

COMPARISON OF ALGORITHM BETWEEN CLASSIFICATION & REGRESSION TREES AND SUPPORT VECTOR MACHINE IN DETERMINING STUDENT ACCEPTANCE IN STATE UNIVERSITIES

M. Anwar Sadat^{*1}, Pujiono², Anggun Pambudi³, Sholihul Ibad⁴

^{1,2,3}Master's Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro Semarang, Indonesia

⁴Information Systems Study Program, Institut Teknologi dan Bisnis Tuban, Indonesia
Email: m.anwarsadat@gmail.com

(Article received: December 04, 2023; Revision: December 28, 2023; published: January 08, 2024)

Abstract

Higher education entrance selection activities are intended to obtain superior student candidates. The opportunity to take part in the selection is given to all high school graduate students and equivalent. The student entrance test at PTN consists of three types of selection routes, namely the SNMPTN or invitation route, the SBMPTN, and the independent examination held by state universities. Starting from the dataset, data selection was carried out from 143 students' data and 7 attribute selections were carried out using preprocessing using data transformation first. The aim of using data transformation is to simplify the data training process for MAN 1 students in Cirebon. Preprocessing for prediction of classification results, accuracy of testing data for 143 students is implemented in the program and the resulting calculation process will be more efficient. After going through the preprocessing stage, the data is divided into training data and testing data using 10-fold cross validation. Next, for the classification process, a comparison of two methods will be used, namely for the first method using CART, the second method using SVM by adding Gain ratio weighting. The results of the research show that in the first experiment the researcher carried out a comparative trial of cross validation and classification performance and used the CART and SVM algorithms. The results comparison using the CART algorithm gets an accuracy of 86.10% and the SVM algorithm method for classifying students entering PTN was 86.71%.

Keywords: Cart Algorithm, PTN participants, SVM algorithm.

KOMPARASI ALGORITMA ANTARA CLASSIFICATION & REGRESSION TREES DAN SUPPORT VECTOR MACHINE DALAM PENENTUAN PENERIMAAN SISWA DIDIK DI PERGURUAN TINGGI NEGERI

Abstrak

Kegiatan seleksi masuk perguruan tinggi (PT) dimaksudkan untuk memperoleh calon mahasiswa yang unggul. Kesempatan untuk mengikuti seleksi diberikan kepada seluruh siswa tamatan SMU dan setara. Tes masuk mahasiswa di PTN terdiri atas tiga macam jalur seleksi, yaitu SNMPTN atau jalur undangan, jalur SBMPTN, dan jalur ujian mandiri yang diadakan oleh universitas negeri. Dimulai dari dataset dilakukan seleksi data dari data 143 siswa dan 7 atribut seleksi sejumlah preprocessing menggunakan data transformation terlebih dahulu. Adapun tujuan menggunakan data transformation agar mempermudah proses data training nilai siswa MAN 1 kota Cirebon. Preprocessing prediksi hasil klasifikasi akurasi data testing 143 siswa diimplementasi pada program serta proses perhitungan yang dihasilkan akan lebih efisien. Setelah melalui tahap preprocessing, data tersebut dibagi menjadi data training dan data testing menggunakan cross validation sebanyak 10-fold cross validation. Selanjutnya untuk proses klasifikasi akan digunakan perbandingan dua metode, yaitu untuk metode pertama menggunakan CART, metode kedua menggunakan SVM dengan menambahkan pem bobotan Gain ratio. Hasil penelitian menunjukkan bahwa pada eksperimen yang pertama peneliti melakukan uji coba perbandingan cross validation dan performance klasifikasi dan menggunakan algoritma CART dan SVM dengan jumlah data 143 siswa dengan atribut 8 dan perbandingan antara data training testing Masuk pengisian data peserta PTN masuk atau tidak masuk, Hasil eksperimen perbandingan menggunakan algoritma CART mendapatkan akurasi 86.10%. Dari hasil optimasi algoritma CART dengan SVM berbasis algoritma untuk klasifikasi siswa masuk PTN metode algoritma CART untuk mengklasifikasi siswa masuk perguruan tinggi negeri sebesar 86.10%, metode algoritma SVM untuk mengklasifikasi siswa masuk perguruan tinggi negeri sebesar 86.71%.

Kata kunci: *Algoritma Cart, Algoritma SVM, Peserta PTN.*

1. PENDAHULUAN

Setiap siswa yang akan menyelesaikan studinya di SMA akan diperhadapkan pada berbagai pilihan, yaitu apakah akan melanjutkan pendidikan ke perguruan tinggi, mengikuti kursus, mencari pekerjaan. Bagi siswa yang akan menetapkan pilihan untuk melanjutkan ke perguruan tinggi, tentu akan diperhadapkan lagi pada pilihan perguruan tinggi mana yang akan dimasuki dan jurusan apa yang akan dipilih [1], [2]. Hal ini tidak semua siswa memiliki minat melanjutkan studi ke perguruan tinggi.

Kegiatan seleksi masuk perguruan tinggi dimaksudkan untuk memperoleh calon mahasiswa yang unggul. Kesempatan untuk mengikuti seleksi diberikan kepada seluruh siswa/siswi tamatan SMU/SMK, dan MA. Tes masuk mahasiswa di perguruan tinggi negeri terdiri atas tiga macam jalur seleksi, yaitu Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) [3], atau Jalur Undangan, Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN), dan Jalur Ujian Mandiri yang diadakan oleh Universitas negeri masing-masing.

PTN Jalur SNMPTN undangan merupakan jalur seleksi yang pertama diadakan oleh setiap perguruan tinggi negeri di seluruh Indonesia. Seleksi jalur ini mengundang siswa/siswi yang berprestasi disetiap Sekolah Menengah Atas (SMA), Sekolah Menengah Kejuruan (SMK) dan Madrasah Aliyah (MA) dengan mendaftar melalui sekolah masing-masing. Pada SNMPTN [4], jalur undangan sekolah sebagai satuan pendidikan diberikan wewenang melakukan seleksi awal calon mahasiswa yang berprestasi dan selanjutnya dilakukan seleksi dari perguruan tinggi. Akhmad Sudrajat dalam pengantar SNMPTN 2020 menjelaskan bahwa “penerimaan mahasiswa baru melalui SNMPTN Jalur Undangan merupakan kerja sama antara universitas dengan pihak sekolah”. Pihak sekolah memberikan data siswa yang berprestasi dan layak diterima menjadi mahasiswa SNMPTN [5] Jalur Undangan. Pendaftaran SNMPTN Undangan dilakukan di sekolah asal calon mahasiswa, syarat untuk calon mahasiswa.

PMDK dibuktikan dengan meningkatnya nilai raport dari semester 1 (satu) sampai dengan semester 5 (lima). Minat siswa SMA melanjutkan pendidikan ke perguruan tinggi tentu cukup beragam. Mungkin saja ada yang memiliki minat yang tinggi, minat yang sedang rendah atau bahkan sama sekali tidak berminat untuk melanjutkan pendidikan ke perguruan negeri [6].

Data mining merupakan serangkaian proses untuk mendapatkan informasi yang berguna dari data yang besar. Data mining juga dapat diartikan sebagai pergekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Dalam data mining terdapat

banyak jenis teknik, diantaranya dengan algoritma CART [7].

Permasalahan dalam penelitian ini berangkat dari penggunaan data mining dalam pendidikan yang dikenal dengan *Educational Data Mining* (EDM) yang dapat menimbulkan sejumlah masalah khusus yang harus diatasi untuk menghasilkan wawasan yang berharga dan membantu pengambilan keputusan pendidikan yang lebih baik. Data pendidikan seringkali tidak lengkap, tidak terstruktur, atau berisi. Data berkualitas buruk dapat membuat model yang tidak akurat dan informasi yang salah. Data yang dikumpulkan dalam konteks pendidikan biasanya melibatkan informasi pribadi tentang siswa, guru, dan staf sekolah. Menjaga kerahasiaan dan menghormati aturan etika dalam penggunaan data ini merupakan isu penting.

Data pendidikan bisa berukuran besar dengan banyak atribut atau karakteristik, seperti data perilaku, penilaian, keterlibatan siswa. Hal ini dapat mempersulit pemilihan fitur yang relevan dan mengurangi dimensi data. Data yang melacak kemajuan siswa dari waktu ke waktu (data vertikal) dapat sangat berharga dalam memberikan wawasan tentang perubahan perilaku dan kinerja. Namun, kurangnya data tersebut seringkali menjadi kendala. Interaksi dalam pendidikan bisa sangat kompleks, melibatkan siswa, guru, program, dan lingkungan belajar [8]. Menggambarkan dan menganalisis interaksi ini dalam model penambangan data bisa jadi rumit. Setiap siswa unik dan belajar secara berbeda. Model EDM harus mampu menangani variasi individual ini untuk memberikan informasi yang berguna. Salah satu tujuan EDM adalah untuk memprediksi perilaku atau kinerja siswa. Namun, prediksi dalam setting pendidikan bisa menjadi rumit karena banyak faktor yang mempengaruhi hasil [9].

Pemahaman tentang konteks pendidikan dan variabel terkait sangat penting untuk membentuk model EDM yang akurat dan bermakna. Keterbatasan pemahaman ini dapat membatasi hasil yang diperoleh. Penerapan solusi GED secara luas dalam lingkungan pendidikan seringkali melibatkan tantangan teknis dan organisasi, seperti integrasi sistem, pelatihan staf, dan koordinasi. Penting untuk diingat bahwa EDM adalah bidang yang berkembang dan tantangan ini dapat diatasi dengan mengembangkan metode analitis yang lebih baik, kolaborasi interdisipliner, dan perhatian yang cermat terhadap regulasi dan etika [10].

Terdapat penelitian sebelumnya terkait dengan kelulusan siswa. Lely Dwi Bhekti P., Wahyu W, pada penelitiannya dengan meningkatkan model Algoritma Classification and Regression Trees (CART) accuracy rate (1-APER) yang paling tinggi yaitu sebesar 92,3% untuk data learning dan 91,4% untuk data testing [11]. Nurul Indah P., Widodo, M.

Ficky Duskarnaen, pada penelitiannya dengan Algoritma Classification and Regression Tree CART yang menghasilkan nilai Hasil Algoritma CART rata-rata akurasi 80% dibanding algoritma lainnya [12]. Ali Fauzi, pada penelitian Pemanfaatan Classification And Regression Trees dalam Memprediksi Kelulusan Siswa Hasil Algoritma CART yang diujikan pada data dengan perbandingan 80% data training dan 20% data testing diperoleh hasil prediksi pada tingkat akurasi 75,9% [13].

Penelitian yang dilakukan Wulandari, Rika Rosnelly, Wanayumini. Analisis Metode Decision Tree Dalam Memprediksi Kelulusan Mahasiswa” dengan nilai akurasi algoritma Algoritma Decesion Tree Mengeluarkan keputusan kelulusan mahasiswa tersebut lulus tepat waktu atau tidak lulus tepat waktunya [14]. Penelitian oleh M. Fando Rizalno, A. Johar, Funny F. Coastera yang melakukan Analisis Prediksi Masa Studi Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritme Cart (Classification And Regression Trees)” Hasil dari penelitian ini adalah nilai parameter yang berpengaruh pada kelulusan mahasiswa dan dapat digunakan untuk proses prediksi [11]. Penelitian yang dilakukan oleh M. Fibo Donya Ikhbal, Denny Kurniadi dengan Menentukan Penjurusan Siswa Dengan Menggunakan Metode Decision Tree Algoritma C4.5 Hasil akurasi sebesar 68.42% dengan 304 sampel data [12].

Penelitian yang dilakukan oleh M. Rizki Qisthiano Klasifikasi Terhadap Prediksi Kelulusan Menggunakan Metode Support Vector Machine (SVM) Hasil Algoritma Support Vector Machine (SVM) memiliki hasil akurasi sebesar 85.06%, Penelitian yang dilakukan oleh Lingga Kurnia R., Bajeng Nurul W. Perbandingan Metode Klasifikasi Naïve Bayes dengan Support Vector Machine Pada Predikat Kelulusan menggunakan Algoritma Naïve Bayes dan Support Vector Machine Algoritma Naïve Bayes Memiliki akurasi lebih besar yaitu 96.52% dengan tingkat eror (Missclassification Rate) 0.03%, sedangkan metode Support Vector Machine memiliki akurasi 86.93%. Penelitian yang dilakukan oleh Ronny S., Wiratmoko Y., Elly Purwantini Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima Kartu Indonesia Pintar Kuliah Menggunakan Regresi Logistik Biner Hasil Algoritma Regresi Logistik Binerrerata recall sebesar 87,93%, accuracy sebesar 88,01%, precision sebesar 97,92%, dan AUC sebesar 84,6%.

Penelitian yang dilakukan oleh R. Thaniket, Kusrini, Emha Taufik L. Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma Support Vector Machin menggunakan algoritma Support Vector Machine (SVM) menghasilkan Accuracy 95,00%. Penelitian yang dilakukan oleh Saeful Bahri. Implementasi Data Mining Untuk Menentukan Jurusan siswa Pada Perguruan Tinggi Hasil Algoritma k-NN dan Naïve bayes dengan tingkat akurasi 75.38%, Penelitian yang dilakukan

oleh Gentur Wahyu N. W., Zaenal Arifin, M. A. Romli, Nurul I. Amal Prediksi Kelanjutan Studi Siswa Ke Perguruan Tinggi Dengan Naive Bayes Nilai akurasi pengklasifikasian dari data yang diujikan sehingga menghasilkan nilai prediksi 86.53%. Penelitian yang dilakukan oleh Ananda Fiqri F., Rohmat Saedudin, Rachmadita Andeswari Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Kelulusan Algoritma naive bayes memperoleh tingkat akurasi tertinggi sebesar 91.49%. Penelitian yang dilakukan oleh Azahari, Y., Dewi R., dan S. Mallala Komparasi Data Mining Naive Bayes Dan Neural Network dalam Memprediksi Masa Studi Mahasiswa S1 Prediksi naive bayes Menghasilkan akurasi hanya 57,63%. akurasi prediksi neural network adalah 72,58% [15].

Perbedaan Menggunakan CART (Classification and Regression Trees) dan SVM (Support Vector Machine) bergantung pada karakteristik data dan tujuan analisis yang ingin dicapai. Berikut adalah beberapa alasan mengapa harus menggunakan CART dan SVM. Model CART membuat pohon keputusan yang mudah dipahami dan ditafsirkan. Aturan keputusan yang dihasilkan oleh CART memberikan wawasan tentang faktor kunci yang mempengaruhi variabel target. CART dapat mengatasi kegagalan data karena partisi data didasarkan pada pemisahan variabel. Ini menawarkan peningkatan paling signifikan dalam pemisahan data. Kemampuan untuk menangani data non-linear. CART dapat dengan mudah menangani hubungan nonlinier antara variabel independen dan variabel target dengan melakukan pemisahan berbasis aturan berdasarkan pemisahan yang dilakukan pada setiap node [12].

SVM efektif untuk menyelesaikan masalah dimensi tinggi dimana dataset memiliki banyak fitur. SVM menggunakan fungsi kernel untuk memetakan data ke dimensi yang lebih tinggi di mana pemisahan kelas yang optimal lebih memungkinkan. SVM dapat mengatasi diskriminasi nonlinier dengan menggunakan fungsi kernel yang dapat mengubah data menjadi ruang fitur yang lebih kompleks. SVM mencoba menemukan margin maksimum antara kelas yang berbeda. Ini berarti bahwa SVM umumnya menawarkan batasan keputusan yang lebih baik dan secara umum bekerja lebih baik saat mengklasifikasikan data baru. Toleransi untuk data yang tidak seimbang [16].

SVM dapat memecahkan masalah ketidakseimbangan data dengan menetapkan bobot yang lebih tinggi untuk sampel yang lebih sedikit. Pilihan antara CART dan SVM bergantung pada karakteristik data, kompleksitas masalah, kemampuan interpretasi yang diinginkan, dan performa yang diharapkan. Jika menginginkan model yang mudah diinterpretasikan dan toleran terhadap varians, CART adalah pilihan yang baik. Jika berurusan dengan masalah diskriminasi dimensi tinggi atau nonlinier dan sedang mencari model yang berkinerja baik dalam mengklasifikasikan data, SVM

mungkin merupakan pilihan yang lebih baik [13], [16].

Metode CART adalah metode keputusan berbasis pohon yang dapat digunakan untuk tugas klasifikasi dan regresi. Metode CART membagi data menjadi himpunan bagian yang lebih kecil berdasarkan aturan pemisahan variabel independen, yang sangat meningkatkan pemisahan data. Pemisahan ini dilakukan secara iteratif hingga diperoleh subset yang seragam (untuk tugas klasifikasi) atau subset dengan karakteristik yang sama (untuk tugas regresi) di kelas target. Pohon keputusan yang akhirnya terbentuk dapat digunakan untuk mengklasifikasikan atau memprediksi nilai variabel target dalam data baru [12].

Metode SVM merupakan metode pemisahan kelas untuk menentukan klasifikasi. SVM mencoba menemukan hyperplane pembeda terbaik antara dua kelas dengan memaksimalkan jarak antara hyperplane dan sampel data terdekat untuk setiap kelas. Ketika keterpisahan linier tidak memungkinkan, SVM menggunakan fungsi kernel untuk memetakan data ke dimensi yang lebih tinggi di mana keterpisahan linier dimungkinkan. SVM juga dapat menangani data yang tidak seimbang dengan menetapkan bobot yang lebih tinggi untuk sampel yang lebih sedikit. CART membuat pohon keputusan yang mudah ditafsirkan, sementara SVM membuat hyperplane yang dipartisi dalam bentuk fungsi keputusan. CART menggunakan penugasan berbasis aturan yang mempertimbangkan penugasan inkremental di setiap node, sementara SVM berusaha untuk penugasan yang optimal dengan memaksimalkan margin antar kelas [11], [16].

Mesin vektor pendukung (SVM) dan pohon klasifikasi dan regresi adalah dua jenis algoritma pembelajaran biasa digunakan untuk masalah klasifikasi dan regresi. Keduanya memiliki fitur dan manfaat yang berbeda, yang mungkin membuatnya cocok untuk situasi tertentu. SVM dapat bekerja secara efisien dalam ruang fitur multidimensi, meskipun jumlah sampel terbatas. Memang, konsep hyperplane memisahkan lapisan data dengan jarak maksimum. Tujuan utama SVM adalah menemukan hyperplane dengan margin terbesar antar kelas data. Ini berarti bahwa SVM cenderung menghasilkan model yang sering ketinggalan zaman. SVM juga dapat memanfaatkan trik perkalian, memungkinkan data dipetakan ke dimensi yang lebih tinggi tanpa menghitung ruang fitur secara eksplisit. Ini membantu SVM untuk menangani masalah yang lebih kompleks dan bahkan data diskrit non-linear. Tentu saja, SVM cenderung kurang rentan terhadap outlier, karena berfokus pada menemukan hyperplane dengan amplitudo maksimum [16].

Pohon Klasifikasi dan Regresi Pohon keputusan adalah model yang mudah dijelaskan kepada manusia. Setiap cabang mewakili keputusan berdasarkan fitur tertentu, dan setiap daun mewakili label klasifikasi atau nilai regresi. Pohon keputusan

dapat menangani data non-linier dan tidak memerlukan asumsi tertentu tentang distribusi data. Ini membuatnya cocok untuk data dengan hubungan yang kompleks atau tidak terstruktur. Pohon keputusan dapat dengan mudah memproses data dengan kombinasi fitur kategori dan numerik tanpa memerlukan pemrosesan awal yang rumit. Pohon keputusan juga digunakan sebagai dasar untuk algoritma himpunan populer seperti Random Forest dan Gradient Boosting. Ini memungkinkan penggabungan beberapa pohon keputusan untuk meningkatkan akurasi dan mengurangi redundansi [11].

Memilih antara Metode SVM dan Menggunakan Metode CART bergantung pada karakteristik data, tujuan analisis, dan preferensi interpretabilitas. memprioritaskan interpretabilitas dan memiliki data yang kompleks, CART mungkin merupakan pilihan yang lebih baik. Namun, jika berurusan dengan data spesial yang sangat fungsional atau mencari pemisahan lapisan maksimum, maka SVM mungkin merupakan pilihan yang lebih baik.

Berdasarkan latar belakang yang telah diuraikan, makalah ini membahas masalah dalam penelitian ini yaitu: Algoritma manakah yang terbaik antara Cart dan SVM? Dengan tujuan penelitian untuk mengetahui klasifikasi algoritma CART dan SVM dalam menentukan siswa berprestasi masuk PTN dengan Komparasi algoritma klasifikasi dalam penentuan diterima siswa didik MAN 1 Kota Cirebon di PTN. Adapun kontribusi dalam penelitian ini adalah: komparasi Algoritma Klasifikasi Dalam Penentuan Diterima Siswa Didik Madrasah Aliyah Negeri (MAN) 1 Kota Cirebon Di Perguruan Tinggi Negeri.

2. METODE PENELITIAN

Pada bagian ini membahas metode atau langkah-langkah yang dilakukan penulis untuk mengumpulkan data dan informasi dalam pemilihan atribut dan mencari nilai Performace Klasifikasi pada algoritma Klasifikasi terbaik dalam penentuan masuk PTN Jalur Prestasi. Langkah-langkah yang dilakukan penulis sebagai berikut:

2.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini terdiri dari:

2.1.1. Data Primer

Data primer yaitu data yang diperoleh dari lapangan atau studi kasusnya [17]. Berikut data Peserta didik MAN 1 Kota Cirebon kurun waktu 3 tahun terakhir Data primer yaitu data yang didapatkan dari lapangan atau studi kasusnya. MAN 1 Kota Cirebon pada waktu 3 tahun terakhir telah meluluskan 297 siswa. Pada tahun 2020 meluluskan 149 siswa dengan prosentasi siswa yang masuk PTN (Perguruan Tinggi Negeri) 30 %. Pada tahun 2021 meluluskan

140 siswa dengan prosentasi siswa yang masuk Perguruan Tinggi Negeri 25 %, dan Pada tahun 2020 meluluskan 247 siswa dengan prosentasi siswa yang masuk Perguruan Tinggi Negeri 30 %. Pada penelitian ini dataset yang akan digunakan merupakan data siswa kelas XII yang didapatkan dari bagian kesiswaan di MAN 1 Kota Cirebon Tahun 2022/2023 (yang akan lulus tahun 2023).

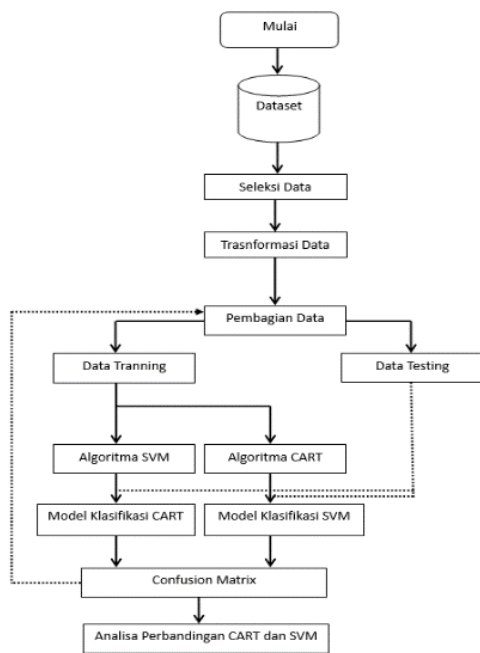
Sedangkan Pada penelitian ini dataset yang akan digunakan merupakan data peserta didik yang didapatkan dari bagian TU dan waka kesiswaan pada MAN 1 kota Cirebon angkatan tahun 2020, 2021, 2022 dan 2023. (Sumber: TU dan Waka Kurikulum MAN 1 Kota Cirebon).

2.1.2. Data Sekunder

Data sekunder yaitu data yang diperoleh dari luar penelitian yang gunanya untuk pendukung data primer [18]. Pada penelitian ini penulis mengumpulkan data sekunder dengan cara studi pustaka dan mengumpulkan referensi-referensi seperti buku, jurnal, tesis dan alamat website yang berhubungan dengan data mining, klasifikasi, algoritma, dan metode yang digunakan.

2.2. Metode Yang Diusulkan

Penelitian ini akan melalui beberapa tahap atau metode, dimana setiap tahap saling berdekatan dan saling mempengaruhi satu sama lain. Metode penelitian yang diusulkan dalam penelitian ini dapat dilihat pada Gambar 1 berikut:



Gambar 1. Tahap penelitian

Dimulai dari dataset akan dilakukan seleksi data dari data 143 siswa dan 7 atribut seleksi sejumlah preprocessing menggunakan data transformation terlebih dahulu. Adapun tujuan menggunakan data transformation agar mempermudah proses Data

Training nilai siswa MAN 1 kota Cirebon Preprocessing Prediksi Hasil Klasifikasi Akurasi Data Testing 143 siswa di implementasi pada program serta proses perhitungan yang dihasilkan akan lebih efisien. Setelah melalui tahap preprocessing, data tersebut dibagi menjadi data training dan data testing menggunakan Cross Validation sebanyak 10-fold cross validation [9]. Selanjutnya untuk proses klasifikasi akan digunakan perbandingan dua metode CART dengan SVM, yaitu untuk metode pertama menggunakan CART [13], metode kedua menggunakan SVM dengan menambahkan pembobotan Gain Ratio. Untuk pengukuran performansi akan dibandingkan tingkat akurasi antara CART yang menggunakan pembobotan Gain Ratio dengan SVM tanpa menggunakan pembobotan Gain Ratio menggunakan Confusion Matrix [16].

2.2.1. Dataset

Dataset penelitian ini menggunakan dataset data nilai siswa MAN 1 Kota Cirebon tahun 2020, 2021 dan 2022 dari semester 1 sampai dengan semester 5. Terdiri dari 7 Atribut Jenis Kelamin, Rata-rata semester 1, Rata-rata semester 2, Rata-rata semester 3, Rata-rata semester 4, Rata-rata semester 5, Predikat Masuk/Tidak.

2.2.2. Transformasi data

Transformasi Data adalah upaya yang dilakukan dengan tujuan utama untuk mengubah skala pengukuran data asli menjadi bentuk lain sehingga data dapat memenuhi asumsi-asumsi yang mendasari analisis ragam [19].

2.2.3. Seleksi Data

Seleksi Data merupakan proses meminimalkan jumlah data yang digunakan untuk proses mining dengan tetap merepresentasikan data aslinya [20] berupa data nilai siswa MAN 1 kota Cirebon.

2.2.4. Pembagian Data

Setelah melalui tahap preprocessing, tahap selanjutnya yaitu membagi data tersebut menjadi dua bagian, yaitu data training dan data testing. Dimana data training dan data testing akan dibagi menggunakan metode Cross Validation [20]. Data tersebut dibagi sebanyak 10-fold cross validation karena 10 adalah jumlah yang tepat untuk mendapatkan estimasi yang terbaik. Setiap k-fold akan mendapatkan jumlah data yang sama. Berikut pembagian data training dan data testing menggunakan 10- fold cross validation.

2.2.5. Algoritma CART

CART merupakan salah satu metode atau algoritma yang digunakan untuk mengeksplorasi data dengan penggambaran pohon keputusan.. Tujuan

utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penjelasan lebih lanjut dari suatu pengklasifikasian [11]. Keluaran dari algoritma ini berupa pohon klasifikasi jika variabel targetnya adalah data kategorik. Sedangkan jika variabel targetnya adalah data numerik atau kontinu maka keluarannya berupa pohon regres Algoritma. Rumus 1:

$$i(t) = \sum_{j \neq k} p(j | t)p(k | t) \quad (1)$$

dengan $p(j | t)$ adalah proporsi kelas j pada simpul t dan $p(k | t)$ adalah proporsi kelas k pada simpul t . Pengevaluasian pemilah s pada simpul t dilakukan berdasarkan kriteria *goodness of split* ($\Phi(s, t)$) sesuai dengan persamaan [12], [13].

$$\Phi(s, t) = \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2)$$

dengan $i(t_L)$ adalah *impurity* pada simpul kiri dan $i(t_R)$ adalah *impurity* pada simpul kanan dimana pemilah yang menghasilkan *goodness of split* tertinggi merupakan pemilah yang terbaik [11].

2.2.6. Algoritma SVM

Algoritma SVM yang berupa klasifikasi dengan cara membagi data menjadi dua kelas menggunakan garis vektor yang disebut hyperplane [16]. Pada permasalahan yang kompleks atau permasalahan dengan parameter yang banyak, metode ini sangat baik untuk digunakan.

Rumus 1:

Rumus Support Vector Machine (SVM) terdiri dari rumus dasar untuk memprediksi label kelas dan rumus untuk menentukan margin maksimum:

Rumus dasar SVM: $y(x) = \text{tanda}(w \cdot x + b)$

Dalam rumus ini, $y(x)$ adalah header prediksi untuk sampel x , w adalah vektor bobot, x adalah vektor fitur, b adalah bias, dan $\text{tanda}()$ adalah fungsi bertanda tangan yang mengembalikan -1 atau 1 bergantung pada apakah argumennya kurang dari atau kurang dari 0.

Rumus untuk margin maksimum:

Margin maksimum SVM dapat dihitung dengan menggunakan rumus: $\text{Margin} = 2/\|w\|$

Dalam rumus ini $\|w\|$ adalah norma Euclidean dari vektor bobot w .

2.2.7. Model Klasifikasi CART

Model Klasifikasi CART adalah sebuah algoritma pembelajaran mesin yang digunakan untuk membangun pohon keputusan untuk melakukan klasifikasi pada data [13]. CART dapat digunakan baik untuk masalah klasifikasi (memprediksi kelas diskrit) maupun masalah regresi (memprediksi nilai kontinu).

Penggunaan Performace Klasifikasi dalam pemilihan pemilah memiliki kelebihan yaitu proses perhitungan yang sederhana dan relative cepat, serta mudah dan sesuai untuk diterapkan dalam berbagai

kasus. Pemilah terbaik dipilih berdasarkan nilai penuruanan rata-rata pesemester yang paling tinggi dari semua kemungkinan pemilahan yang dilakukan. Fungsi performace klasifikasi dengan metode CART [3] dan SVM [4] dituliskan dalam persamaan.

2.2.8. Model Klasifikasi SVM

Model klasifikasi SVM (Support Vector Machine) adalah sebuah metode pembelajaran mesin yang digunakan untuk melakukan klasifikasi pada data [16]. SVM dapat digunakan untuk masalah klasifikasi biner (dua kelas) maupun multikelas.

2.2.9. Confusion Matrix

Confusion matrix adalah sebuah tabel yang digunakan untuk mengevaluasi performa model klasifikasi dengan membandingkan prediksi yang dihasilkan oleh model dengan nilai sebenarnya dari data yang diamati [21]. Confusion matrix umumnya digunakan untuk masalah klasifikasi biner, di mana terdapat dua kelas yang akan diprediksi : positif dan negatif.

2.2.10. T-Test

Adalah uji statistik yang digunakan untuk membandingkan rata-rata dua kelompok independen [22]. Dapat menggunakan uji-t di RapidMiner sebagai berikut:

1. Impor data ke RapidMiner menggunakan operator baca atau melalui tautan ke sumber data eksternal.
2. Pastikan data terstruktur dengan benar dan telah memilih fungsi yang digunakan dalam uji-t.

2.3. Atribut Penelitian

Atribut atau atribut penelitian yang digunakan yaitu data akademik siswa MAN 1 Kota Cirebon angkatan tahun 2020, 2021 dan 2022 (yang akan masuk PTN tahun 2023). Berikut Tabel 1 tentang variabel yang digunakan dalam penelitian ini.

Tabel 1. Atribut Penelitian

No	Variabel	Keterangan
1	Jenis Kelamin	Laki-laki atau Perempuan
2	Rata-rata semester 1	Indeks Rata-rata Semester 1
3	Rata-rata semester 2	Indeks Rata-rata Semester 2
4	Rata-rata semester 3	Indeks Rata-rata Semester 3
5	Rata-rata semester 4	Indeks Rata-rata Semester 4
6	Rata-rata semester 5	Indeks Rata-rata Semester 5
7	Predikat	Masuk SMPTN/PTN atau Tidak Masuk SNMPTN/PTN

2.4. Jenis Penelitian

Jenis penelitian ini adalah penelitian *experimental research* yaitu penelitian yang dilakukan dengan cara menguji kebenaran sebuah hipotesis dengan statistik yang melibatkan beberapa variabel dengan menggunakan tes tertentu dan menghubungkannya dengan masalah penelitian [23].

2.5. Instrumen Penelitian

Instrumen yang digunakan pada penelitian ini adalah:

2.5.1. Perangkat Keras (Hardware)

Untuk peralatan perangkat keras yang digunakan dalam pelaksanaan penelitian ini adalah komputer atau laptop yang memiliki spesifikasi sebagai berikut:

- a. Intel Core Processor Core i3
- b. Ram 8 GB
- c. Windows 10 64-bit OS, X64 based processor.

2.5.2. Perangkat Lunak (Software)

Software yangl dibutuhkan pada komparasi algoritma prediksi untuk prediksi mahasiswa yang berpotensi *cumlaude* yaitu windows 8 sebagai sistem operasi [24], microsoft excel 2010 sebagai pengolahan data dan aplikasi *RapidMiner* sebagai penjabaran proses data mining.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

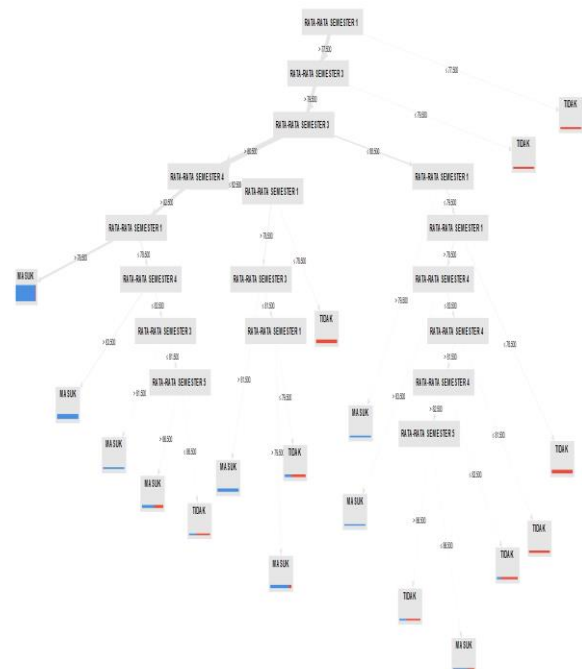
3.1.1. Penentuan Dataset

Pengambilan Data nilai siswa MAN 1 kota Cirebon dari Waka Kurikulum dari tahun 2021,2022 dan 2023 Semester 1 sampai dengan Semester 5, Salah satu faktor yang mempengaruhi nilai akreditasi adalah jumlah lulusan peserta didik masuk PTN jalur prestasi. Oleh sebab itu, jumlah lulusan siswa mempunyai peranan penting dalam menarik animo calon siswa baru. MAN 1 Kota Cirebon merupakan MAN 1 yang berada di Kota Cirebon yang memiliki akreditasi A untuk periode tahun 2021-2024. Salah satu permasalahan yang muncul adalah jumlah siswa yang masuk di terima di (PTN) perguruan tinggi negeri masih sedikit dan jumlah siswa nya yang lulus tidak sesuai dengan yang di diharapkan. Jika permasalahan tersebut tidak diatasi dapat mengakibatkan permasalahan di masa yang akan datang yaitu menurunnya nilai akreditasi dan penilaian dari masyarakat akan mempengaruhi jumlah siswa yang akan masuk ke MAN 1 Kota Cirebon. Untuk bisa siswa banyak yang masuk ke perguruan tinggi negeri. Salah satu solusi yang peneliti tawarkan dengan menggunakan Klasifikasi data mining.

Dalam penelitian ini akan di klasifikasi masuk PTN dengan metode *Cart* dan SVM. Dataset yang digunakan adalah data primer dari waka kesiswaan MAN 1 Kota Cirebon berjumlah 143 siswa kemudian di datanya dibersihkan menjadi 143 siswa dengan menggunakan atribut data nilai siswa 8 Atribut seperti Jenis Kelamin, Rata-rata semester 1, Rata-rata Semester 2, Rata-rata semester 3, Rata-rata semester 4, Rata-rata semester 5, Jumlah Rata-rata Semester 1 -5,

Predikat Keterangan masuk PTN dan tidak masuk PTN.

Pada eksperimen yang pertama peneliti melakukan uji coba perbandingan Cross Validation dan Performance Klasifikasi dan menggunakan algoritma CART dan SVM dengan jumlah data 143 siswa dengan atribut 8 dan perbandingan antara data training testing Masuk pengisian data peserta PTN masuk atau tidak masuk, Hasil Eksperimen perbandingan menggunakan algoritma CART mendapatkan akurasi 86.10%. Menggunakan algoritma SVM [4] yang menghasilkan akurasi 86.71%. Gambar 2, Gambar 3, dan Gambar 4 tentang Graph CART Decision:



Gambar 2. Graph CART (Decision Tree)

Name	Type	Waring	Statistik	Nilai	File (100 - 10 attributes)	Support (100 rows)
JENIS KELAMIN	Ername	C	0	0	0 (100%) 1 (0%)	
PREDIKAT	Ername	C	MASUK	TEKAK	MASUK (64) TEKAK (45)	
prediction(PREDIKAT)	Ername	C	MASUK	TEKAK	MASUK (65) TEKAK (47)	
confidence(MASUK)	Rial	C	0	1	0.947	
confidence(TEKAK)	Rial	C	0	1	0.333	
RATA-RATA SEMESTER 1	Integer	C	77	82	70.800	
RATA-RATA SEMESTER 2	Integer	C	77	82	70.128	
RATA-RATA SEMESTER 3	Integer	C	79	83	81.032	
RATA-RATA SEMESTER 4	Integer	C	81	85	82.748	

Gambar 3. Support Vector Machine (SVM)

```

RATA-RATA SEMESTER 1 > 77.500
| RATA-RATA SEMESTER 3 > 79.500
| | RATA-RATA SEMESTER 3 > 80.500
| | | RATA-RATA SEMESTER 4 > 82.500
| | | | RATA-RATA SEMESTER 1 > 78.500: MASUK (MASUK=47, TIDAK=1)
| | | | RATA-RATA SEMESTER 1 ≤ 78.500
| | | | | RATA-RATA SEMESTER 4 > 83.500: MASUK (MASUK=13, TIDAK=0)
| | | | | RATA-RATA SEMESTER 4 ≤ 83.500
| | | | | | RATA-RATA SEMESTER 3 > 81.500: MASUK (MASUK=3, TIDAK=0)
| | | | | | RATA-RATA SEMESTER 3 ≤ 81.500
| | | | | | | RATA-RATA SEMESTER 5 > 86.500: MASUK (MASUK=4, TIDAK=3)
| | | | | | | RATA-RATA SEMESTER 5 ≤ 86.500: TIDAK (MASUK=1, TIDAK=2)
| | | | RATA-RATA SEMESTER 4 ≤ 82.500
| | | | RATA-RATA SEMESTER 1 > 78.500
| | | | | RATA-RATA SEMESTER 3 > 81.500: MASUK (MASUK=8, TIDAK=0)
| | | | | RATA-RATA SEMESTER 3 ≤ 81.500
| | | | | | RATA-RATA SEMESTER 1 > 79.500: MASUK (MASUK=7, TIDAK=1)
| | | | | | RATA-RATA SEMESTER 1 ≤ 79.500: TIDAK (MASUK=2, TIDAK=4)
| | | | | | | RATA-RATA SEMESTER 1 ≤ 78.500: TIDAK (MASUK=0, TIDAK=9)
| | | RATA-RATA SEMESTER 3 ≤ 80.500
| | | | RATA-RATA SEMESTER 1 > 79.500: MASUK (MASUK=3, TIDAK=0)
| | | | RATA-RATA SEMESTER 1 ≤ 79.500
| | | | | RATA-RATA SEMESTER 1 > 78.500
| | | | | | RATA-RATA SEMESTER 4 > 83.500: MASUK (MASUK=2, TIDAK=0)
| | | | | | RATA-RATA SEMESTER 4 ≤ 83.500
| | | | | | | RATA-RATA SEMESTER 4 > 81.500
| | | | | | | | RATA-RATA SEMESTER 4 > 82.500
| | | | | | | | | RATA-RATA SEMESTER 5 > 86.500: TIDAK (MASUK=1, TIDAK=2)
| | | | | | | | | RATA-RATA SEMESTER 5 ≤ 86.500: MASUK (MASUK=2, TIDAK=1)
| | | | | | | | | | RATA-RATA SEMESTER 4 ≤ 82.500: TIDAK (MASUK=1, TIDAK=4)
| | | | | | | | | | RATA-RATA SEMESTER 4 ≤ 81.500: TIDAK (MASUK=0, TIDAK=5)
| | | | | | | RATA-RATA SEMESTER 1 ≤ 78.500: TIDAK (MASUK=0, TIDAK=9)
| RATA-RATA SEMESTER 3 ≤ 79.500: TIDAK (MASUK=0, TIDAK=4)
RATA-RATA SEMESTER 1 ≤ 77.500: TIDAK (MASUK=0, TIDAK=4)
    
```

Gambar 4. Penjelasan CART (Decision Tree)

Hasil dari Klasifikasi dengan menggunakan SVM sehingga menghasilkan bobot atribut sesuai dengan Tabel 2 bahwa dari 8 atribut yang digunakan atribut yang paling berpengaruh adalah Jenis Kelamin, Rata-rata Semester 1, Rata-rata Semester 2, Rata-rata Semester 3, Rata-rata Semester 4, Rata-rata Semester 5. Berdasarkan Klasifikasi menggunakan Performace Klasifikasi dengan algoritma Klasifikasi SVM faktor atau atribut yang paling dominan mempengaruhi Siswa masuk PTN sesuai dengan minat siswa dan nilai siswa adalah Predikat siswa apakah siswa tersebut bisa masuk PTN atau status hanya siswa tidak masuk PTN

3.1.2. Pengumpulan Data

Pada penelitian ini teknik pengumpulan data yang digunakan adalah studi literatur melalui buku dan jurnal. Sedangkan data yang digunakan untuk pengolahan algoritma dengan menggunakan data primer yang diambil dari data akademik siswa MAN 1 Kota Cirebon angkatan tahun 2020, 2021, 2022. Dari pengumpulan dataset diperoleh jumlah data 143 dari waka kurikulum.

3.1.3. Pengolahan Data Awal (Pre Processing)

Setelah dataset diperoleh kemudian data tersebut dibersihkan dari data yang *missing value* dan data yang *redundan*. Untuk membersihkan data peneliti menggunakan software *microsoft excel*. Hasil dari data *Klasifikasi performace* diperoleh data sejumlah 143 dari data tersebut tahap selanjutnya akan diolah dengan menggunakan *RapidMiner*.

3.1.4. Pemodelan Algoritma CART dan SVM

Dataset yang digunakan dalam penelitian ini adalah data primer dari MAN 1 Kota Cirebon yang berjumlah 143 orang kemudian di datanya dibersihkan menjadi 143 orang dengan menggunakan 8 atribut yaitu : Jenis Kelamin, Rata-rata Semester 1, Rata-rata Semester 2, Rata-rata Semester 3, Rata-rata Semester 4, Rata-rata Semester 5, Predikat masuk PTN dan tidak Masuk PTN. Jika dibuat kategorialnya sebagai berikut:

Tabel 2. Kategorikal

Atribut	Variabel
Jenis Kelamin	X1
Rata-Rata Semester 1	X2
Rata-Rata Semester 2	X3
Rata-Rata Semester 3	X4
Rata-Rata Semester 4	X5
Rata-Rata Semester 5	X6
Predikat	X8

Pengkonversian semua teks dan kategori menggunakan angka agar dapat diolah di *rapidminer*. Berikut pengkonversian pada atribut:

Pengkonversian Jenis Kelamin:

Laki-Laki = 1

Perempuan = 0

Mendefinisikan perbedaan (pembagian) variabel:

Pilih salah satu variabel prediktor sebagai variabel perbedaan untuk membangun pohon keputusan. Pilih variabel yang memiliki pengaruh signifikan terhadap variabel atau target yang diinginkan. Perhitungan jumlah polusi: Gunakan pengukuran ketidak murnian seperti indeks Gini atau entropi untuk memperkirakan ketidak murnian atau ketidak sempurnaan pada setiap node.

Rumus Indeks Gini : $Gini(p) = 1 - \sum(\pi_i^2)$, di mana π_i adalah pecahan dari setiap kelas di node. rumus entropi: $Entropi(p) = -\sum(\pi_i * \log_2(\pi_i))$, di mana π_i adalah pecahan dari setiap kelas di node. Pertukaran informasi : Gunakan variabel pemisahan yang telah ditentukan sebelumnya untuk membagi data menurut kriteria yang sesuai. Misalnya, jika variabel pemisahannya adalah GENDER, data dibagi menjadi dua himpunan bagian: Satu untuk jenis kelamin P dan satu untuk jenis kelamin L.

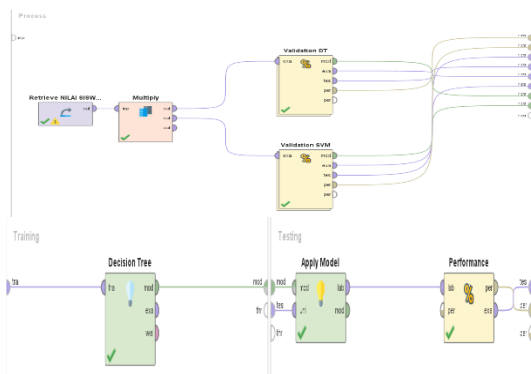
Perhitungan tingkat pengotor di setiap subset: Hitung ukuran (indeks Gini atau entropi) di setiap subset yang dihasilkan dari partisi data. Gunakan rumus yang sama seperti pada langkah 2 untuk menghitung jumlah dalam himpunan. Untuk

menemukan variabel perbedaan terbaik, Bandingkan pengukuran di setiap subset dan tentukan yang terbaik. Semakin banyak jumlah yang berkurang semakin baik variabel. Struktur pohon keputusan:

Lakukan langkah-langkah di atas secara rekursif untuk setiap subset yang dihasilkan dari pemisahan data. Bangun pohon keputusan dengan menambahkan node baru dan hubungannya dengan node sebelumnya. Untuk menentukan kondisi berhenti. Tentukan kondisi penghentian untuk membatasi pembentukan pohon keputusan, mis. B. mencapai tingkat kedalaman maksimum, jumlah minimum sampel di node atau kontaminasi di bawah ambang batas tertentu. Prediksi dan Evaluasi: Gunakan pohon keputusan yang dihasilkan untuk membuat prediksi dari data baru. Evaluasi kinerja model menggunakan metrik yang sesuai seperti presisi, akurasi, daya ingat, atau MSE (mean squared error) dalam kasus regresi.

3.1.5. Pemodelan Data Mining dengan Menggunakan Algoritma CART dan SVM

Pada pemodelan awal dengan menggunakan algoritma CART dan SVM dikarenakan dataset bertipe data nominal dan numerik maka dataset yang akan diolah dengan algoritma CART dan SVM harus dirubah ke *numerik*. Untuk pola pemodelan dengan *RapidMiner* dapat dilihat pada Gambar 5.



Gambar 5. Pemodelan data mining pada algoritma CART

Dataset yang ada harus dirubah ke numerik kemudian pembagian data training dan data testing sehingga dihasilkan akurasi 86.10% Gambar 6 dan Gambar 7.

accuracy 86.10% +/- 1.85% (micro average: 86.10%)			
	true MASUK	true TIDAK	class precision
pred. MASUK	85	11	88.54%
pred. TIDAK	9	38	80.85%
class recall	90.43%	77.55%	

Gambar 6. Hasil Akurasi Algoritma CART

accuracy 86.71% +/- 1.11% (micro average: 86.71%)			
	true MASUK	true TIDAK	class precision
pred. MASUK	85	11	89.07%
pred. TIDAK	9	38	81.25%
class recall	90.43%	77.55%	

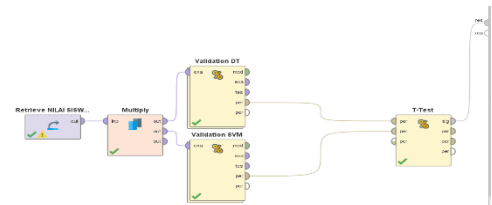
Gambar 7. Hasil Akurasi Algoritma SVM

3.1.6. Evaluasi Pemodelan Data Mining dengan Algoritma CART dan SVM

Berdasarkan hasil eksperimen yang dilakukan menggunakan algoritma Klasifikasi CART dengan dataset 143 dengan 8 atribut menghasilkan akurasi 86.10%. dibandingkan dengan algoritma SVM menghasilkan akurasi 86.71% untuk prediksi keberhasilan masuk PTN.

3.1.7. Pemodelan Algoritma Klasifikasi CART dan SVM dengan T-Test

Setelah melakukan Klasifikasi menggunakan algoritma CART dan SVM yang menghasilkan akurasi CART 86.10% dan akurasi SVM 86.71%. kemudian tahap berikutnya adalah menambahkan metode T-Test mencari hasil perbandingan antara metode CART dengan SVM klasifikasi keberhasilan masuk PTN.



Gambar 8. Pemodelan Algoritma CART (Decision Tree) dan SVM dengan menambahkan metode T-Test

Gambar 8 menggunakan Model T-Test dengan algoritma CART menghasilkan 0.861+/-0.071 algoritma SVM menghasilkan nilai 0.867+/-0.051 untuk perbedaannya 0.826.

3.1.8. Evaluasi Klasifikasi CART dan SVM dengan T-Test

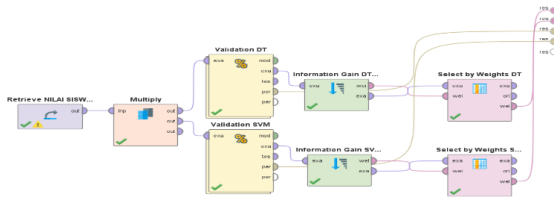
Berdasarkan hasil eksperimen yang dilakukan menggunakan algoritma Klasifikasi CART dan SVM dengan Model T-Test menghasilkan nilai CART 0.861+/-0.071 algoritma SVM menghasilkan nilai 0.867+/-0.051 untuk perbedaannya 0.826..

3.1.9. Pemodelan Algoritma Klasifikasi CART dan SVM dengan Informasi Gain

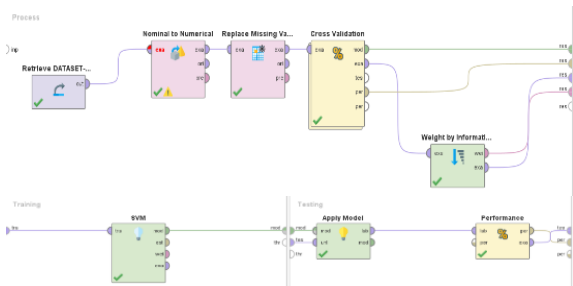
Tabel 3. Hasil Pemodelan Algoritma CART

	true masuk	true tidak	Class precision
pred. Masuk	85	11	88.54%
pred. Tidak	9	38	80.85%
Class recall	90.43%	77.55%	

Tabel 3 menjelaskan bahwa setelah melakukan Klasifikasi menggunakan algoritma CART dan SVM yang menghasilkan akurasi CART 86.10% dan akurasi SVM 86.71%. kemudian tahap berikutnya adalah menambahkan metode Informasi gain mencari bobot perbandingan antara metode CART dengan SVM klasifikasi keberhasilan masuk PTN, terlihat juga di Gambar 9, dan Gambar 10.



Gambar 9. Pemodelan Algoritma CART (Decision Tree) dan SVM Dengan Metode Seleksi Atribut Information Gain dengan Rapidminer



Gambar 10. Pemodelan Algoritma SVM

3.1.10. Evaluasi dan Validasi Hasil Pemodelan Algoritma CART dan SVM

Data yang sudah diolah menggunakan algoritma CART dan SVM melalui tool RapidMiner. Selanjutnya, di evaluasi dan validasi hasil menggunakan confusion matrix sebagai performance yang akan menghasilkan nilai Accuracy, Precision, Recall.

Berdasarkan hasil performance vector pada Tabel 5 diatas, maka dapat disimpulkan data yang diprediksi dengan benar melalui algoritma CART sejumlah tabel 143, sebanyak 85 siswa benar prediksi klasifikasi masuk PTN dan benar tidak sesuai masuk PTN 11 siswa. Kemudian sebanyak 9 orang diprediksi tidak masuk PTN dan sebanyak 38 orang diprediksi benar tidak masuk. Sehingga dapat disimpulkan siswa yang masuk sejumlah 85 siswa yang masuk PTN atau masuk tidak sesuai sejumlah 11 siswa . Untuk perhitungan akurasi sebagai berikut:

a. Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

$$Accuracy = \frac{38 + 85}{38 + 85 + 11 + 9} \times 100\%$$

$$Accuracy = 86.10\%$$

Berdasarkan pemodelan algoritma CART dengan metode seleksi atribut information gain dengan RapidMiner didapatkan nilai presentase akurasi 86.10 % sesuai dengan perhitungan manual akurasi. Kemudian untuk perhitungan manual Precision sebagai berikut:

b. Precision

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

$$Precision = \frac{38}{38 + 11} \times 100\%$$

$$Precision = 77.55\%$$

Berdasarkan pemodelan algoritma CART dengan metode seleksi atribut information gain dengan RapidMiner didapatkan nilai presentase precision 77.55% sesuai dengan perhitungan manual precision. Kemudian untuk perhitungan manual Recall sebagai berikut:

c. Precision

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

$$Recall = \frac{39}{38 + 9} \times 100\%$$

$$Recall = 80.85\%$$

Berdasarkan pemodelan algoritma neural network di dalam Gambar 8 dan Tabel 6 dengan metode seleksi atribut information gain dengan RapidMiner didapatkan nilai presentase recall 80.85% sesuai dengan perhitungan manual recall.

Hasil penelitian ini menghasilkan algoritma klasifikasi terbaik antara CART dan SVM, Decision Tree Kemudian setelah diketahui algoritma klasifikasi terbaik Akurasi dari masing-masing metode menggunakan algoritma klasifikasi CART dan SVM . Hasil uji coba menunjukkan penerapan teknik feature selection dihasilkan akurasi CART 86.10% dibanding dengan metode SVM penelitian yang dilakukan oleh peneliti dengan atribut yang sama dan klasifikasi yang sama tentang masuk PTN akurasi 86.71%.

3.1.11. Hasil Klasifikasi

Hasil Performace Vector klasifikasi siswa masuk PTN menggunakan algoritma CART dan SVM dengan confusion matrix dengan data tranning dan testing sejumlah 143 menghasilkan klasifikasi masuk seperti Gambar 11, Gambar 12, Gambar 13, Gambar 14.

PerformanceVector

```

PerformanceVector:
accuracy: 86.10% +/- 7.15% (micro average: 86.01%)
ConfusionMatrix:
True:  MASUK  TIDAK
MASUK:  85    11
TIDAK:   9    38
precision: 83.64% +/- 15.47% (micro average: 80.85%) (positive class: TIDAK)
ConfusionMatrix:
True:  MASUK  TIDAK
MASUK:  85    11
TIDAK:   9    38
recall: 77.50% +/- 18.41% (micro average: 77.55%) (positive class: TIDAK)
ConfusionMatrix:
True:  MASUK  TIDAK
MASUK:  85    11
TIDAK:   9    38
AUC (optimistic): 0.952 +/- 0.049 (micro average: 0.952) (positive class: TIDAK)
AUC: 0.891 +/- 0.088 (micro average: 0.891) (positive class: TIDAK)
AUC (pessimistic): 0.837 +/- 0.137 (micro average: 0.837) (positive class: TIDAK)
    
```

Gambar 11. Hasil Performance Confusion Matrix menggunakan algoritma CART

Row No.	JENIS KEJL.	PREDIKAT	predictionP	confidence	confidence	RATA-RATA	RATA-RATA	RATA-RATA	RATA-RATA	RATA-RATA
1	0	TEKAK	TEKAK	0.010	0.982	79	79	79	81	87
2	1	MASUK	MASUK	0.721	0.289	79	79	81	83	86
3	1	TEKAK	MASUK	0.651	0.349	80	80	81	81	86
4	0	TEKAK	TEKAK	0.289	0.721	79	79	80	82	87
5	0	MASUK	MASUK	0.881	0.119	79	79	83	82	87
6	0	MASUK	MASUK	0.798	0.202	78	78	82	84	88
7	0	TEKAK	TEKAK	0.165	0.835	77	78	81	83	87
8	0	MASUK	MASUK	0.983	0.017	80	81	81	83	86
9	1	MASUK	MASUK	0.721	0.289	79	79	81	83	87
10	1	TEKAK	TEKAK	0.119	0.881	78	79	80	82	86
11	1	MASUK	MASUK	0.721	0.289	79	79	81	83	87
12	1	MASUK	MASUK	0.881	0.119	79	79	81	84	88
13	0	MASUK	TEKAK	0.500	0.500	78	79	81	83	87
14	0	MASUK	MASUK	0.721	0.289	79	79	82	82	87
15	0	MASUK	MASUK	0.721	0.289	79	79	81	83	87
16	0	TEKAK	TEKAK	0.047	0.953	78	79	80	81	87
17	1	MASUK	MASUK	0.721	0.289	79	79	81	83	86

Gambar 12. Hasil confidence masuk dan tidak menggunakan algoritma CART

PerformanceVector

```

PerformanceVector:
accuracy: 86.71% +/- 5.11% (micro average: 86.71%)
ConfusionMatrix:
True:  MASUK  TIDAK
MASUK:  85    10
TIDAK:   9    39
precision: 85.05% +/- 15.16% (micro average: 81.25%) (positive class: TIDAK)
ConfusionMatrix:
True:  MASUK  TIDAK
MASUK:  85    10
TIDAK:   9    39
recall: 81.33% +/- 19.13% (micro average: 79.59%) (positive class: TIDAK)
ConfusionMatrix:
True:  MASUK  TIDAK
MASUK:  85    10
TIDAK:   9    39
AUC (optimistic): 0.978 +/- 0.029 (micro average: 0.978) (positive class: TIDAK)
AUC: 0.977 +/- 0.032 (micro average: 0.977) (positive class: TIDAK)
AUC (pessimistic): 0.975 +/- 0.035 (micro average: 0.975) (positive class: TIDAK)
    
```

Gambar 13. Hasil confidence masuk dan tidak menggunakan algoritma CART

Row No.	JENIS KEJL.	PREDIKAT	predictionP	confidence	confidence	RATA-RATA	RATA-RATA	RATA-RATA	RATA-RATA	RATA-RATA
1	0	TEKAK	TEKAK	0.010	0.982	79	79	79	81	87
2	1	MASUK	MASUK	0.721	0.289	79	79	81	83	86
3	1	TEKAK	MASUK	0.651	0.349	80	80	81	81	86
4	0	TEKAK	TEKAK	0.289	0.721	79	79	80	82	87
5	0	MASUK	MASUK	0.881	0.119	79	79	83	82	87
6	0	MASUK	MASUK	0.798	0.202	78	78	82	84	88
7	0	TEKAK	TEKAK	0.165	0.835	77	78	81	83	87
8	0	MASUK	MASUK	0.983	0.017	80	81	81	83	86
9	1	MASUK	MASUK	0.721	0.289	79	79	81	83	87
10	1	TEKAK	TEKAK	0.119	0.881	78	79	80	82	86
11	1	MASUK	MASUK	0.721	0.289	79	79	81	83	87
12	1	MASUK	MASUK	0.881	0.119	79	79	81	84	88
13	0	MASUK	TEKAK	0.500	0.500	78	79	81	83	87
14	0	MASUK	MASUK	0.721	0.289	79	79	82	82	87
15	0	MASUK	MASUK	0.721	0.289	79	79	81	83	87
16	0	TEKAK	TEKAK	0.047	0.953	78	79	80	81	87
17	1	MASUK	MASUK	0.721	0.289	79	79	81	83	86

Gambar 14. Hasil confidence masuk dan tidak menggunakan algoritma SVM

3.2. Pembahasan Penelitian

Lingga Kurnia Ramadhani, Bajeng Nurul Widyaningrum, setiap perguruan tinggi pasti mengharapkan mahasiswa yang lulus akan sesuai jumlahnya dengan mahasiswa yang masuk [15]. Dalam penelitian ini bermaksud untuk melakukan klasifikasi dengan menggunakan data dari mahasiswa yang lulus dengan predikat kelulusan. Peneliti mencoba membandingkan 2 algoritma yang cukup populer dalam proses klasifikasi yaitu Naive Bayes dan Support Vector Machine. Pada penelitian - penelitian sebelumnya Naive Bayes cenderung menghasilkan performa yang lebih baik dari akurasi dibandingkan dengan Support Vector Machine [2], [16]. Setelah dilakukan pengujian dengan menggunakan data kelulusan mahasiswa ternyata Naive Bayes memiliki akurasi lebih besar yaitu 96.52% dengan tingkat eror (Missclassification Rate) 0.03%, sedangkan metode Support Vector Machine memiliki akurasi 86.93%.

Lely Dwi Bhakti Pratiwi, Wahyu Wibowo, Ismaini Zain dari Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS), berpendapat bahwa hasil karakteristik nilai peminat yang diterima baik yang berasal dari SMA Negeri maupun SMA Swasta serta jenis kelamin laki-laki maupun perempuan mempunyai kesamaan dalam hal rata-rata nilai tertinggi dan rata-rata nilai terendah dimana rata-rata nilai tertinggi adalah nilai Matematika Dasar dan rata-rata terendah adalah nilai Biologi [25]. Pengklasifikasian peminat berdasarkan status penerimaan dengan metode CART menggunakan kombinasi data learning dan data testing sebesar 85 % dan 15%. Hal tersebut dikarenakan kombinasi data learning dan testing tersebut menghasilkan nilai total accuracy rate (1-APER) yang paling tinggi yaitu sebesar 92,3 % untuk data learning dan sebesar 91,4 % untuk data testing [12], [26], [27].

Nurul Indah Prabawati, Widodo, M. Ficky Duskarnaen berjudul Kinerja Algoritma Classification and Regression Tree (Cart) dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta, menyimpulkan bahwa Organisasi kemahasiswaan adalah fasilitas yang disediakan oleh perguruan tinggi sebagai wadah untuk mengembangkan kemampuan non akademis, minat dan bakat mahasiswa. Namun, dalam kenyataannya banyak mahasiswa yang mengikuti organisasi mengalami penurunan prestasi hingga tidak dapat lulus tepat waktu. Di Universitas Negeri Jakarta belum adanya sistem yang dapat mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi [11]. Sebelum membangun system pengambilan keputusan, diperlukan penelitian mengenai akurasi suatu algoritma agar sistem keputusan yang dibuat memiliki tingkat akurasi yang tinggi. Penelitian ini menggunakan algoritma data mining yaitu algoritma Classification and Regression Tree (CART). CART merupakan metode pohon

keputusan biner. *CART* dikembangkan untuk melakukan analisis klasifikasi pada peubah respon baik yang nominal, ordinal, maupun kontinu. Metode klasifikasi *CART* terdiri dari dua metode yaitu metode pohon regresi dan pohon klasifikasi. Data mahasiswa yang mengikuti organisasi yang lulus tepat waktu dan tidak lulus tepat waktu akan diolah menggunakan algoritma *CART* [11], [28].

Setelah diklasifikasikan data tersebut akan dihitung hasil akurasi menggunakan *K-fold Cross Validation* dengan nilai $K = 5$, $k = 10$, dan $K = 20$. Berdasarkan hasil contoh data mahasiswa yang mengikuti organisasi menunjukkan bahwa hasil perhitungan akurasi algoritma *CART* terbaik diperoleh ketika nilai $K = 20$. Algoritma *CART* telah mampu mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta. Algoritma *CART* menghasilkan rata-rata akurasi 80% [12], [29].

Ali Fauzi Berjudul *Classification And Regression Trees (CART) Memprediksi Kelulusan Siswa Pada Suatu Mata Pelajaran Di E-Learning SMA Negeri 1 Pare Kediri*. Penelitian ini menyebutkan bahwa *pemanfaatan* Teknologi dalam dunia pendidikan saat ini terus dikembangkan, salah satunya adalah E-learning yang diterapkan di SMAN 1 Pare. Software yang digunakan untuk menangani sistem E-learning di SMA tersebut adalah Moodle, dan di E-learning tersebut siswa dapat melakukan beberapa macam aktifitas. Dengan memanfaatkan Algoritma Classification And Regression Trees (*CART*) dan berdasarkan data-data yang ada di Database Moodle tersebut, diharapkan mampu untuk melakukan prediksi kelulusan pada suatu mata pelajaran (TIK) yang ada di E-learning SMA Negeri 1 Pare Kediri. Pengolahan data diawali dengan mengambil data sebanyak 10 aktifitas siswa yang tersimpan didalam database moodle, dan kemudian data tersebut di lakukan normalisasi dengan metode Min-Max Normalization. Dari data yang sudah dinormalisasi tersebut diproses dalam distribusi frekuensi sehingga membentuk data interval untuk mempermudah analisa pada Algoritma *CART*. Pada Algoritma *CART*, dilakukan pemilahan untuk setiap decision node menjadi dua cabang yang digunakan untuk membentuk candidate split. Candidate split dipilih untuk penyusunan inisial partisi pada root node dan decision node. Kriteria pemilihan tersebut berdasarkan nilai goodness of split yang terbesar. Pada penelitian ini, dengan rule yang diperoleh dari Algoritma *CART* yang diujikan pada data dengan perbandingan 80% data training dan 20% data testing diperoleh hasil prediksi dengan tingkat akurasi 75,9% [11], [12], [30].

Tentang Analisis Ketepatan Waktu Lulus Berdasarkan Karakteristik Mahasiswa FEM Dan FAPERTA IPB Menggunakan Metode Chart. Fira Nurahmah Al Amin, Indahwati, Yenni Angraini [9]. Pada pohon klasifikasi FAPERTA persentase tepat waktu tertinggi sebesar 26.1% yang memiliki ciri-ciri

yaitu mahasiswa berasal dari Departemen AGH dan PTN, dan memiliki IPK lebih besar dari 3.335. Adapun persentase tepat waktu terendah sebesar 1.6% yang memiliki ciri-ciri yaitu mahasiswa berasal dari Departemen MSL dan ARL, dan berjenis kelamin Laki-laki. Nilai ketepatan pohon klasifikasi FAPERTA sebesar 88.2%. Pada pohon klasifikasi FEM persentase tepat waktu tertinggi sebesar 75% memiliki ciri-ciri yaitu mahasiswa berasal dari Departemen IE dan MAN. Adapun persentase tepat waktu terendah sebesar 22.4% yang memiliki ciri-ciri yaitu mahasiswa berasal dari Departemen AGB dan ESL, memiliki IPK kurang dari sama dengan 3.325, dan melalui Jalur Masuk SNMPTN. Nilai ketepatan pohon klasifikasi FEM sebesar 71.4% [31], [32].

Tentang Implementasi Data Mining Untuk Menentukan Minat Siswa Dalam Menentukan Jurusan Pada Perguruan Tinggi Saeful Bahri, *ITB Ahmad Dahlan, Jakarta, menyebutkan bahwa* dari hasilpengujian tiga algoritma yang digunakan, kinerja algoritma *decision tree* menjadi yang terbaik dibandingkan dengan algoritma *k-NN* dan *Naive bayes* dengan tingkat akurasi 75.38% dan nilai AUC 0.689 [31], [33].

Berkenaan dengan Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima KIP Kuliah Menggunakan Regresi Logistik Biner yang dilakukan oleh Ronny Susetyoko, Wiratmoko Yuwono, Elly Purwantini, dari *Politeknik Elektronika Negeri Surabaya, Indonesia, menyebutkan bahwa* seleksi mahasiswa baru penerima Kartu Indonesia Pintar Kuliah (KIP Kuliah) dilakukan oleh setiap institusi untuk memilih mahasiswa yang benar-benar memiliki potensi akademik yang baik dan keterbatasan ekonomi. Agar mahasiswa penerima KIP Kuliah tepat sasaran maka perlu dibuat rumusan atau model yang tepat dengan mempertimbangkan fitur-fitur yang representatif. Penelitian ini bertujuan untuk membuat model klasifikasi mahasiswa baru penerima KIP Kuliah menggunakan Metode Regresi Logistik Biner (Logit Biner) [34], [35].

Beberapa fitur yang diduga berpengaruh dalam model logit biner ini adalah status orang tua, penghasilan, status rumah, jumlah rumah, jumlah motor, jumlah mobil, dan daya listrik. Pada tahap *preprocessing* data, dilakukan pembersihan data dan *encoding* beberapa fitur. Beberapa model dibandingkan kinerjanya, baik yang menggunakan data set asli, data set hasil normalisasi, data set hasil *undersampling*, data set hasil *oversampling*, maupun data set hasil kombinasi *oversampling* dan *undersampling*. Sebelum pemodelan, dilakukan pemisahan antara data *training* dan data *testing*. Evaluasi model berdasarkan signifikansi parameter di dalam model logit dan kinerja klasifikasi dari matriks konfusi. Berdasarkan hasil pengujian parameter, beberapa fitur yang signifikan dalam model adalah penghasilan, status_rumah, jumlah_motor, dan jumlah mobil. Dari perbandingan tujuh model logit biner, model yang terbaik adalah model yang

menggunakan data set asli dengan rerata F1 Score 92,40%, rerata *recall* sebesar 87,93%, *accuracy* sebesar 88,01%, *precision* sebesar 97,92%, dan AUC sebesar 84,6% [36].

Data tentang Prediksi Kelanjutan Studi Siswa Ke Perguruan Tinggi Dengan Naive Bayes Gentur Wahyu Nyipto Wibowo, Zaenal Arifin, Muhammad Anwarudin Romli, Nurul Ikhsanul Amal dari Fakultas Sains dan Teknologi Universitas Islam Nahdlatul Ulama Jepara, *menyebutkan bahwa* Metode Naive Bayes dapat menghasilkan probabilitas di setiap kriteria dengan class yang berbeda, sehingga nilai-nilai probabilitasnya dapat digunakan untuk mengoptimalkan prediksi kelanjutan studi siswa ke perguruan tinggi mengacu pada pengklasifikasian yang dilakukan Naive Bayes sendiri, dan pemilihan pembobotan pada atribut dapat meningkatkan nilai akurasi pengklasifikasian dari data yang diujikan sehingga menghasilkan nilai prediksi 86.53% [15], [37].

Data tentang Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa tepat waktu dilakukan oleh Ananda Fiqri Firdaus, Rohmat Saedudin, Rachmadita Andeswari. Dengan melakukan prediksi akan probabilitas kelulusan mahasiswa diharapkan para dosen dapat fokus memberikan perhatiannya kepada mahasiswa yang terprediksi kelulusannya rendah. Dengan mengimplementasikan klasifikasi data mining, Memprediksi kelulusan mahasiswa akan jauh lebih mudah dan efisien. Tujuan Penelitian ini adalah untuk mengimplementasikan klasifikasi dengan metode naive bayes untuk memprediksi kelulusan mahasiswa. Penelitian ini juga menguji rasio antara data training dan data tes mana yang tepat untuk digunakan agar hasil yang diperoleh lebih optimal. Berdasarkan hasil penelitian ini diperoleh bahwa dengan menggunakan 80% data sebagai data training dihasilkan hasil akurasi paling optimal dengan tingkat akurasi 90.78%, presisi data sebesar 88% dan 88.4% untuk recall. Dan naive bayes memperoleh tingkat akurasi tertinggi sebesar 91.49% dengan menggunakan 70% data sebagai data training [16], [38].

Tentang Prediksi kelulusan dilakukan oleh Azahari, Yulindawati, Dewi Rosita, dan Syamsuddin Mallala. Manajemen perguruan tinggi dalam menentukan kebijakan preventif terkait pencegahan dini kasus drop out. Dengan menggunakan data mining algoritma naive bayes dan neural network dapat dilakukan prediksi kelulusan mahasiswa di STMIK Widya Cipta Dharma (WiCiDa) Samarinda. Atribut yang digunakan yaitu, umur saat masuk kuliah, klasifikasi kota asal Sekolah Menengah Atas, pekerjaan ayah, program studi, kelas, jumlah saudara, dan Indeks Prestasi Kumulatif (IPK). Dari data testing diperoleh tingkat akurasi hanya 57,63%. Hasil penelitian menunjukkan banyaknya kelemahan dari hasil prediksi naive bayes dikarenakan tingkat akurasi kevalidannya tergolong tidak terlalu tinggi.

Sedangkan akurasi prediksi neural network adalah 72,58% [3], [15].

Data tentang Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma Support Vector Machine. dilakukan oleh Rooy Thaniket, Kusri, Emha Taufik Luthf. Kelulusan adalah keinginan setiap mahasiswa untuk dapat menyelesaikan studi Untuk Mencapai kelulusan Mahasiswa harus menyelesaikan tahapan seperti menempuh 8 semester 144 sks dan itu sebagai aturan dalam perguruan tinggi [39]. Pada penelitian kali ini peneliti menggunakan data mahasiswa lulusan dari tahun 2015-2019 yang peneliti ambil disalah satu universitas metode yang digunakan peneliti adalah .data mining. Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Algoritma yang peneliti gunakan dalam memprediksi kelulusan adalah algoritma Support Vector Machine (SVM) karna mampum mempredik kelulusan yang baik. Dalam prediksi kelulusan menghasilkan rata-rata akurasi sebesar Accuracy 95,00% [16], [26].

Data tentang Analisis Metode Decision Tree Dalam Memprediksi Kelulusan Mahasiswa dilakukan oleh Wulandari, Rika Rosnelly, Wanayumini, Universitas merupakan sarana penyelenggaraan pendidikan akademik bagi mahasiswa. Salah satu faktor yang menentukan mutu perguruan tinggi adalah persentase ketuntasan belajar mahasiswa dalam suatu masa studi. Mahasiswa merupakan salah satu parameter penting untuk menentukan kualitas dan mengevaluasi rencana pembelajaran dan pelaksanaan perguruan tinggi. Khusus untuk standar dan evaluasi pascasarjana, unsur penilaiannya adalah rata-rata waktu studi dan IPK. Perlu dilakukan penelitian untuk memprediksi kelulusan mahasiswa, dengan menggunakan menggunakan metode Decision Tree dan algoritma C4.5. Hasil yang diperoleh mendapatkan rule prediksi nilai yang benar dan salah. Nilai akurasi yang diperoleh didasarkan pada nilai pengakuan berhasil dan tidak berhasil [16].

Berkenaan Analisis Prediksi Masa Studi Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritme Cart (Classification And Regression Trees) [3] dilakukan oleh Muhammad Fando Rizalno Asahar Johar Funny F. Coastera [17]. Metode yang digunakan untuk menganalisisnya adalah Decission tree dengan menggunakan algoritme CART [3]. Melakukan prediksi dibutuhkan dataset yang teratur, Dataset yang digunakan masih mengandung missing values sehingga dalam penelitian ini tahap preprocessing data dilakukan. Tahap preprocessing menggunakan sistem data warehouse dengan penerapa ETL. Aplikasi ini dibuat dengan menggunakan Visual studio code bahasa pemrograman PHP framewort laravel. Hasil dari penelitian ini berupa aplikasi pengelompokkan data alumni lulus tepat waktu dan prediksi lama masa studi utuk mahasiswa baru. Pengujian software menggunakan metode Evaluasi Accuracy dan Black

box. Hasil dari penelitian ini adalah nilai parameter yang berpengaruh pada kelulusan mahasiswa dan dapat digunakan untuk proses prediksi masa studi mahasiswa [14], [15].

Pembahasan dalam menentukan Penjurusan Siswa Dengan Menggunakan Metode Decision Tree Algoritma C4.5 dilakukan oleh Muhammad Fibo Donya Ikbal, Denny Kurniadi. Penelitian ini bertujuan untuk mengklasifikasi penjurusan siswa dan menerapkan metode data mining menggunakan algoritma decision tree C4.5. Klasifikasi ini menggunakan 6 atribut sebagai atribut kontrol yang merupakan nilai rata-rata rapor SMP yang meliputi mata pelajaran B. Indonesia, B. Inggris, Matematika dan IPA, serta rekomendasi guru BK SMP. Berdasarkan pengujian yang dilakukan terhadap siswa tahun masuk ajaran 2020/2021 diperoleh akurasi sebesar 68.42% dengan 304 sampel data dengan atribut rekomendasi guru BK SMP yang mempengaruhi keputusan penjurusan siswa. Demikian dapat disimpulkan bahwa menentukan penjurusan dengan menggunakan data mining algoritma decision tree C4.5 dapat mempercepat dan akurat dalam pengambilan keputusan penjurusan siswa SMA Negeri 2 Padang [13], [16], [26].

Dari pembahasan dapat dijelaskan bahwa banyak algoritma klasifikasi yang digunakan dalam penelitian di dunia pendidikan. Meskipun algoritma tersebut memiliki kemampuan yang serupa, tetapi pastinya algoritma tersebut memiliki tingkat keakuratan yang berbeda. Maka dari itu perbandingan terhadap algoritma klasifikasi tersebut dibutuhkan untuk memberikan informasi tentang performace yang paling baik sehingga menghasilkan klasifikasi yang lebih akurat dalam hal mengklasifikasi menentukan siswa berprestasi masuk Perguruan Tinggi Negeri. Maka dari itu peneliti akan melakukan pendekatan algoritma klasifikasi dalam menentukan siswa berprestasi diterima di Perguruan Tinggi Negeri dengan metode pendekatan *classification and regression trees* (CART) dan Menggunakan metode *support vector machine* (SVM).

4. KESIMPULAN

Dari hasil optimasi algoritma CART dengan SVM Berbasis algoritma untuk klasifikasi siswa masuk PTN (Perguruan Tinggi Negeri) dapat disimpulkan ke dalam kesimpulan umum dan kesimpulan khusus. Kesimpulan umum dari penelitian ini adalah bahwa perbandingan menggunakan algoritma SVM lebih baik dan lebih tinggi persentasinya dibanding menggunakan algoritma CART.

Adapun kesimpulan khusus penelitian ini sebagai berikut:

- a. Metode Algoritma CART untuk mengklasifikasi siswa masuk perguruan tinggi negeri sebesar 86.10%

- b. Metode Algoritma SVM untuk mengklasifikasi siswa masuk perguruan tinggi negeri sebesar 86.71%.

5. SARAN

Dari hasil penelitian ini masih banyak kekurangan algoritma klasifikasi CART dibanding dengan SVM untuk mengklasifikasikan siswa masuk perguruan tinggi negeri yaitu:

1. Membandingkan hasil penelitian ini dengan metode yang tingkat akurasi lebih baik.
2. Penelitian selanjutnya bisa diharapkan penerapan algoritma bisa ditingkatkan lagi akurasi dengan beberapa eksperimen dan parameter maupun dataset yang lebih baik.

UCAPAN TERIMA KASIH

Ucapan terima kasih kepada kedua dosen pembimbing dan seluruh dosen mata kuliah serta bapak & ibu, dan keluarga.

DAFTAR PUSTAKA

- [1] O. Arifudin, "Manajemen Sistem Penjaminan Mutu Internal (Spmi) sebagai upaya meningkatkan mutu perguruan tinggi," *J. Ilm. MEA Manajemen, Ekon. Akunt.*, vol. 3, no. 1, pp. 161–169, 2019.
- [2] A. Karim, N. F. Mardhotillah, and M. I. Samadi, "Ethical leadership transforms into ethnic: Exploring new leader's style of Indonesia," *J. Leadersh. Organ.*, vol. 1, no. 2, pp. 146–157, 2019.
- [3] S. Abbas, *Manajemen Perguruan Tinggi*. Jakarta, Indonesia: Prenada Media Group, 2008.
- [4] J. Mantik, R. Zaeni, A. Syam, and W. Achmad, "Online Learning in Higher Education: Analysis during the Pandemic Covid-19," *J. Mantik*, vol. 5, no. 4, pp. 2256–2261, 2022.
- [5] A. S. Maadi, "Digitalisasi manajemen pendidikan Islam dan ekonomi syariah di Perguruan Tinggi," *Fikrotuna J. Pendidik. dan Manaj. Islam*, vol. 7, no. 1, pp. 741–759, 2018.
- [6] M. N. Husunun, *Menuju Kepustakawanan Perguruan Tinggi Muhammadiyah Aisyiyah berkemajuan*. Surakarta: UMS Library, 2017.
- [7] S. R. Ratnawati and W. Werdiningsih, "Pemanfaatan e-learning sebagai inovasi media pembelajaran PAI di era revolusi industri 4.0," *Belajera J. Pendidik. Islam*, vol. 5, no. 2, pp. 199–220, 2020, doi: 10.29240/belajera.v5i2.1429.
- [8] W. Cayeni and A. S. Utari, "Penggunaan Teknologi Dalam Pendidikan: Tantangan Guru Pada Era Revolusi Industri 4.0," in

- Prosiding Seminar Nasional Pendidikan Program Pascasarjana Universitas PGRI Palembang 03 Mei 2019*, 2019, pp. 658–667. [Online]. Available: <https://jurnal.univpgri-palembang.ac.id/index.php/Prosidingpps/article/view/3096>
- [9] I. M. W. Jagantara, P. B. Adnyana, and N. L. P. M. Widiyanti, “Pengaruh model pembelajaran berbasis proyek (PjBP) terhadap hasil belajar Biologi ditinjau dari gaya belajar siswa SMA,” *e-Journal Progr. Pascasarj. Univ. Pendidik. Ganesha*, vol. 4, no. 3, pp. 1–13, 2014.
- [10] K. A. Winata, Sahudi, Nurwadjah, and A. Suhartini, “Implementasi konsepsi tujuan pembelajaran PAI untuk menghadapi era revolusi 4.0,” *EduTeach J. Edukasi dan Teknol. Pembelajaran*, vol. 1, no. 2, pp. 11–21, 2020, doi: 10.37859/eduteach.v1i2.1964.
- [11] N. I. Prabawati, Widodo, and M. Ficky, “Duskarnaen. Kinerja Algoritma Classification and Regression Tree (Cart) dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta,” *J. Pinter*, vol. 3, no. 2, 2019.
- [12] L. D. B. Pratiwi, W. Wibowo, and I. Zain, “Klasifikasi Nilai Peminat SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri) ITS dengan Pendekatan Classification and Regression Trees (CART),” *J. Sains Dan Seni Pomits*, vol. 4, no. 2, p. 15, 2020.
- [13] K. W. Lamm, H. S. Carter, and A. J. Lamm, “A theory based model of interpersonal leadership: An integration of the literature,” *J. Leadersh. Educ.*, vol. 15, no. 4, pp. 183–205, 2016, [Online]. Available: doi:
- [14] Wulandari, R. Rosnelly, and Wanayumini, “Analisis Metode Decision Tree Dalam Memprediksi Kelulusan Mahasiswa132,” *CSRID J.*, vol. 13, no. 3, pp. 130–140, 2021, doi: <https://www.doi.org/10.22303/csr.13.3.2021.130-140> e-ISSN :2460-870X.
- [15] G. W. N. Wibowo, Z. Arifin, M. Anwarudin, and N. I. Romli, “Prediksi Kelanjutan Studi Siswa Ke Perguruan Tinggi Dengan Naive Bayes,” *J. DISPROTEK*, vol. 11, no. 1, 2020.
- [16] M. R. Qisthiano, “Klasifikasi Terhadap Prediksi Kelulusan Mahasiswa Dengan Menggunakan Metode Support Vector Machine (SVM),” in *2 st Proceeding STEKOM*, Jakarta, Indonesia: Institut Teknologi dan Bisnis Nasional (ITBN), 2022.
- [17] J. W. Cresswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (4th ed.)*. Thousand Oaks, CA: Sage, 2014.
- [18] G. Guthrie, *Auerbach Qualitative Data*. 2014. doi: 10.4135/9788132105961.n15.
- [19] Sugiyono, *Quantitative, Qualitative, and R&D Research Methods*. Bandung, Indonesia: Alfabeta, 2010.
- [20] T. Perumal, “Quantitative Research Methods,” *Course Mater.*, vol. Chapter 8, no. Leedy 1993, pp. 87–104, 2014, doi: <http://dx.doi.org/10.5210/fm.v8i1.1023>.
- [21] A. Çelik *et al.*, “Science Integration,” *J. Mater. Process. Technol.*, vol. 1, no. 1, pp. 1–8, 2018, doi: 10.1016/j.cirp.2016.06.001.
- [22] J. Afriana, A. Permanasari, and A. Fitriani, “Penerapan project based learning terintegrasi STEM untuk meningkatkan literasi sains siswa ditinjau dari gender,” *J. Inov. Pendidik. IPA*, vol. 2, no. 2, p. 202, 2016, doi: 10.21831/jipi.v2i2.8561.
- [23] E. Ersanilli, J. Carling, and H. De Haas, “Methodology for quantitative data collection.” *Eumagine, Emagining Europe from the Outside*, pp. 1–27, 2011.
- [24] Jubilee, *SPSS Komplet untuk Mahasiswa*. Jakarta, Indonesia: Elex Media Komputindo, 2018.
- [25] G. Galih, “Data Mining di Bidang Pendidikan untuk Analisa Prediksi Kinerja Mahasiswa dengan Komparasi 2 Model Prediksi pada STMIK Jabar,” *J. Teknol. Sist. Inf. dan Apl.*, vol. 2, no. 1, p. 23, 2019, doi: 10.32493/jtsi.v2i1.2643.
- [26] M. F. D. Ikhbal and D. Kurniadi, “Menentukan Penjurusan Siswa Dengan Menggunakan Metode Decision Tree Algoritma C4.5 (Studi Kasus: SMA Negeri 2 Padang),” *J. VOKASI Inform.*, doi: <https://doi.org/10.24036/javit.v1i3.40> ISSN: 2775 - 6807.
- [27] A. Karim, “Pembaharuan pendidikan Islam multikulturalis,” *J. Pendidik. Agama Islam - Ta’lim*, vol. 14, no. 1, pp. 19–35, 2016, [Online]. Available: <http://jurnal.upi.edu/taklim/view/3880/pembaharuan-pendidikan-islam-multikulturalis-.html>
- [28] S. L. Qodriah, W. Hartati, and A. Karim, “Self-leadership and career success: Motivation of college lecturers,” *J. Leadersh. Organ.*, vol. 1, no. 2, pp. 79–95, 2019.
- [29] Fikriyah, A. Karim, M. K. Huda, and A. Sumiati, “Spiritual leadership: The case of instilling values in students through the Kiai’s program in the globalization era,” *J. Leadersh. Organ.*, vol. 3, no. 1, pp. 16–30, 2021, doi: <https://doi.org/10.22146/jlo.63922>.
- [30] A. Karim and D. Afnan, “Kiai interpersonal

- managerial: Henry Mintzberg perspective,” *J. Leadersh. Organ.*, vol. 2, no. 2, pp. 75–90, 2020, doi: <https://doi.org/10.22146/jlo.56290>.
- [31] S. Bahri, “Implementasi Data Mining Untuk Menentukan Minat Siswa Dalam Menentukan Jurusan Pada Perguruan Tinggi,” *ITB Ahmad Dahlan*, vol. 3, no. 1, pp. 23–33, 2022.
- [32] D. Afnan *et al.*, “Phubbing and social interaction: An analysis of smartphone usage in higher education,” *J. Adv. Res. Dyn. Control Syst.*, vol. 12, no. 6, pp. 2325–2334, 2020, doi: [10.5373/JARDCS/V12I6/S20201191](https://doi.org/10.5373/JARDCS/V12I6/S20201191).
- [33] A. Karim and F. Wajdi, “Propaganda and da’wah in digital era (A case of hoax cyber-bullying against ulama),” *KARSA J. Sos. dan Budaya Keislam.*, vol. 27, no. 1, pp. 171–202, 2019, doi: [10.19105/karsa.v27i1.1921](https://doi.org/10.19105/karsa.v27i1.1921).
- [34] E. Priyanto, “Dakwah dan kesalehan sosial: kiprah dakwah Roostien Ilyas,” Universitas Islam Negeri Syarif Hidayatullah Jakarta, 2015.
- [35] D. Widiantari, M. I. Bin Samadi, and A. Karim, “Charismatic Leadership Effects of Teachers in Fostering Graduate Quality of Senior High School,” *J. Leadersh. Organ.*, vol. 4, no. 2, pp. 179–190, 2022, doi: [10.22146/jlo.74872](https://doi.org/10.22146/jlo.74872).
- [36] N. Anwar and I. Riadi, “Analisis Investigasi Forensik WhatsApp Messenger Smartphone Terhadap WhatsApp Berbasis Web,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 3, no. 1, p. 1, 2017, doi: [10.26555/jiteki.v3i1.6643](https://doi.org/10.26555/jiteki.v3i1.6643).
- [37] M. R. Syabibi, A. Karim, S. Kulkarni, and A. Sahil, “Communicative cultural dakwah of Abdurrahman Wahid in pluralistic society,” *Karsa J. Soc. Islam. Cult.*, vol. 29, no. 2, pp. 1–33, 2021, doi: [10.19105/karsa.v29i2.5220](https://doi.org/10.19105/karsa.v29i2.5220).
- [38] A. Karim, A. Faiz, N. Nur’Aini, and F. Y. Rahman, “The policy of organization, the spirit of progressivism Islam, and its association with social welfare educators,” *Tatar Pas. J. Diklat Keagamaan*, vol. 16, no. 1, pp. 69–75, 2022.
- [39] F. Mansir and A. Karim, “Islamic education learning approaches in shaping students’ emotional intelligence in the digital age,” *HAYULA Indones. J. Multidiscip. Islam. Stud.*, vol. 4, no. 1, pp. 67–86, 2020, doi: <https://doi.org/10.21009/004.01.04..>