i [(w^T \cdot x_i) + b] \geq 1 - \xi_i, \quad i = 1, 2, \dots, \ell$$

fungsi *lagrange* sebagai berikut [11].

$$L(w, b, \alpha, \xi, \beta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\ell} \xi_i - \sum_{i=1}^{\ell} \alpha_i (y_i [(w^T \cdot x_i) + b] - 1 + \xi_i) \sum_{i=1}^{\ell} \beta_i \xi_i$$

2.7.4 Klasifikasi *Non Linear*

Algoritma mengenai pemetaan dengan kernel [6].

$$\Phi : x_i^q \rightarrow x_i^r$$

$$x_i \rightarrow \Phi(x_i)$$

dengan:

- Φ = fungsi kernel yang digunakan untuk pemetaan
 - q = set fitur dalam data yang lama
 - r = set fitur yang baru sebagai hasil pemetaan untuk setiap data *training*
 - x_i = data *training*, dimana $x_1, x_2, \dots, x_\ell \in x_i^q$ merupakan fitur yang akan dipetakan ke fitur berdimensi tinggi r
- macam-macam fungsi *kernel* sebagai berikut [6].

1. *Kernel Linear*

$$K(x, y) = x \cdot y$$

2. *Kernel Polynomial*

$$K(x, y) = (x \cdot y + c)^d$$

3. *Kernel Gaussian RBF*

$$K(x, y) = \exp\left(\frac{-\|x - y\|^2}{2 \cdot \sigma^2}\right)$$

2.8 Konsep SVMM (Support Vector Machine Multiclass)

Ada dua pendekatan utama untuk SVMM. Pertama, menemukan dan menggabung beberapa fungsi pemisah persoalan klasifikasi dua kelas untuk menyelesaikan persoalan klasifikasi multi kelas. Yang kedua, secara langsung menggunakan semua data dari semua kelas dalam satu formulasi persoalan optimisasi [4]. Satu lawan semua (*One against all*, OAA) dan Satu lawan satu (*One against one*, OAO) [12].

Dengan metode ini diperlukan untuk menemukan nilai $\frac{k(k-1)}{2}$ fungsi pemisah dimana setiap fungsi dilatih dengan data dari dua kelas. Untuk melatih data dari kelas ke-i dan ke-j, dilakukan dengan menyelesaikan persoalan klasifikasi 2-kelas berikut:

$$\min_{w^{ij}, b^{ij}, \xi^{ij}} \frac{1}{2} (w^{ij})^T w^{ij} + C \sum_t \xi_t^{ij}$$

dengan:

w^{ij} = Margin antara 2 grup data i dan j

b^{ij} = Parameter i dan j yang akan dicari nilainya

ξ^{ij} = Variabel *Slack* i dan j

C = Nilai parameter

2.9 Pengukuran Uji Ketepatan Klasifikasi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua himpunan data dengan benar, tetapi tidak dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi harus diukur tingkat ketepatan klasifikasinya menggunakan matriks konfusi [6].

Tabel 1. Matriks Konfusi untuk Klasifikasi Tiga Kelas

| | | Kelas hasil prediksi (j) | | |
|----------------|-----------|--------------------------|-----------|-----------|
| | | kelas = 1 | kelas = 2 | kelas = 3 |
| Kelas asli (i) | kelas = 1 | f_{11} | f_{12} | f_{13} |
| | kelas = 2 | f_{21} | f_{22} | f_{23} |
| | kelas = 3 | f_{31} | f_{32} | f_{33} |

Untuk menghitung akurasi klasifikasi digunakan formula

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{Jumlah prediksi yang dilakukan}}$$

$$= \frac{f_{11}+f_{22}+f_{33}}{f_{11}+f_{12}+f_{13}+f_{21}+f_{22}+f_{23}+f_{31}+f_{32}+f_{33}}$$

Untuk menghitung laju *error* (kesalahan prediksi) digunakan formula

$$\begin{aligned} \text{Laju Error} &= \frac{\text{Jumlah data yang diprediksi secara salah}}{\text{Jumlah prediksi yang dilakukan}} \\ &= \frac{f_{12}+f_{13}+f_{21}+f_{23}+f_{31}+f_{32}}{f_{11}+f_{12}+f_{13}+f_{21}+f_{22}+f_{23}+f_{31}+f_{32}+f_{33}} \end{aligned}$$

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data yang diambil dari rekapitulasi data jurusan dan rata-rata nilai mata pelajaran yang diambil dari database SMA Negeri 1 Jepara tahun pelajaran 2013/2014 yang banyaknya rekapitulasi data ada 348 buah. Selanjutnya data ini digunakan sebagai data simulasi klasifikasi perbandingan menggunakan *Decision Tree* dan *Support Vector Machine Multiclass (SVMM)*.

Tahapan analisis yang dilakukan adalah sebagai berikut:

3.1 Melakukan Klasifikasi Menggunakan *Decision Tree*

Pada tahap ini dilakukan persiapan data *training* dengan algoritma C4.5 sebagai berikut :

1. Menyiapkan data *training* dan *testing*. Data *training* diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam 3-kelas kategori mata pelajaran IPA, IPS dan BAHASA.
2. Menentukan akar dari pohon, dihitung terlebih dahulu *entropy*.
3. Selanjutnya akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi menjadi akar pertama.
4. Ulangi langkah ke-2 hingga semua *record* terpartisi.
5. Proses partisi *decision tree* akan berhenti saat :
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
 - c. Tidak ada *record* di dalam cabang yang kosong.
6. Studi Kasus *Decision Tree* menggunakan *software Matlab* selanjutnya diinterpretasikan dengan membuat aplikasi berbasis *Decision Tree* dengan GUI (*Graphical User Interface*).

3.2 Melakukan Klasifikasi Menggunakan SVMM

Pada tahap ini dilakukan persiapan data *training* dan *testing* dengan algoritma SVMM sebagai berikut :

1. Melakukan analisis data sesuai dengan SVMM
2. Melakukan pendekatan utama pada data *training* dan *testing* dengan cara menemukan dan menggabung beberapa fungsi pemisah persoalan klasifikasi SVM dua kelas untuk menyelesaikan klasifikasi SVMM dengan metode Satu Lawan Satu (SLU)
3. Studi Kasus SVMM menggunakan *software Matlab* yaitu dilakukan pembuatan persamaan antara kelas IPA dan IPS, persamaan antara kelas IPA dan Bahasa, serta persamaan antara kelas IPS dan Bahasa.
4. Membuat aplikasi klasifikasi SVMM berbasis text dan GUI (*Graphical User Interface*).

3.3. Analisis Hasil Perbandingan

Pada tahap ini dapat dilakukan evaluasi dari hasil *training* dan *testing*, dimana dilakukan perbandingan ketepatan klasifikasi dengan metode *Decision Tree* dan *Support Vector Machine Multiclass (SVMM)*.

4. HASIL DAN PEMBAHASAN

4.1 Analisis Klasifikasi Metode *Decision Tree*

Dilakukan analisis *Decision Tree* mengenai penjurusan siswa Sekolah Menengah Atas (SMA) dalam jurusan IPA, IPS dan Bahasa. Penentuan penjurusan berdasarkan Data Mining akan berbeda dengan yang bukan Data Mining.

4.2 Menentukan Akar Dari Pohon Keputusan

Dari data *traning* sebanyak 278 dan data *testing* sebanyak 70 buah, dihitung nilai *entropy*, *information gain*, dan *gain ratio* yang memiliki fitur berjumlah 3 menggunakan tipe numerik dengan rincian sebagai berikut :

- IPA = kontinu
- IPS = kontinu
- BAHASA = kontinu

Tabel 2. Hasil perhitungan *entropy* dan *gain* untuk node akar data *training*

| SIMPUL AKAR | | Total | 80 | | 85 | | Entropy Total |
|-------------|--------------------|-------------|----------|-------------|----------|-----|---------------|
| | | | <=80 | >80 | <=85 | >85 | |
| | | 278 | | | | | |
| N.IPA | J.IPA | 148 | 22 | 126 | 119 | 29 | 1,25861 |
| | J.IPS | 113 | 90 | 23 | 110 | 3 | |
| | J.BAHASA | 17 | 7 | 10 | 16 | 1 | |
| | Entropy Tiap Kelas | 1,197389 | 1,115429 | 0,373091 | 0,840189 | | |
| | GAIN | 0,108097651 | | 0,830072482 | | | |
| N.IPS | J.IPA | 148 | 20 | 128 | 105 | 43 | 1,25861 |
| | J.IPS | 113 | 58 | 55 | 103 | 10 | |
| | J.BAHASA | 17 | 5 | 12 | 15 | 2 | |
| | Entropy Tiap Kelas | 1,403353 | 1,041475 | 0,632507 | 1,190899 | | |
| | GAIN | 0,10909244 | | 0,515629854 | | | |
| N.BAHASA | J.IPA | 148 | 13 | 135 | 127 | 21 | 1,25861 |
| | J.IPS | 113 | 37 | 76 | 107 | 6 | |
| | J.BAHASA | 17 | 3 | 14 | 13 | 4 | |
| | Entropy Tiap Kelas | 1,277248 | 0,736541 | 0,559935 | 1,115775 | | |
| | GAIN | 0,418984368 | | 0,636692966 | | | |

$$\begin{aligned}
 Gain Ratio(S, J) &= \frac{Gain (S, J)}{Split Information (S, J)} \\
 &= \frac{0,830072482}{0,52562704} = 1,579204299
 \end{aligned}$$

Jadi pemilihan fitur yang tepat untuk menjadi pemecah pada *node* data *training* sebesar 1,579204299

Tabel 3. Hasil perhitungan *entropy* dan *gain* untuk node akar data *testing*

| SIMPUL AKAR | | 80 | | | Entropy Total | |
|-------------|--|--------------------|------|--------------|---------------|--|
| | | Total | <=80 | | | |
| | | | >80 | | | |
| N.IPA | | 70 | | | 1,24318 | |
| | | J.IPA | 38 | 6 | 32 | |
| | | J.IPS | 28 | 27 | 1 | |
| | | J.BAHASA | 4 | 3 | 1 | |
| | | Entropy Tiap Kelas | | 0,78234 | 0,880472 | |
| | | GAIN | | 0,413175803 | | |
| N.IPS | | J.IPA | 38 | 6 | 32 | |
| | | J.IPS | 28 | 18 | 10 | |
| | | J.BAHASA | 4 | 3 | 1 | |
| | | Entropy Tiap Kelas | | 1,330245 | 1,050569 | |
| | | GAIN | | 0,084736291 | | |
| N.BAHASA | | J.IPA | 38 | 10 | 28 | |
| | | J.IPS | 28 | 15 | 13 | |
| | | J.BAHASA | 4 | 1 | 3 | |
| | | Entropy Tiap Kelas | | 1,489234 | 1,239291 | |
| | | GAIN | | -0,088946583 | | |

$$\begin{aligned}
 \text{Gain Ratio}(S, J) &= \frac{\text{Gain}(S, J)}{\text{Split Information}(S, J)} \\
 &= \frac{0,413175803}{0,999411065} = 0,41341928
 \end{aligned}$$

Jadi pemilihan fitur yang tepat untuk menjadi pemecah pada *node* data training sebesar 0,41341928

4.3 Analisis Klasifikasi Metode SVMM

Persoalan klasifikasi menggunakan metode SVMM adalah melakukan klasifikasi lebih dari satu kelas. Dalam kasus ini digunakan persoalan 3 kelas yang terdiri dari jurusan IPA, IPS, dan Bahasa.

Jumlah fungsi pemisah dapat dirumuskan dengan $\frac{k(k-1)}{2}$ dengan $k=3$ kelas.

$$\text{Jadi nilai } \frac{k(k-1)}{2} = \frac{3(3-1)}{2} = 3$$

Fungsi pemisah per kelas yaitu :

- Kelas 1 = pemisah antara jurusan IPA dan jurusan IPS (ρ_{12})
 - Data *training* ρ_1 diberi label +1
 - Data *training* ρ_2 diberi label -1
- Kelas 2 = pemisah antara jurusan IPA dan jurusan Bahasa (ρ_{13})
 - Data *training* ρ_1 diberi label +1
 - Data *training* ρ_3 diberi label -1
- Kelas 3 = pemisah antara jurusan IPS dan jurusan Bahasa (ρ_{23})
 - Data *training* ρ_2 diberi label +1
 - Data *training* ρ_3 diberi label -1

4.4 Analisis Hasil Perbandingan

Hasil dari prediksi tingkat *error* dari klasifikasi *decision tree* didapatkan hasil sebagai berikut:

Tabel 4. Hasil Prediksi Klasifikasi Data *Training Decision Tree*

| fij | | Kelas Hasil Prediksi (j) | | |
|-------------------|--------|--------------------------|-----|--------|
| | | IPS | IPA | BAHASA |
| Kelas Asli (i) | IPS | 100 | 13 | 0 |
| | IPA | 17 | 131 | 0 |
| | BAHASA | 6 | 9 | 2 |

Tabel 5. Hasil Prediksi Klasifikasi Data *Testing Decision Tree*

| fij | | Kelas Hasil Prediksi (j) | | |
|-------------------|--------|--------------------------|-----|--------|
| | | IPS | IPA | BAHASA |
| Kelas Asli (i) | IPS | 27 | 1 | 0 |
| | IPA | 5 | 33 | 0 |
| | BAHASA | 1 | 1 | 2 |

Tabel 6. Hasil Prediksi Data *Training* Klasifikasi SVMM

| fij | | Kelas Hasil Prediksi (j) | | |
|-------------------|--------|--------------------------|-----|--------|
| | | IPA | IPS | BAHASA |
| Kelas Asli (i) | IPA | 134 | 14 | 0 |
| | IPS | 39 | 74 | 0 |
| | BAHASA | 9 | 5 | 3 |

Tabel 7. Hasil Prediksi Data *Testing* Klasifikasi SVMM

| fij | | Kelas Hasil Prediksi (j) | | |
|-------------------|--------|--------------------------|-----|--------|
| | | IPA | IPS | BAHASA |
| Kelas Asli (i) | IPA | 36 | 2 | 0 |
| | IPS | 4 | 24 | 0 |
| | BAHASA | 1 | 2 | 1 |

Tabel 8. Perbandingan Data *Training*

| Analisis Perbandingan Data Training | | |
|-------------------------------------|-----------------|--------------------|
| Metode | Akurasi dalam % | Laju Error dalam % |
| Decision Tree | 83,81% | 16,19% |
| SVMM | 75,90% | 24,10% |

Tabel 9. Perbandingan Data *Testing*

| Analisis Perbandingan Data Testing | | |
|------------------------------------|-----------------|--------------------|
| Metode | Akurasi dalam % | Laju Error dalam % |
| Decision Tree | 88,57% | 11,43% |
| SVMM | 87,14% | 12,86% |

5. KESIMPULAN

Hasil analisis menunjukkan metode klasifikasi *decision tree* data dibangun berdasarkan objek yang dimiliki dalam data latih. *Decision tree* secara penuh merefleksikan semua isi himpunan data latih daripada klasifikasi SVMM karena harus menyimpan sebagian kecil data latih untuk digunakan kembali pada saat proses prediksi. Tingkat akurasi dan laju error lebih baik pada *decision tree* dimana nilai total akurasi sebesar 88,57% dan error 11,43% untuk klasifikasi *decision tree* dan total akurasi sebesar 87,14% dan error 12,86% untuk klasifikasi SVMM.

6. DAFTAR PUSTAKA

- [1] Gunn, S.R. 1998. *Support Vector Machines for Classification and Regression*. University of Southampton : Southampton.
- [2] Han, J dan Kamber, M. 2006. *Data Mining Concepts and Techniques – 2nd Ed*. San Fransisco: Elsevier Inc.
- [3] Hermawati, A.F. 2013. *Data Mining*. Surabaya: Universitas 17 Agustus 1945 & ANDI.
- [4] Hsu, C.W. dan Lin, C.J. 2002. *A Comparison of Methods for Multiclass Support Vector Machines*. IEE Transactions on Neural Network, 13(2), 415-425.
- [5] Kerami, D. dan Murfi,H. 2004. *Kajian Kemampuan Generalisasi Support Vector Machine dalam Pengenalan Jenis Splice Site pada Barisan DNA*. Jurnal Makara Sains Vol 8 No 3. Hal 89-95.
- [6] Prasetyo, E. 2012. *Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB*. Yogyakarta: ANDI.
- [7] Prasetyo, E. 2014. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: ANDI.
- [8] Santosa, B. dan Trafalis, T. 2004. *Multiclass Procedure for Minimax Probability Machine*. ASME Press, 14, 447-452.
- [9] Santosa, B. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- [10] Sumathi, S. 2006. *Introduction to Data Mining and Its Applications*. Germany: Springer Verlag berlin Heidelberg.
- [11] Vapnik, V. dan Cortes, C. 1995. *Support Vector Networks*. Machine Learning, 20, 273-297
- [12] Weston, J. dan Watkins, C. 1998. *Support Vector Machine for Multiclass Pattern Recognition. Proceeding of the seventh European Symposium on Articial Neural Networks ICANN'97*.
- [13] Widodo, P. P., Handayanto, R. T. dan Herlawati. 2013. *Penerapan Data Mining dengan Matlab*. Bandung: Rekayasa Sains.