

# Algoritma C4.5 Berbasis Forward Selection Untuk Klasifikasi Bidang Minat Studi Mahasiswa Teknik Informatika

<sup>1</sup> Sirli Fahriah, <sup>2</sup>Wiktasari

<sup>1,2</sup> Teknik Informatika, Jurusan Teknik Elektro, Politeknik Negeri Semarang  
E-mail : <sup>1</sup> sirlifahriah@polines.ac.id, <sup>2</sup> wikta@polines.ac.id

## Abstrak

Bidang minat merupakan faktor terpenting dalam dunia pendidikan. Pada umumnya mahasiswa atau pelajar tidak mengerti bidang minat yang sesuai dengan bakat mereka. Salah satu dampak yang kemungkinan akan terjadi, dampak lainnya akibat kurang tepat memilih bidang minat yaitu mahasiswa enggan mempelajari bidang studi yang sudah terlanjur dipilih sehingga mengakibatkan turunnya nilai akademik yang berdampak lulus tidak tepat waktu. Pada fakultas ilmu komputer di sebuah Universitas terdapat program studi sarjana Teknik Informatika yang dibagi menjadi dua bidang minat yaitu bidang minat sistem cerdas dan bidang minat rekayasa perangkat lunak design. Dalam pemilihan bidang minat, Sebagian besar mahasiswa merasa kesulitan untuk menentukan bidang minat mereka. Dalam menangani hal ini, perlu adanya gagasan yang dapat merekomendasikan mahasiswa untuk memilih bidang minat mereka. Pada kasus ini, diperlukan metode atau algoritma yang tepat berdasarkan kriteria data yang akan digunakan sebagai variabel. Metode atau algoritma yang akan dipakai pada penelitian ini adalah algoritma C4.5 berbasis forward selection. Algoritma C4.5 berbasis forward selection ini dibandingkan dengan algoritma C4.5. Algoritma C4.5 merupakan Algoritma Klasifikasi dari cabang ilmu Data mining. Metode klasifikasi akan memberi keputusan berdasarkan atribut data yang dipilih. Hasil dari klasifikasi pemilihan bidang minat menggunakan data yang sudah seimbang dan atribut yang telah di seleksi beberapa fiturnya serta dilakukan iterasi pada cross validation sehingga menghasilkan akurasi yang tepat. Berdasarkan hasil pengujian dengan dua metode, yaitu pengujian yang hanya menggunakan Algoritma C4.5 menghasilkan akurasi sebesar 84.33% dan pengujian menggunakan algoritma C4.5 berbasis forward selection menghasilkan akurasi sebesar 85.00%. Dari dua pengujian ini membuktikan bahwa terdapat peningkatan akurasi dengan penambahan fitur seleksi pada algoritma C4.5.

**Kata kunci :** Data Mining, Klasifikasi, Algoritma C4.5, feature selection, bidang minat

## Abstract

*The area of interest is the most important factor in the world of education. In general, students or students do not understand the areas of interest that match their talents. One of the possible impacts is due to the inaccurate choice of fields of interest, namely students are reluctant to study the field of study that has already been chosen, resulting in a decrease in academic scores which results in an untimely graduation. In the faculty of computer science at a university, there is an undergraduate study program in Informatics Engineering which is divided into two areas of interest, namely midwives with an interest in intelligent systems and an area of interest in software design engineering. In selecting areas of interest, most students find it difficult to determine their areas of interest. In dealing with this, it is necessary to have ideas that can recommend students to choose their area of interest. In this case, an appropriate method or algorithm is needed based on the data criteria to be used as a variable. The method or algorithm that will be used in this research is the C4.5 algorithm based on forward selection. The C4.5 algorithm based on forward selection is compared to the C4.5 algorithm. The C4.5 algorithm is a classification algorithm from the branch of data mining. The classification method will make decisions based on the selected data attributes. The results of the classification of the selection of areas of interest use balanced data and attributes that have been selected for several features and cross validation iteration is carried out to produce the right accuracy. Based on the results of testing with two methods, testing using only the C4.5 algorithm produces an accuracy of 84.33% and testing using the C4.5 algorithm based on forward selection produces an accuracy of 85.00%. These two tests prove that there is an increase in accuracy with the addition of a selection feature to the C4.5 algorithm.*

**Keywords :** Data Mining, Classification, Algorithm C4.5; feature selection, areas of interest.

## I. PENDAHULUAN

Bidang minat merupakan faktor terpenting dalam dunia pendidikan. Pada umumnya mahasiswa atau pelajar tidak mengerti bidang minat yang sesuai dengan bakat mereka. Menurut beberapa ahli, Bakat atau bidang minat merupakan kondisi dimana seseorang memungkinkan dapat mencapai suatu kecakapan, pengetahuan dan keterampilan khusus melewati latihan yang khusus [1]. Bakat adalah kondisi ataupun kualitas yang dimiliki masing – masing individu, yang kemungkinan individu tersebut berkembang pada masa yang akan datang [2]. Salah satu dampak yang timbul akibat pemilihan bidang minat yang kurang tepat yaitu kurangnya minat belajar mahasiswa. Dari semua dampak yang kemungkinan akan terjadi, dampak lainnya akibat kurang tepat memilih bidang minat yaitu mahasiswa enggan mempelajari bidang studi yang sudah terlanjur dipilih sehingga mengakibatkan turunnya nilai akademik yang berdampak lulus tidak tepat waktu. Pada sebagian besar penelitian psikologi pendidikan, minat dan bakat siswa cukup terkait dengan prestasi akademik [3].

Berkaitan dengan bidang minat, mahasiswa program studi teknik informatika fakultas ilmu komputer pada sebuah Universitas dibagi menjadi dua bidang minat yaitu sistem cerdas dan rekayasa perangkat lunak design. Untuk mendapatkan sebuah pola dari data yang akan dianalisis sehingga menjadi sebuah hasil keputusan, diperlukan beberapa proses dan tahapan seperti pemilihan atribut atau parameter dari data yang telah diolah. penentuan atribut berdasarkan beberapa pertimbangan yang benar benar berpengaruh pada hasil keputusan. Berdasarkan beberapa penelitian yang telah dilakukan sebelumnya, dapat menggunakan berbagai macam algoritma data mining untuk menyelesaikan masalah serupa tentang pemilihan jurusan. Algoritma tersebut diantaranya decision tree, naive bayes, clustering dan lain sebagainya. Berdasarkan beberapa penelitian yang telah dilakukan mengatakan bahwa, peneliti melihat kinerja algoritma C4.5 baik untuk penentuan bidang minat [4]. Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utaman dari algoritma lainnya. Kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe dikret dan numerik. Dalam mengkontruksi pohon, algoritma

C4.5 membaca seluruh sampel data training dari storage dan memuatnya ke memori. Hal ini yang menjadi salah satu kelemahan algoritma C4.5 dalam kategori skalabilitas yaitu algoritma ini hanya dapat digunakan jika data training dapat disimpan secara keseluruhan dan pada waktu yang bersamaan di memori [5]. Dalam pendekatan masalah ini, direncanakan akan menggunakan algoritma C4.5 berbasis seleksi fitur. Penerapan algoritma C4.5 berbasis seleksi fitur diterapkan untuk ketepatan pemilihan bidang minat mahasiswa. Dengan menggunakan pendekatan ini diharapkan dapat meningkatkan nilai akurasi dan menghasilkan keputusan yang akurat.

Beberapa penelitian terkait seperti Prafulla Kalapatapu dkk [6] meneliti tentang klasifikasi music di india pada tahun 2016. Hasil penelitian menunjukkan bahwa penggunaan seleksi fitur tidak selalu meningkatkan akurasi, namun tetap dapat meningkatkan akurasi di beberapa situs bergantung pada metode yang tepat. Pemilihan fitur seleksi berbasis informasi gain dari neural network dan SVM merupakan yang terbaik untuk data set lagu india. Pada penelitian lain, Al-Sarem [7] meneliti di tahun 2015. Penelitian ini untuk kepentingan akademik berdasarkan algoritma C4.5. Hasil dari penelitian menunjukkan bahwa algoritma C4.5 dapat di terapkan dalam kasus tersebut dan memberikan keakuratan dalam klasifikasi. Pada tahun 2016, Wiwit Supriyanti dkk [8] meneliti tentang Perbandingan Kinerja Algoritma C4.5 dan Naïve Bayes untuk Ketepatan Pemilihan Konsentrasi Mahasiswa. Uji coba di lakukan dengan menggunakan data yang di dapatkan dari Universitas Muhammadiyah Surakarta. Hasil komparasi antara C4.5 dan Naïve Bayes yang digabungkan dengan forward selection didapatkan hasil akurasi tertinggi yaitu C4.5 sebesar 84,98%.

Begitu pula dengan Yusuf S Nugroho dkk [9] yang membandingkan algoritma naïve bayes, C4.5 dan K-means untuk pemilihan jurusan pada SMA N 1 Boyolali. Data yang digunakan terdiri dari gender, minat, rata rata nilai IPA, rata rata nilai IPS, psikotest IPA, psikotest IPS, asal sekolah dan jurusan. Berdasarkan nilai accuracy dan recall, metode decision tree lebih baik dibandingkan dengan metode yang lain dengan nilai accuracy sebesar 79,14% dan nilai recall sebesar 90,80%. Berdasarkan nilai precision, metode naive bayes lebih baik dibandingkan dengan metode yang lain dengan nilai 77,51%.

## II. METODE PENELITIAN

### 2.1 Desain Penelitian

Untuk menunjang penelitian pada bidang data mining yang akan dilakukan biasanya diterapkan metode CRISP-DM (Cross Industry Standart Process for Data Mining) merupakan model dari proses data mining untuk menggambarkan urutan pada data mining, menyediakan standart proses data mining yang digunakan sebagai strategi dalam menyelesaikan permasalahan bersifat umum. Sementara penelitian pada penelitian ini menggunakan penelitian Eksperimen. Penelitian Eksperimen melibatkan penyelidikan perlakuan pada parameter variabel dan menggunakan tes yang ditentukan oleh peneliti. Metode penelitian dilakukan beberapa langkah seperti pengumpulan data set, pengolahan data set, metode yang diusulkan, evaluasi dan validasi hasil.

### 2.2 Pengumpulan Data Set

Pengumpulan data dalam penelitian ini mengambil data mahasiswa Teknik Informatika dari sebuah Univeristas. Data yang diambil dalam penelitian ini adalah data mahasiswa program studi Teknik Informatika sebanyak 953 record. Data akan dibagi menjadi data training untuk mendapatkan model data dan data testing untuk menguji model data. Pembagian data training dan testing dibagi dengan perbandingan 9 banding 1. Data tersebut diantaranya Id mahasiswa, status mahasiswa, IPS, mata kuliah yang telah ditempuh beserta nilainya dan bidang minat yang diambil. Tidak semua atribut digunakan berdasarkan data yang didapat, namun melalui preprocessing data untuk mendapatkan atribut yang mungkin berpengaruh untuk pemilihan bidang minat studi mahasiswa. Syarat untuk bidang minat yang telah di tetapkan adalah beberapa mata kuliah yang telah di tulis dalam buku panduan akademik yang didapatkan pada saat awal masuk perkuliahan Di tahapan ini atribut data yang digunakan antara lain: Nilai Mata Kuliah dan bidang minat.

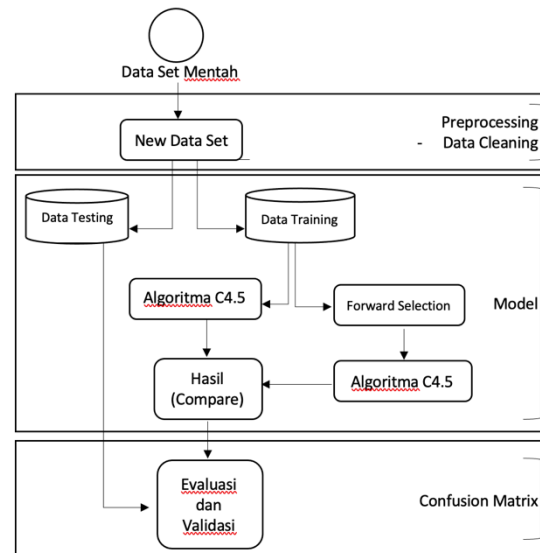
### 2.3 Pengolahan Data Set

Proses pengolahan data awal diperlukan untuk menyiapkan data yang benar benar valid sebelum di proses. Dalam penelitian ini dilakukan dengan membersihkan data atau cleaning yang kosong. Pada tahapan ini dilakukan pengolahan awal terhadap data set untuk menyederhanakan data agar data dapat dikenali dan digunakan sebagai model dari algoritma yang telah diusulkan. Data akan diproses dengan tools bantu data mining.

Apabila terdapat data yang masih inkonsistensi data atau missing value, maka perlu adanya penyeleksian data dengan filter data. Untuk data yang miising value dilakukan dengan penghapusan record pada data yang missing value. Data yang di filter adalah data yang memiliki atribut yang lengkap sesuai yang di tentukan. Peneliti hanya memakai data yang lengkap dengan atribut yang telah dipilih. Setelah data di seleksi dan dibersihkan didapatkan variabel atau atribut yang berpengaruh sehingga variabel menjadi 12 atribut regular dan 1 atribut label dan dapat langsung di olah.

### 2.4 Metode yang diusulkan

Metode yang diusulkan untuk melakukan klasifikasi pemilihan bidang minat studi mahasiswa teknik informatika yaitu penerapan algoritma C4.5 berbasis Forward Selection. Dataset akan dilakukan seleksi dengan metode Forward Selection sebagai metode seleksi fitur terhadap atribut-atribut yang kurang berpengaruh atau tidak relevan pada dataset dengan tujuan dapat meningkatkan akurasi. Selanjutnya untuk validasi menggunakan 10-fold cross validation. Tahapan dan metode digambarkan seperti dibawah:



Gambar 1 Metode yang diusulkan

### 2.5 Validasi dan Evaluasi Hasil

Setelah melewati beberapa tahapan dan proses, hasil dari penelitian adalah sebuah model dan akurasi dari data yang telah di proses. Model atau pola tersebut yang dapat dijadikan klasifikasi untuk pemilihan bidang minat studi mahasiswa teknik informatika maka dari itu perlu adanya evaluasi dan validasi. Selanjutnya dalam tahapan

ini akan dilakukan evaluasi terhadap model yang telah didapatkan pada fase sebelumnya. Tujuan dilakukannya evaluasi yaitu menyesuaikan model yang diperoleh agar tepat dengan sasaran sesuai target ingin dicapai. Pada tahapan ini model yang terbentuk dari penerapan algoritma C4.5 berbasis Forward Selection akan diuji menggunakan diagram Confusion Matrix untuk mengetahui nilai akurasi. Dari hasil diagram Confusion Matrix tersebut akan diketahui hasil prediksi positive yang benar dan salah, serta hasil prediksi negative yang benar dan salah. Kemudian dilakukan proses validasi menggunakan metode k-fold cross validation untuk mengetahui rata-rata keberhasilan dari suatu sistem, caranya yaitu melakukan perulangan dengan meng acak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak [10]. Hasil dari klasifikasi tersebut diharapkan bahwa dengan menggunakan seleksi fitur forward selection dapat mengoptimalkan akurasi dan kinerja yang baik terhadap algoritma C4.5.

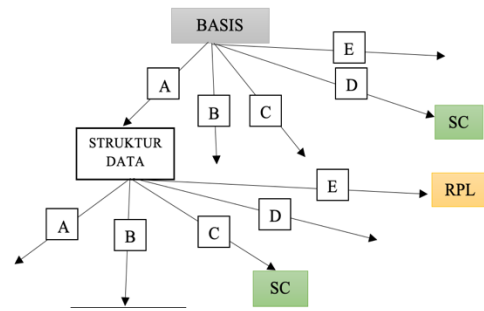
### III. HASIL DAN PEMBAHASAN

#### 3.1 Hasil Eksperimen

Setelah melewati beberapa tahapan dan proses, hasil dari penelitian adalah sebuah model dan akurasi dari data yang telah di proses dan diuji. Model atau pola tersebut yang dapat dijadikan klasifikasi untuk pemilihan bidang minat studi mahasiswa teknik informatika. Hasil dari penelitian diharapkan menghasilkan klasifikasi pemilihan bidang minat studi yang sesuai dan akurat.

##### 3.1.1 Eksperimen Algoritma C4.5

Pada Tahap ini data akan diuji hanya dengan menggunakan algoritma C4.5. Pengujian dengan mengklasifikasikan data sehingga menghasilkan model atau sebuah aturan tertentu (rule) untuk klasifikasi pemilihan bidang minat studi mahasiswa teknik informatika. Pengujian data sebanyak 300 record yang terdiri dari 12 atribut dan 1 label dengan merujuk pada perhitungan entropy dan informasi gain. Dari hasil perhitungan akan menghasilkan pohon keputusan yang merupakan sebuah aturan baru dari data agar lebih mudah untuk dipahami. Di bawah ini merupakan contoh pohon pendek dari Algoritma C4.5:



Gambar 2 Pohon Keputusan

##### 3.1.2 Eksperimen Algoritma C4.5 berbasis Forward Selection

Penggunaan *forward selection* pada penelitian ini untuk menyeleksi atribut yang tidak berpengaruh atau kurang relevan terhadap masalah serta diharapkan dapat meningkatkan hasil akurasi dari algoritma C4.5. Variabel input nilai mata kuliah yang telah ditentukan diantaranya yaitu Nilai Kalkulus I, Nilai Dasar Pemrograman, Nilai Pemrograman Berbasis Objek, Nilai Rekayasa Pereangkat Lunak, Nilai Basis Data, Nilai Logika Informatika, Nilai Alogirtma Pemograman, Nilai Matematika Diskrit, Nilai Struktur Data, Nilai Matriks & Ruang Vektor, Nilai Kalkulus II, Nilai Probabilitas dan Statistik. Dari data tersebut akan dilakukan inisialisasi terhadap atribut – atribut yang ada dalam data.

#### 3.2 Pengujian Metode

##### 3.2.1 Pengujian pada Algoritma C4.5

Untuk mengetahui bagaimana kinerja dari algoritma C4.5 dalam melakukan klasifikasi pada pemilihan bidang minat studi mahasiswa, maka dilakukan pengujian dengan confusion matrix. Dari dataset yang berjumlah 540 data mahasiswa teknik informatika terdiri dari 327 bidang minat SC dan 213 bidang minat RPLD. Dari 327 data bidang minat SC mahasiswa yang diklasifikasikan bidang minat SC adalah benar (TP) sejumlah 314 data dan mahasiswa yang diklasifikasikan bidang minat adalah salah (TN) sejumlah 13 data. Sedangkan dari 213 data bidang minat RPLD mahasiswa yang diklasifikasikan bidang minat RPLD adalah benar (FP) sejumlah 193 data dan mahasiswa yang diklasifikasikan bidang minat RPLD adalah salah (FN) sejumlah 20 data.

3.2.2 Pengujian pada Algoritma C4.5 berbasis Forward Selection

Untuk mengetahui bagaimana kinerja dari algoritma C4.5 berbasis forward selection dalam melakukan klasifikasi pada pemilihan bidang minat studi mahasiswa, maka dilakukan pengujian dengan confusion matrix,

Percobaan diatas dengan menggunakan atribut yang telah melewati proses seleksi. Atribut yang digunakan pada pengujian diatas adalah atribut yang dinilai berpengaruh untuk proses klasifikasi atribut tersebut diantaranya adalah Nilai Kalkulus I, Nilai Dasar Pemrograman, Nilai Rekayasa Pereangkat Lunak, Nilai Basis Data, Nilai Logika Informatika, Nilai Alogirtma Pemograman, Nilai Matriks & Ruang Vektor, Nilai Kalkulus II, Nilai Probabilitas dan Statistik. Hasil dari percobaan dengan algoritma C4.5 berbasis forward selection membuktikan bahwa metode tersebut dapat meningkatkan akurasi dari algoritma C4.5 tanpa forward selection yang semula 84.33% meningkat menjadi 85.00%.

3.3 Evaluasi dan Validasi

Pada tahapan ini dilakukan evaluasi dan validasi untuk menguji algoritma yang telah diterapkan dan terpilih dengan metode evaluasi k-fold cross validation yakni dengan membentuk k subset dari data set yang ada. Metode tersebut diawali dengan membagi data sejumlah k-fold sesuai ketetapan peneliti. Dalam prosesnya data akan menggunakan 600 data dengan penerapan metode adaboost pada algoritma C4.5 berbasis forward selection.

Pada tahap ini data akan dibagi menjadi 10 buah iterasi dengan ukuran yang sama. Selanjutnya proses training dan testing dilakukan sebanyak 10 kali dengan perbandingan 9 banding 1 pada data training dan testing. Metode 10-fold cross validation akan bekerja dengan melakukan pengujian yang diulang sebanyak 10 kali dan hasil pengukuran berupa akurasi rata-rata dari 10 kali pengujian.

3.3.1 Pengujian menggunakan k-fold cross validation

Tahapan awal dari metode 10-fold cross validation adalah membagi 600 data set menjadi 10 subset (bagian). Hasil pembagian 600 yaitu masing – masing subset berjumlah 60 data. Fold pertama terdiri dari kombinasi data 9 subset berbeda yang digabungkan menjadi satu dan dijadikan sebagai data training. Sedangkan 1 subset sisanya berperan sebagai data testing. Kemudian proses dari training dan testing

dilakukan hingga fold ke 10. Tabel pembagian data training dan data testing metode 10-fold cross validation seperti pada tabel dibawah:

TABEL 1 METODE 10-FOLD CROSS VALIDATION

Fold	Data Training		Data Testing		Akurasi
	Subset	Data	Subset	Data	
1	S2, S3, S4, S5, S6, S7, S8, S9, S10	540	S1	60	83.89%
2	S1, S3, S4, S5, S6, S7, S8, S9, S10	540	S2	60	84.07%
3	S1, S2, S4, S5, S6, S7, S8, S9, S10	540	S3	60	86.11%
4	S1, S2, S3, S5, S6, S7, S8, S9, S10	540	S4	60	85.00%
5	S1, S2, S3, S4, S6, S7, S8, S9, S10	540	S5	60	83.89%
6	S1, S2, S3, S4, S5, S7, S8, S9, S10	540	S6	60	84.44%
7	S1, S2, S3, S4, S5, S6, S8, S9, S10	540	S7	60	85.93%
8	S1, S2, S3, S4, S5, S6, S7, S9, S10	540	S8	60	87.59%
9	S1, S2, S3, S4, S5, S6, S7, S8, S10	540	S9	60	84.26%
10	S1, S2, S3, S4, S5, S6, S7, S8, S9	540	S10	60	84.26%
Rata - rata					85,00%

Pada tabel terlihat bahwa nilai akurasi yang didapatkan pada tiap fold stabil. Nilai tertinggi terdapat pada fold ke 8 sebesar 87.89%. akurasi yang didapat berdasarkan hasil perhitungan confusion matrix. Dibawah ini merupakan contoh confusion matrix dari akurasi tertinggi.

TABEL 2 HASIL PERHITUNGAN CONFUSION MATRIX

	True IF	True RPLD	Class precision
Pred. IF	312	52	85,71%
Pred. RPLD	15	161	91,48%
Class recall	95,41%	75,59%	

Dari hasil akurasi yang dihasilkan, dapat dianalisis mengenai performa dari penerapan algoritma C4.5 berbasis forward selection untuk proses klasifikasi pemilihan bidang minat mahasiswa teknik informatika. Algoritma yang diterapkan tersebut sudah dapat dikatakan baik dengan menghasilkan nilai akurasi rata – rata sebesar 85% yang diuji menggunakan pengujian 10-fold cross validation pada 540 data training dan 60 data testing.

3.4 Analisis Pembahasan

Dalam penelitian ini dilakukan klasifikasi pemilihan bidang minat studi Mahasiswa Teknik Informatika dengan Algoritma C4.5 dan C4.5 berbasis forward selection. Dengan menggunakan data dan atribut yang sama yaitu 540 sebagai data training, telah di dapat metode terbaik pada percobaan ini adalah dengan menggunakan algoritma C4.5 berbasis forward selection. Hasil dari penerapan C4.5 berbasis forward selection

menghasilkan akurasi terbesar dengan peningkatan akurasi.

Tujuan dari penelitian ini yaitu untuk meningkatkan nilai akurasi yang dihasilkan oleh algoritma klasifikasi C4.5 dengan mengurangi beberapa atribut pada dataset menggunakan forward selection. Dataset diolah dengan menggunakan perbandingan dari hasil yang didapatkan antara algoritma C4.5 dan C4.5 berbasis forward selection.

Dari perbandingan metode dengan data yang sama, terlihat eksperimen atau percobaan menggunakan metode C4.5 berbasis forward selection dengan 540 data training menghasilkan akurasi terbesar yaitu 85%. Dengan menggunakan metode evaluasi 10-fold cross validation yang digunakan untuk mengetahui rata – rata keberhasilan nilai akurasi yang didapat valid atau tidak, serta mengukur performa dari algoritma didapatkan hasil akurasi rata – rata dari 10 kali iterasi pengujian sebesar 85%.

#### IV. KESIMPULAN

Kesimpulan yang didapatkan berdasarkan hasil penerapan algoritma C4.5 dan C4.5 berbasis forward selection untuk klasifikasi pemilihan bidang minat studi mahasiswa teknik informatika sebagai berikut:

1. Dengan menggunakan seleksi fitur terbukti bahwa metode C4.5 berbasis forward selection dapat meningkatkan akurasi dari algoritma C4.5, dimana atribut yang terlibat semula pada dataset berkurang dengan hanya memilih atribut yang berpengaruh.
2. Hasil pengujian dari kinerja algoritma C4.5 dan C4.5 berbasis forward selection dengan 540 data menghasilkan akurasi tertinggi pada algoritma C4.5 berbasis forward selection dengan hasil akurasi 85% dengan meningkatkan akurasi pada metode algoritma C4.5 yang sebelumnya 84,33% dan C4.5 berbasis forward selection 85%.
3. Hasil dari penerapan algoritma C4.5 berbasis forward selection dikatakan baik dan dapat diterapkan pada kasus ini dengan data mahasiswa teknik informatika.

#### DAFTAR PUSTAKA

- [1] Musrofi M. (2010). Melesatkan Prestasi Akademik Siswa. Yogyakarta: Pedagogia.
- [2] S. Liliana, "Penerapan Algoritma C4.5 untuk penentuan Informatika dan Komputer (STMIK) Indonesia," *Jurnal GEMA AKTUALITA*, vol. 2, no. 1, Juni 2013. Sukardi, Dewa Ketut. (2003). Analisis Tes Psikologi. Jakarta: Rineka Cipta.
- [3] F. A. Prastya, "Penerapan Algoritma C4.5 untuk Prediksi Jurusan Siswa SMAN 3 Rembang," *Skripsi Teknik Informatika Universitas Dian Nuswantoro Semarang*, 2015.
- [4] Z. Quan, G. Lin-gang, W. Chong-jun, W.-j. and C. Shi-fu, "Using An Improved C4.5 for Imbalanced Dataset of Intrusion," *National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093*.
- [5] D. H. Kamagi and S. Hansun, Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa, Tangerang, 2014, pp. 2085-4552.
- [6] P. Kalapatapu, S. Goli, P. Arthum and A. Malapati, "A Study On Feature Selection and Classification Techniques Indian Music," *Procedia Computer Science*, no. 98, p. 125 – 131, 2016.
- [7] M. Al-Sarem, "Building a Decision Tree Model for Academic Advising Affairs Based on the Algorithm C4. 5," *International Journal of Computer Science Issues*, vol. 12, no. 5, pp. 1694-0814, 2015.
- [8] W. Supriyatin , Kusri and A. Amborowati, "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa," *INFORMA Politeknik Indonusa Surakarta*, vol. 1, pp. 2442-7942, 2016.
- [9] Y. S. Nugroho and I. Graha, "Data Mining Jasa Pengiriman Titipan Kilat di PT Citra Van dengan Metode Decision Tree," *Naskah Publikasi Program Studi Teknik Informatika Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta*, Juni 2014.
- [10] A. Syamsudin, "Kebutuhan Penelitian di Bidang Ilmu Pendidikan.," 2004.