

Perbandingan Antara Metode Decision Tree Dan Support Vector Machine Pada Model Rekomendasi Mobil Bekas

Ali Maksum¹, Daniel Swanjaya²

Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail : ¹aliabang622@gmail.com, ²daniel@unpkediri.ac.id.

Abstrak –Perkembangan usaha di bidang penjualan mobil bekas sangat pesat sehingga banyak konsumen memandang dari pengaruh *product quality* terhadap *brand loyalty* melalui kepuasan konsumen sebagai variabel mediasi. Kepuasan konsumen mempengaruhi terhadap signifikan *brand loyalty* dan sebagai variabel mediasi dari *product quality* terhadap *brand quality*. Peneliti mencoba membuat perbandingan antara metode *decision tree* dan *support vector machine* untuk membandingkan akurasi yang lebih baik, hasil dari penelitian ini perbandingan dari *Decision Tree* dan *Support Vector Machine* dari data *Training decision tree* sebesar 2,5 % dan *Support Vector Machine* sebesar 9,1 %, untuk data *testing* dari *Decision Tree* sebesar 6,2 % dan *Support Vector Machine* sebesar 2,2 %. Bisa disimpulkan bahwa pada metode *Decision Tree* akurasi lebih baik daripada metode *Support Vector Machine*.

Kata Kunci— Perbandingan Metode *Decision tree* dan *Support Vector Machine*, Mobil, *Soft Computing*

1. PENDAHULUAN

Perkembangan usaha di bidang penjualan mobil bekas yang semakin pesat dapat dilihat dengan banyaknya dealer atau showroom mobil yang berdiri di Indonesia. Kemajuan aktifitas penjualan mobil yang semakin pesat menimbulkan tingkat persaingan usaha yang ketat. Dengan adanya produk sejenis yang dijual oleh dealer atau usaha jenis showroom jual beli mobil bekas yang berbeda, yang mana keinginan konsumen yang beraneka ragam membuat konsumen bebas untuk menentukan produk yang dapat memuaskan kebutuhannya. Oleh karena itu, perusahaan harus menemukan strategi untuk lebih maju dari pesaingnya, [1].

Showroom Wira Jaya Motor adalah tempat jual beli mobil *second*, dalam pembelian mobil dari pihak penjual sering sekali mencari mobil yang memiliki spesifikasi yang elegan dan keluaran terbaru. Konsumen juga lebih memilih mobil yang memiliki mesin yang masih baik dan *body* mobil yang masih bagus yang dimiliki mobil keluaran terbaru. Pembelian mobil bekas bukan hal yang mudah dilakukan oleh konsumen karena banyak kriteria yang harus diperhatikan dalam membeli mobil bekas, seperti kriteria mesin, tahun, dokumen, harga, *seating*, *interior*, *exterior*, *body type*, *band* (merk mobil).

Konsumen juga harus membandingkan kriteria mana yang harus didahulukan, ditengah situasi yang belum stabil karena masa pandemi ini, membuat para produsen mobil berfikir ulang untuk melakukan inovasi agar dapat terus eksis dalam dunia otomotif dengan memproduksi mobil yang sesuai dengan minat konsumen di Indonesia. Perbandingan antara

masyarakat tingkat menengah keatas sedikit jumlahnya dibandingkan dengan masyarakat tingkat menengah kebawah, kondisi perekonomian masyarakat yang demikian membuat konsumen akan lebih mempertimbangkan untuk membeli mobil yang terjangkau dari sisi harga dan kapasitas daya angkut penumpang yang lebih besar. Harga mobil dengan *budget* 100-500 juta untuk mobil baru merupakan *range* harga yang ideal bagi masyarakat tingkat menengah kebawah. Sekitar 20 *brand* mobil baru yang ada di Indonesia sekarang ini bisa diperoleh dengan *range* harga 100-500 juta dengan kriteria yang diinginkan. [2].

Pengaruh *Product Quality* Terhadap *Brand Loyalty* Konsumen Honda Jazz Dengan Kepuasan Konsumen Sebagai Variabel *Intervening* Di Surabaya. Tujuan penelitian ini adalah menjelaskan pengaruh *product quality* terhadap *brand loyalty* melalui kepuasan konsumen sebagai variabel mediasi. Jenis penelitian ini adalah kuantitatif. Variabel penelitian ini meliputi (variabel bebas). Jumlah sampel penelitian ini sebanyak 100 sampel. Kepuasan konsumen memiliki pengaruh yang signifikan terhadap *brand loyalty*, dan kepuasan konsumen merupakan variabel mediasi dari pengaruh *product quality* terhadap *brand quality*. [3].

Implementasi Decision Tree Untuk Penjurusan Siswa SMA Menggunakan Matlab, Penelitian ini dilakukan untuk menemukan rule (aturan) hubungan riwayat akademik terhadap hasil belajar mahasiswa di STMIK Royal Kisaran Program Studi Sistem Informasi, [4].

Penentuan Kendaraan Mobil Bekas Menggunakan Metode *Topsis*, informasi tentang

mobil bekas yang dapat membantu konsumen dalam membeli kendaraan. Namun terkadang manusia sering lupa, apalagi kalau kriteria tersebut sangat banyak seperti harga mobil, aksesoris mobil, kapasitas penumpang dan tahun produksi sehingga konsumen sering mengalami kesalahan dalam pemilihan kendaraan bekas yang akan dipinangnya . [5].

Berdasarkan pernyataan perbandingan dengan menerapkan metode klasifikasi antara metode *Decision Tree* dan *Support Vector Machine* pada model rekomendasi mobil bekas. Saya memilih metode ini dikarenakan lebih kompleks dan akurat. Penggunaan metode *Decision Tree* untuk mengklasifikasi dan memprediksi data mobil bekas dan proses pada metode *Decision Tree* adalah mengubah data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* sedangkan pada metode *Support Vector Machine* untuk mencari *hyperplane* terbaik dari data mobil bekas berfungsi sebagai proses pemisah antara dua buah kelas pada input space. *Input* data berupa linear dan non linear. Jika input data berupa linear maka pemisah *hyperplane* dapat diberikan dalam persamaan.

2. METODE PENELITIAN

Metode penelitian ini menggunakan metode *Decision Tree* dan metode *Support Vector Machine*.

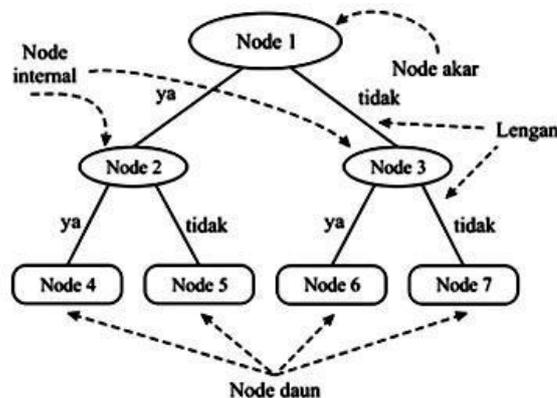
2.1 Metode Decision Tree

Klasifikasi *Decision Tree* merupakan salah satu teknik terkenal dalam data *mining* dan merupakan salah satu metode yang populer dalam menentukan keputusan suatu kasus. Metode ini tidak memerlukan proses pengelolaan pengetahuan terlebih dahulu dan dapat menyelesaikan kasus-kasus yang memiliki dimensi besar, [6].

Algoritma decision tree dibangun dalam suatu metode *rekursif topdown divide and-conquer*, seluruh contoh pelatihan dimulai dari simpul *root*, lalu dilakukan pengujian, kemudian mencabang ke jalur yang benar berdasarkan hasil pengujian. Apakah leaf bisa ditemukan? Jika yes, masukan contoh ini ke kelas target, jika tidak maka kembali ke langkah pertama. Atribut-atribut yang berada dalam suatu kategori (jika bernilai *kotinu*, maka nilai-nilai tersebut didiskritkan dahulu). Contoh-contoh dipartisi secara rekursif berdasarkan atribut terpilih. Atribut-atribut uji dipilih berdasarkan *heuristic* atau pengukuran *statistic* (misal, *information gain*).

Decision Tree merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode *Decision Tree* mengubah fakta yang sangat besar menjadi sebuah pohon keputusan yang

mempresentasikan aturan-aturan. Aturan dapat dengan mudah dipahami dengan Bahasa alami. Proses pada *Decision Tree* adalah mengubah data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*, .[7].



Gambar 1 Pohon Keputusan Decission Tree

Pada gambar 1 dapat dijelaskan sejumlah elemen sebagai berikut :

- Node akar** : tidak mempunyai lengan masukan dan mempunyai nol atau lebih lengan keluaran.
- Node internal** : adalah setiap node yang bukan daun(nonterminal) yang mempunyai tepat satu lengan masukan dan dua atau lebih lengan keluarannya. Node ini menyatakan pengujian yang didasarkan pada nilai fitur.
- Lengan** : setiap cabang menyatakan nilai hasil pengujian di node bukan daun.
- Node daun (terminal)** : adalah node yang mempunyai tempat satu lengan masukan dan tidak mempunyai lengan keluaran. Node ini menyatakan sebuah label kelas (keputusan).

Pohon keputusan mempunyai tiga klasifikasi diantaranya :

- Pohon klasifikasi, digunakan untuk melakukan prediksi ketika ada data baru yang belum diketahui label kelasnya. Pendekatan ini yang paling banyak digunakan
- Pohon regresi, ketika hasil prediksi dianggap segala nilai nyata yang mungkin akan didapatkan.
- CART (C&T), ketika masalah klasifikasi dan regresi digunakan bersamaan. Pohon keputusan yang diinduksi tidak selalu sama dengan beberapa percobaan karena urutan atau cara pemilihan fitur sebagai pemecah cabang. Ada banyak pilihan algoritma untuk menginduksi pohon keputusan seperti *Hunt*, CART (C&T), ID3, C4.5, SLIQ, SPRINT, QUEST, DTREG, THAID, dan sebagainya. Saat menyusun sebuah *Decision Tree* pertama yang

harus dilakukan adalah menentukan atribut mana yang akan dijadikan simpul akar dan atribut mana yang akan dijadikan simpul selanjutnya. Pemijilihan atribut yang baik memungkinkan untuk mendapatkan *Decision Tree* yang paling kecil ukurannya. Atau atribut yang bisa memisahkan obyek menurut kelasnya. Atribut yang dipilih adalah atribut yang menghasilkan simpul yang paling bersih. Ukuran *piurity* dinyatakan dengan tingkat *impurity*, dan untuk menghitungnya dapat dilakukan dengan menggunakan konsep *Entropy*, *Entropy* menyatakan suatu kesimpulan objek. Jika diberikan sekumpulan objek dengan label/output Y yang terdiri dari objek berlabel 1, 2 sampai n, *Entropy* dari objek dengan n kelas ini dapat dihitung dengan persamaan 1.

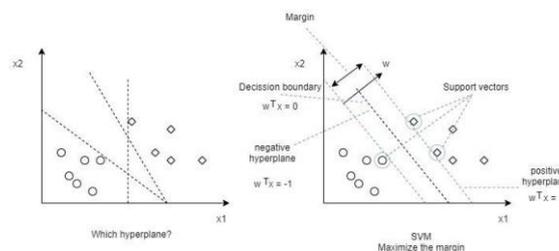
$$Entropy(y) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_n \log_2 p_n, \dots \dots \dots (1)$$

1. Setelah menghitung *Entropy*, hitung *information Gain*.
 2. *Information gain* adalah kriteria yang paling populer untuk pemilihan atribut. *Information gain* dapat dihitung dari *output* data atau *variabel dependent* Y yang dikelompokkan berdasarkan atribut A, dinotasikan dengan *gain* (y,A). *information gain*, *gain* (y,A), dari atribut A *relative* terhadap *output* data y adalah : dimana nilai (A) adalah semua nilai atribut A, dan adalah subset dari y dimana nilai c.
- Algoritma C4.5 sendiri merupakan pengembangan dari algoritma ID3, dimana pengembangan dilakukan dalam hal : bisa mengatasi missing data, bisa mengatasi data kontinyu, purning. Secara umum langkah-langkah algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :
- a. Pilih atribut sebagai akar.
 - b. Buat cabang untuk tiap-tiap nilai
 - c. Bagi kasus dalam cabang.
 - d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.
 - e. Menentukan rule untuk mengetahui aturan-aturannya.

2.2 Metode Support Vector Machine

Support Vector Machine adalah usaha untuk mencari *hyperlane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada *input space*. *Input* data berupa linear dan non linear. Jika *input* data berupa linear maka pemisah *hyperlane* dapat diberikan dalam persamaan, [8].

Algoritma *Support Vector Machine* adalah Algoritma *supervised* yang berupa klasifikasi dengan cara membagi data menjadi dua buah kelas menggunakan garis *hyperlane*. Pada permasalahan yang kompleks atau permasalahan dengan parameter yang banyak, metode ini juga sangat baik untuk digunakan, (Octaviani, et al., 2014).[9]



Gambar 2 Grafik SVM Mencari Hyperlane Terbaik

Pada gambar 2 posisinya berada ditengah-tengah antara dua kelas, maksudnya jarak antara *hyperlane* dengan objek-objek data berada dengan kelas yang berdekatan (terluar) yang diberi tanda bulat kosong dan positif. Dalam *SVM* objek data terluar yang paling dekat dengan *hyperlane* disebut *Support Vector*. Objek yang disebut sebagai *support vector* paling sulit diklasifikasikan dikarenakan posisi yang hampir tumpang tindih (*overlap*) dengan kelas lain. Mengingat sifatnya yang kritis, *support vector* inilah yang diperhitungkan untuk menemukan *hyperlane* yang paling optimal oleh *SVM*.

$$f(X) = wTx + b, (3) \dots \dots \dots (2)$$

Dimana w adalah n-dimensi bobot *vector* dan b adalah pengali skala nilai bias. Persamaan ini menemukan maksimum margin untuk memisahkan kelas dari kelas positif dari kelas *negatif*. Fungsi keputusan ditujukan dalam persamaan. Pemilihan parameter pada *Support Vector Machine* untuk mendapatkan kinerja yang tinggi, beberapa parameter dari algoritma *Support Vector Machine* harus diperbaiki, termasuk :

- A. Pemilihan fungsi karnel.
- B. Kinerja *SVM* tergantung pada pemilihan fungsi karnel, besaran parameter dan penentuan karnel C. fungsi karnel yang berbeda memperoleh tingkat keberhasilan yang berbeda untuk jenis data aplikasi. Ketika nilai penentuan parameter C yang dipilih terlalu besar atau terlalu kecil, *generalisasi SVM* semakin berkurang. Jika parameter karnel dan penentuan karnel yang tepat dipilih, kinerja *SVM* akan optimal.
- C. Parameter karnel (s);
- D. Parameter regularasi (C,v, ε) untuk *tradeoff* antara kompleksitas model dan akurasi *mode*.

2.3 Perancangan Sistem

2.3.1 Pengumpulan Data

Pada tahap ini adalah proses pengumpulan data yang dibutuhkan untuk, serta melakukan analisa atau pengamatan pada data yang sudah terkumpul untuk selanjutnya diolah lebih lanjut. Data yang didapat antara lain:

Tabel 1 Data Mentah Mobil Bekas

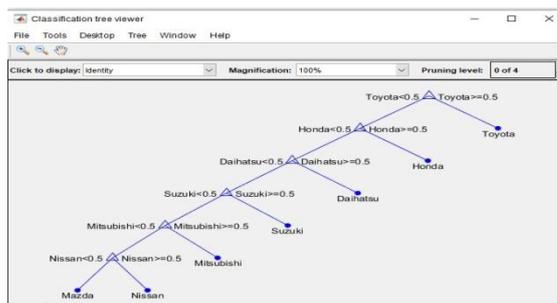
BAND
Suzuki
Audi
Honda
Daihatsu
Mitsubishi
Honda
Honda
Toyota
Mazda
Nissan

Pada tabel 1 adalah data mentah yang diambil dari data mobil bekas yaitu kolom band yang diambil sampel 10 data untuk pengujian dimana data tersebut harus dinormalisasi terlebih dahulu untuk menjadikan nilai 0-1. Yang mana akan diproses dengan metode *Decision Tree*, untuk mengetahui klasifikasi dan akurasi.

Suzuki	Audi	Honda	Daihatsu	Mitsubishi	Toyota	Mazda	Nissan	Datsun	Mercedes-Benz	BAND
1	0	0	0	0	0	0	0	0	0	Suzuki
0	1	0	0	0	0	0	0	0	0	Audi
0	0	1	0	0	0	0	0	0	0	Honda
0	0	0	1	0	0	0	0	0	0	Daihatsu
0	0	0	0	1	0	0	0	0	0	Mitsubishi
0	0	1	0	0	0	0	0	0	0	Honda
0	0	1	0	0	0	0	0	0	0	Honda
0	0	0	0	0	1	0	0	0	0	Toyota
0	0	0	0	0	0	1	0	0	0	Mazda
0	0	0	0	0	0	0	1	0	0	Nissan

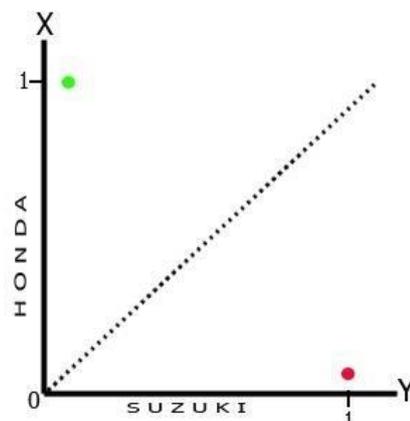
Gambar 4 Data Hasil Normalisasi

Gambar 4 adalah data yang sudah ternormalisasi kemudian masuk pada proses decision tree yang pertama menentukan atribut mana yang akan menjadi akar, kemudian membuat cabang untuk tiap-tiap nilai, kemudian bagi kasus dalam setiap cabang, ulangi kasus pada setiap cabang sampai cabang memiliki *class* yang sama.



Gambar 5 Pohon Keputusan

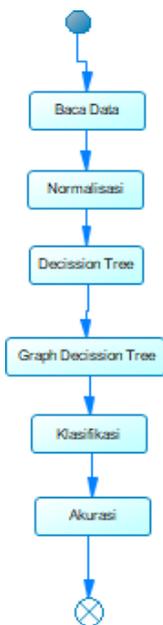
Dari gambar 5 adalah graph pada Decision Tree yaitu sebuah pohon keputusan yang menampilkan susunan yang sudah diatur oleh sistem *matlab*, yang mana pada setiap cabangnya akan menghasilkan *output* yang berupa kelas.



Gambar 6 Grafik SVM Dengan Hypelane Terbaik

Pada gambar 3 disimpulkan bahwa antara posisinya berada ditengah-tengah antara dua kelas, maksudnya jarak nantara *hyperlane* dengan objek-objek data berada dengan kelas yang berdekatan (terluar) yang diberi tanda bulat. Dalam *Support Vector Machine* objek data terluar yang paling dekat dengan *hyperlane* disebut *support vector*.

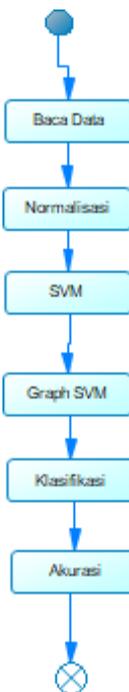
a) Activity diagram UML



Gambar 3 Diagram Activity Decision Tree

Pada gambar 8 merupakan proses *activity* diagram pada metode *Decision Tree* yang diawali dengan membaca data, setelah itu data akan dinormalisasi berupa nilai 0-1, kemudian memulai pemrosesan dengan metode *Decision Tree* dengan proses menentukan atribut dan *class*, yang bertujuan dengan hasil sebuah pohon keputusan dengan aturan-aturan, yang mana akan diakhiri dengan mengetahui akurasi dari metode *Decision Tree* tersebut.

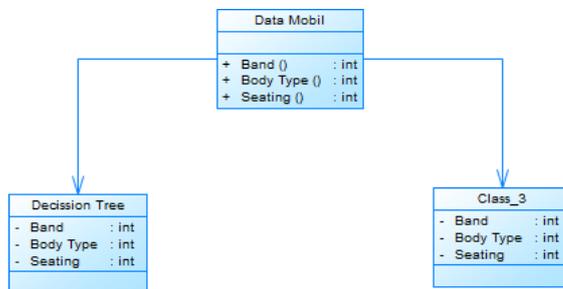
Pada gambar 7 adalah proses dari metode *Support Vector Machine* yang juga diawali dengan membaca data yang sudah dinormalisasi kemudian masuk ke pemrosesan dengan metode *Support Vector Machine* untuk mengetahui *hyperlane* terbaik antara dua buah *class* dengan menampilkan *graph*, kemudian akan diketahui klasifikasinya dengan berupa akurasi terbaik.



Gambar 4 Diagram Activity Support Vector Machine

b) Class diagram

Class diagram merupakan diagram yang bisa digunakan untuk menentukan alur struktur sistem dengan cara memodelkan kelas, atribut, operasi dan hubungan antara objek satu sama lain.

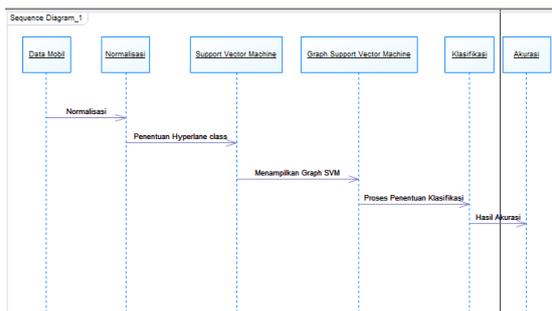


Gambar 5 Diagram Class

Pada gambar 9 dijelaskan bahwa data mobil yang sudah didapatkan akan diolah menggunakan metode *Decision Tree* dan metode *Support Vector Machine*. Kemudian hasil dari kedua metode akan dibandingkan.

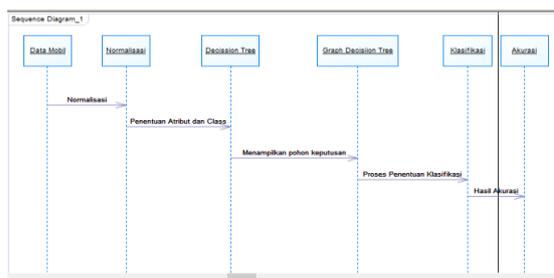
c) Sequence Diagram

Sequence diagram merupakan suatu diagram yang menggambarkan interaksi objek dan mengindikasikan (memberi petunjuk atau tanda) komunikasi antara objek.



Gambar 6 Diagram Sequence Metode Decision Tree

Pada gambar 10 dijelaskan bahwa proses pertama yang dilakukan adalah membaca data mobil kemudian dilakukan Normalisasi data sehingga data dapat diolah, selanjutnya menentukan atribut dan class sebagai bagian dari Decision Tree kemudian data diolah menggunakan algoritma Decision Tree sehingga dapat dihasilkan graph sebuah Pohon keputusan dan akurasi.



Gambar 7 Diagram Sequence Metode Support Vector Machine

Pada gambar 11 juga dijelaskan untuk proses pertama membaca data mobil bekas, kemudian data dinormalisasi untuk bisa diolah, kemudian data yang sudah dinormalisasi diolah menggunakan metode Support Vector Machine untuk mengetahui Hyperlane (batas keputusan) kemudian svm menampilkan Graph Support Vector Machine dan diakhiri dengan menampilkan akurasi.

3. HASIL DAN PEMBAHASAN

Untuk hasil dari kedua metode selanjutnya akan dilakukan perbandingan metode mana yang lebih akurat untuk memperoleh akurasi terbaik pada data mobil bekas. Simulasi perbandingan antara metode Decision Tree dan metode Support Vector Machine, berikut ini adalah tabel hasil perbandingan dari metode Decision Tree dan Support Vector Machine.

Tabel 2 Hasil Perbandingan Data Training

Metode	Akurasi (%)
Decision tree	2,50
SVM	9,10

1. Perbandingan data training

Hasil perbandingan dari Decision Tree dan Support Vector Machine dari data Training Decision Tree sebesar 2,5 % dan Support Vector Machine sebesar 9,1 % , untuk data testing dari Decision Tree sebesar 6,2 % dan Support Vector Machine sebesar 2,2 % . Bisa disimpulkan bahwa pada metode Decision Tree akurasi lebih baik daripada metode Support Vector Machine.

Tabel 3 Hasil Perbandingan Data Testing

Hasil perbandingan data Testing	
Metode	Akurasi dalam %
Decision tree	6,2 %
SVM	2,2 %

4. SIMPULAN

Dari Hasil analisis menunjukkan metode klasifikasi Decision Tree data dibangun berdasarkan objek yang dimiliki dalam data training. Decision Tree secara penuh merefleksikan semua data training dan data testing daripada klasifikasi SVM karena harus menyimpan sebagian kecil data training dan data testing untuk digunakan kembali pada saat proses prediksi. Tingkat akurasi lebih baik pada Decision Tree dimana Hasil perbandingan dari Decision Tree dan Support Vector Machine dari data Training Decision Tree sebesar 2,5 % dan Support Vector Machine sebesar 9,1 % , untuk data testing dari Decision Tree sebesar 6,2 % dan Support Vector Machine sebesar 2,2 % . Bisa disimpulkan bahwa pada metode Decision Tree akurasi lebih baik daripada metode Support Vector Machine.

5. SARAN

Semoga dengan penelitian ini bisa menjadi referensi untuk pengembangan sistem perbandingan antara metode Decision Tree dan metode Support Vector Machine yang lebih baik lagi, sebagai acuan untuk pengembangan dalam sebuah sistem perbandingan. Decision Tree dan Support Vector Machine dari data Training Decision Tree sebesar 2,5 % dan Support Vector Machine sebesar 9,1 % , untuk data testing dari Decision Tree sebesar 6,2 %

dan *Support Vector Machine* sebesar 2,2 %. Bisa disimpulkan bahwa pada metode *Decision Tree* akurasi lebih baik daripada metode *Support Vector Machine*.

DAFTAR PUSTAKA

- [1] Utama, P. (2016). BAB I PENDAHULUAN A .Latar Belakang Penelitian. [http://repository.unpas.ac.id/12341/4/BAB I draft.pdf](http://repository.unpas.ac.id/12341/4/BAB_I_draft.pdf).
- [2] Saputra, A. (2011). Sistem Informasi Pemilihan Mobil Bekas Menggunakan *Decision Support System Analytical Hierarchy Process* Pada *Showroom Yokima Motor Bandung*. *Jurnal Teknik Informatika Dan Sistem Informasi*, Vol. 1, No. 2, Hal:88–101. <https://doi.org/10.28932/jutisi.v1i2.370>.
- [3] Hartono, B. S. (2018). Pengaruh *Product Quality* Terhadap *Brand Loyalty* Konsumen Honda Jazz Dengan Kepuasan Konsumen Sebagai *Variabel Intervening* Di Surabaya. *Jurnal Strategi Pemasaran*.
- [4] Sembiring, M. A. (2016). Penerapan *Metode Decision Tree* Algoritma C45 Untuk Memprediksi Hasil Belajar Mahasiswa Berdasarkan Riwayat Akademik. *Jurnal Teknologi Dan Sistem Informasi*.
- [5] Muharram, F., Sukabumi, U. M., & Barat, J. (2020). Penentuan Kendaraan Mobil Bekas Menggunakan. Vol. 4, No. 2, Hal:194–199.
- [6] Putranto, R. A., Wuryandari, T., & Sudarno. (2015). Perbandingan Analisis Klasifikasi antara *Decision Tree* dan *Support Vector Machine Multiclass* untuk Penentuan Jurusan pada Siswa SMA. *JURNAL GAUSSIAN*.
- [7] Prasetyo, K. G., & Pahlevi, S. M. (2019). Analisis Perbandingan Algoritma *Decision Tree* Dengan *Support Vector Machine* Untuk Mendeteksi Kompetensi Mahasiswa Konsentrasi Informatika Komputer Studi Kasus : Politeknik Lp3I Jakarta , Kampus Depok. Vol. 5, No. 60, 11–26.
- [8] Jacobus, A., Winarko, E. (2014). Penerapan Metode *Support Vector Machine* pada Sistem Deteksi Intrusi secara *Real-time*. *IJCCS Indonesian Journal of Computing and Cybernetics Systems*, Vol. 18, No. 1, Hal:13. <https://doi.org/10.22146/ijccs.3491>