

REKOMENDASI PENGAMBILAN MATA KULIAH PILIHAN UNTUK MAHASISWA SISTEM INFORMASI MENGGUNAKAN ALGORITME DECISION TREE

¹I Putu Pradnyana Iswara, ²Falih Farhan, ³Wira Kumara, ⁴Ahmad Afif Supianto

^{1,2,3}Sistem Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya

⁴Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya

Email: ¹putupradnyana@student.ub.ac.id, ²falihfarhan@student.ub.ac.id, ³kumara.wira@student.ub.ac.id, ⁴afif.supianto@ub.ac.id

(Naskah masuk: 29 Juni 2018, diterima untuk diterbitkan: 03 Mei 2019)

Abstrak

Program studi Sistem Informasi di Universitas Brawijaya belum terdapat konsentrasi ilmu secara umum, sehingga terkadang pemilihan mata kuliah pilihan oleh mahasiswa menjadi kurang tepat. Penelitian ini bertujuan untuk mengklasifikasikan konsentrasi ilmu tersebut dengan pendekatan data *mining* untuk selanjutnya dapat digunakan sebagai rekomendasi pemilihan mata kuliah pilihan oleh mahasiswa. Pada penelitian ini kami mengimplementasikan algoritme *decision tree* untuk memberikan rekomendasi kepada mahasiswa terkait keminatan apa yang lebih cocok diambil berdasarkan nilai-nilai mata kuliah prasyarat di semester-semester sebelumnya. Diharapkan dengan adanya rekomendasi ini semakin jelas pembatas antara disiplin ilmu yang ada pada Program Studi Sistem Informasi Universitas Brawijaya dimana terdapat 3 jenis jalur keminatan mata kuliah pilihan yaitu *Database*, Logika & pemrograman dan Manajemen Sistem Informasi/Teknologi Informasi. Data set yang terdiri dari data *training* dan data *testing* merupakan data akademik dari mahasiswa angkatan 2015 yang sudah mengambil mata kuliah pilihan, data target dari penelitian ini adalah data akademik mahasiswa angkatan 2016. Algoritme *decision tree* yang digunakan adalah *Rule Induction*, *CHAID*, *Random Forest*, dan *ID3*. Komposisi dari data *training* dan *testing* diubah-ubah untuk mengetahui pengaruh perubahan komposisi tersebut. Kelima algoritme tersebut diuji sebanyak 5 kali. Dari seluruh hasil pengujian didapatkan rata-rata akurasi dari kelima metode yang diusulkan berturut-turut adalah 66,48%, 67,49%, 80,62%, dan 86,90%. Hasil tersebut menunjukkan bahwa algoritme dengan rata-rata akurasi tertinggi dimiliki oleh algoritme *ID3* dikarenakan algoritmanya yang fleksibel dan pada kasus penelitian ini menjadi algoritme yang paling baik digunakan dibandingkan algoritme-algoritme yang lain.

Kata kunci: *decision tree*, *ID3*, *klasifikasi*, *komparasi*

ELECTIVE COURSES RECOMMENDATION OF INFORMATION SYSTEM'S STUDENTS USING DECISION TREE ALGORITHM

Abstract

The Information Systems study program at the University of Brawijaya does not yet have a concentration of knowledge in general, so sometimes the selection of elective courses by students becomes inappropriate. This study aims to classify the concentration of science with a data mining approach which can then be used as a recommendation for selecting elective courses by students. In this study, we implemented a decision tree algorithm to provide recommendations to students regarding what interests are more suitable to be taken based on the values of the prerequisite courses in previous semesters. It is hoped that this recommendation will make clearer the boundaries between the existing scientific disciplines in the UB Information System Study Program where there are three types of paths of elective courses, namely Database, Logic & Programming and Management of Information System / Information Technology. The data set consisting of training data, and testing data is academic data from 2015 class students who have taken elective courses, the target data from this study is the 2016 academic student data. The decision tree algorithm used is Rule Induction, CHAID, Random Forest, and ID3. The composition of training and testing data is changed to determine the effect of changes in the composition. The five algorithms were tested five times. From all the test results, the average accuracy of the five proposed methods is 66.48%, 67.49%, 80.62%, and 86.90% respectively. These results indicate that the ID3 algorithm owns the algorithm with the highest average accuracy because the algorithm is flexible and in the case of this study is the algorithm that is best used compared to other algorithms.

Keywords: *classification*, *comparation*, *decision tree*, *ID3*

1. PENDAHULUAN

Di usianya yang menyentuh enam tahun ini Program Studi (Prodi) Sistem Informasi sudah cukup terkenal baik secara akademis maupun kemahasiswaan di Indonesia khususnya di Universitas Brawijaya, dengan visi Menjadi program studi unggulan dalam pengembangan pendidikan dan pengetahuan di bidang Sistem Informasi di tingkat nasional dan internasional melalui integrasi Tri Dharma Perguruan Tinggi. Dengan mulai dipandanginya Prodi sistem informasi ini sebagai pilihan pertama atau kadang sebagai cadangan makin tinggi juga jumlah mahasiswa yang menjadi mahasiswa di Prodi ini. Prodi sistem informasi Universitas Brawijaya sudah cukup banyak memiliki alumni, kurikulumnya pun tiap empat tahun sekali diperbaharui mengikuti jaman sebagai usaha untuk menghasilkan sarjana-sarjana yang berdaya saing.

Meskipun sudah ada beberapa mata kuliah yang menjurus kepada disiplin bidang seperti pemrograman & logika, manajemen, dan *database*, kurikulum yang sedang digunakan oleh Prodi sistem informasi ini belum ada konsentrasi yang secara spesifik menggolongkan mahasiswa ke dalam bidang yang memang ingin ditekuninya. Tidak ada batas yang jelas antara kelompok bidang ilmu mata kuliah satu dengan yang lainnya, sehingga masih banyak mahasiswa yang asal-asalan dalam memilih mata kuliah karena mereka tidak paham dengan minatnya dan tidak ada aturan yang membatasi antar disiplin ilmu yang ada pada program studi sistem informasi.

Salah satu teknik yang dapat digunakan untuk mengatasi permasalahan tersebut adalah penerapan data *mining*, khususnya metode klasifikasi untuk membantu dalam pengambilan keputusan, seperti *Decision Tree* (baik yang jenis ID3, CHAID, *rule induction*, maupun *random forest*). Pada penelitian oleh Hanif (2016) bertujuan untuk mengetahui kualitas pelayanan kesehatan pada pasien BPJS, berdasarkan penelitian tersebut, algoritme yang memiliki akurasi paling tinggi setelah melakukan pengujian untuk membandingkan dengan 2 algoritme lainnya. Pada penelitian oleh Susanto (2014), algoritme CHAID adalah salah satu algoritme yang digunakan untuk memprediksi prestasi siswa berdasarkan faktor sosial ekonomi, motivasi, kedisiplinan serta prestasi masa lalu. Terakhir pada penelitian oleh Saefudin (2017), algoritme *random forest* menjadi salah satu metode yang digunakan untuk seleksi calon mahasiswa pada penerimaan mahasiswa baru di Universitas Pamulang, dari hasil pengujian untuk membandingkan akurasi algoritme pada penelitian tersebut algoritme *random forest* memiliki akurasi yang cukup tinggi dibandingkan algoritme lainnya. Selain itu, dalam skripsi yang ditulis oleh Nuraryudha (2014), ditunjukkan kegiatan penelitian yang bertujuan untuk mendapatkan rekomendasi mata kuliah menggunakan metode data *mining*.

Kedua penelitian tersebut sama-sama menggunakan nilai *confidence* yang dihasilkan dari proses *mining* untuk mendapatkan rekomendasi mata kuliah pilihan.

Di sisi lain, penelitian yang mencoba membandingkan beberapa algoritme telah diusulkan untuk melakukan prediksi keberhasilan pemasaran produk layanan perbankan (Dewi, 2016). Dalam penelitian tersebut terdapat lima algoritme yang diuji untuk mendapatkan akurasi tertinggi. Algoritme yang diuji pada penelitian tersebut sama dengan algoritme yang digunakan pada penelitian ini yaitu *Decision Tree*, yang lalu selanjutnya metode terpilih digunakan untuk proses prediksi tersebut. Penelitian perbandingan algoritme data *mining* pada bidang perbankan juga dilakukan oleh Oktanisa dkk. (2018) untuk melakukan klasifikasi pada permasalahan *Bank Direct Marketing*. Selain itu, dalam penelitian oleh Deny (2015) dan Usman (2016), sama-sama membandingkan tiga algoritme data *mining*, hanya saja tujuan penelitian yang berbeda, penelitian pertama bertujuan untuk menentukan strategi penjualan produk, sementara yang kedua bertujuan untuk menentukan klasifikasi hipertensi.

Untuk metode penentuan atribut pada pengujian akurasi algoritme, kami menggunakan acuan pada penelitian oleh Prasojo (2015). Dimana pada penelitian tersebut pada data yang akan digunakan, dipilih atribut NIM mahasiswa, mata kuliah serta nilai abjad dari mata kuliah tersebut. Dari nilai abjad masing-masing mata kuliah, maka dapat dilakukan klasifikasi serta mencari hubungan/asosiasi yang terdapat pada data tersebut sehingga didapatkan hasil *mining* yang selanjutnya dapat digunakan untuk rekomendasi.

Dari uraian penelitian-penelitian terdahulu tersebut di atas terlihat bahwa terdapat gap penelitian yang menjadi usulan penelitian ini. Penelitian ini melakukan rekomendasi pemilihan mata kuliah pilihan melalui perbandingan tingkat akurasi beberapa algoritme *decision tree* untuk menentukan algoritme terbaik dalam memberikan rekomendasi terkait pengambilan mata kuliah pilihan, khususnya untuk mahasiswa jurusan Sistem Informasi Oleh karena itu, pada penelitian ini ada beberapa hal yang ingin diangkat menjadi permasalahan, yaitu apakah hasil data *mining* menggunakan kelima algoritme ini dapat memberikan rekomendasi keminatan kepada mahasiswa, algoritme mana yang paling sesuai untuk kasus ini berdasarkan tingkat akurasinya.

Data yang digunakan dalam penelitian ini adalah data akademik dari mahasiswa Prodi Sistem Informasi angkatan 2015 sebagai data *training* dan data *testing*, serta angkatan 2016 sebagai target yang akan diklasifikasi keminatannya. Algoritme *decision tree* yang digunakan adalah algoritme ID3, karena berdasarkan hasil pengujian dari 5 metode yang berbeda, algoritme tersebut yang mendapatkan nilai akurasi tertinggi untuk penelitian ini. Dengan adanya

pemodelan dari klasifikasi menggunakan algoritme ID3 ini diharapkan dapat menjadi acuan bagi mahasiswa Prodi Sistem Informasi yang masih bimbang untuk menentukan mata kuliah pilihan yang sesuai dengan kapabilitasnya untuk diambil berdasarkan nilai mata kuliah wajib yang menjadi prasyarat mata kuliah pilihan.

2. DATA MINING

Data *mining* merupakan suatu proses untuk mengekstraksi lalu mengidentifikasi informasi dengan cara statistik, kecerdasan buatan, *machine learning*, maupun matematika (Riadi, 2017). Dengan adanya data *mining* maka didapatkan kemudahan untuk mencari informasi bisnis yang penting dari *database* yang jumlah datanya sangat besar. Adapun poin utama yang dapat dihasilkan dari data *mining* yaitu yang pertama prediksi tren, yang dapat berguna bagi proses bisnis yang dilakukan saat ini. Lalu selanjutnya data *mining* dapat digunakan untuk menemukan pola-pola unik dalam informasi. Terakhir, data *mining* dapat digunakan untuk membuat keputusan kritis dalam strategi.

2.1 Algoritme ID3

Algoritme ID3 merupakan algoritme yang dipergunakan untuk membangun sebuah *decision tree* atau pohon keputusan (Rokach, 2005). Algoritme ini ditemukan oleh Quinlan (1979), dengan memanfaatkan Teori Informasi atau *Information Teori* milik Shanon. ID3 sendiri merupakan singkatan dari *Iterative Dichotomiser 3*. Algoritme ID3 tidak dapat melakukan *backtracking* untuk merevisi keputusan pemilihan atribut yang telah dilakukan sebelumnya. ID3 hanya menangani nilai-nilai dari atribut yang sedikit dan diskrit. Persamaan (1) digunakan untuk mencari *entropy* per atribut, lalu persamaan (2) digunakan untuk mendapatkan gain per atribut.

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots\dots\dots(1)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots\dots\dots(2)$$

2.2 Algoritme Random Forest

Random Forest adalah sebuah algoritme yang ditemukan pertama oleh Tin Kam Ho, digunakan untuk klasifikasi, regresi dan tugas lainnya. Cara kerja dari *random forest* adalah membangun banyak *decision tree* saat *training time*, dan membuat keluaran kelas yang merupakan node kelas ataupun prediksi rata-rata dari masing-masing *tree* (Wu, 2009). *Random forest* tepat digunakan untuk *decision tree* yang mengalami *overfitting* pada *training* setnya. Berikut adalah persamaan dari *decision tree*. Persamaan (3) adalah untuk

menemukan rata-rata prediksi dari *tree* individual, lalu persamaan (4) adalah untuk menentukan standar deviasi dari prediksi *tree* individual.

$$f = \frac{1}{B} \sum_{b=1}^B f_b(x') \dots\dots\dots(3)$$

$$\sigma = \sqrt{\frac{\sum_{b=1}^B (f_b(x') - f)^2}{B-1}} \dots\dots\dots(4)$$

2.3 Algoritme CHAID

CHAID (*Chi-square Automatic Interaction Detection*) merupakan algoritme prediktif yang dapat digunakan menganalisis data dengan cara membentuk segmentasi yang membagi menjadi beberapa kelompok yang berbeda berdasarkan variabel independen yang mempengaruhi variabel respons dengan menggunakan statistik *chi-square*. Alur kerja CHAID adalah memeriksa setiap variabel independen dengan menggunakan perhitungan *chi-square* sehingga didapatkan variabel independen yang paling signifikan, sampai semua pembagian teridentifikasi(Liao, 2007). Persamaan (5) adalah perhitungan statistik *chi-square*.

$$X_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \dots\dots\dots(5)$$

2.4 Rule Induction

Rule Induction adalah sebuah algoritme pembelajaran yang dimana membentuk sebuah *rule*/aturan formal yang didapatkan dari serangkaian observasi terhadap data (Han, 2006). Aturan tersebut dapat mewakili model ilmiah dari sebuah data ataupun dapat memprediksi pola yang ada pada data. Berikut adalah persamaan yang digunakan, (6) adalah cara untuk mendapatkan nilai *support* dari item A, (7) adalah mencari nilai *support* dari kedua item, dan (8) untuk mendapatkan nilai *confidence*.

$$Support(A) = \frac{jumlah\ transaksi\ berisi\ A}{total\ transaksi} \dots\dots\dots(6)$$

$$Support(A, B) = P(A \cap B) \\ P(A \cap B) = \frac{jumlah\ transaksi\ berisi\ A\ \&\ B}{total\ transaksi} \dots\dots\dots(7)$$

$$Confidence(A \rightarrow B) = P(A|B) \\ P(A|B) = \frac{jumlah\ transaksi\ berisi\ A\ \&\ B}{jumlah\ transaksi\ berisi\ A} \dots\dots\dots(8)$$

3. METODE PENELITIAN

3.1 Data

3.1 Data Training dan Testing

Data ini digunakan sebagai acuan untuk proses *mining* dan pengujian. Data berupa sampel dari data akademik mahasiswa sistem informasi angkatan 2015. Data tersebut digunakan karena mahasiswa sistem informasi angkatan 2015 sudah mengambil

mata kuliah wajib yang dasar untuk penentuan keminatan, serta sudah banyak mengambil mata kuliah pilihan. Data ini memiliki atribut berupa :

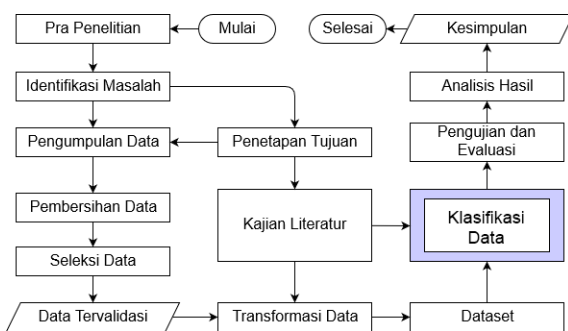
Tabel 1. Atribut Data *Training* dan *Testing*

No.	Nama Atribut	Penjelasan
1	NIM	Nilai induk Mahasiswa yang diposisikan sebagai id yang mendeskripsikan identitas mahasiswa
2	Nilai huruf mata kuliah DBD	Nilai huruf matakuliah DBD yang menjadi prasyarat dari keminatan "DATABASE"
3	Nilai huruf mata kuliah ABD	Nilai huruf matakuliah DBD yang menjadi prasyarat dari keminatan "DATABASE"
4	Nilai huruf mata kuliah ASD	Nilai huruf matakuliah DBD yang menjadi prasyarat dari keminatan "LOGIKA & PROG."
5	Nilai huruf mata kuliah ADSI	Nilai huruf matakuliah DBD yang menjadi prasyarat dari keminatan "LOGIKA & PROG."
6	Nilai huruf mata kuliah DPI	Nilai huruf matakuliah DBD yang menjadi prasyarat dari keminatan "LOGIKA & PROG."
7	Nilai huruf mata kuliah SFB	Nilai huruf matakuliah DBD yang menjadi prasyarat dari keminatan "MANAJEMEN SI/TI"
8	Nilai huruf mata kuliah PPB	Nilai huruf matakuliah DBD yang menjadi prasyarat dari keminatan "MANAJEMEN SI/TI"
9	Keminatan	Keminatan yang diambil oleh mahasiswa yang diposisikan menjadi label

3.2 Data Target

Data berupa sampel dari data akademik mahasiswa sistem informasi angkatan 2016. Data ini memiliki atribut yang sama seperti data *training* dan *testing*, namun belum memiliki label atau kategori yang berisikan keminatan yang diambil karena angkatan 2016 belum menentukan keminatan.

3.2 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Dalam pelaksanaan penelitian ini, kegiatan dilakukan sesuai dengan Gambar 1. Secara garis besar penelitian ini terdiri dari pra-penelitian, pengumpulan dan pengolahan data, kajian literatur, implementasi metode klasifikasi, serta analisis hasil dan kesimpulan penelitian. Pada tahap Pra

penelitian, kami melaksanakan konsultasi dengan dosen pembimbing berdasarkan permasalahan yang kami temukan, lalu dilakukan identifikasi permasalahan dan menetapkan tujuan penelitian. Agar lebih akurat, kami melakukan kajian literatur dengan menggunakan jurnal atau *paper* penelitian sebelumnya yang berhubungan dengan penelitian kami. Pada tahap pengolahan ada beberapa kegiatan sesuai dengan tahapan yang ada pada data *mining*, yaitu pembersihan data (*cleansing*), seleksi data, transformasi data, dan pembentukan data set yang dalam penelitian akan digunakan sebagai data *training* dan data *testing*. Dan terakhir setelah melakukan proses *mining*, maka hasil dapat dianalisis dan ditarik kesimpulan untuk mendapatkan rekomendasi.

4. HASIL DAN PEMBAHASAN

4.1 Persiapan Data

Pada tahapan ini, data *testing* dan data *training* berasal dari data sampel mahasiswa Sistem Informasi angkatan 2015 berjumlah 100. Data ini telah melewati proses pembersihan (*cleansing*), yaitu menghilangkan data yang memiliki atribut tidak lengkap, dalam hal ini data dihilangkan jika atribut mata kuliah wajib yang telah ditentukan tidak ada. Dari data tersebut sudah keminatan mahasiswanya sudah ditentukan.

4.2 Pemodelan dengan *RapidMiner*

Untuk melakukan proses *mining*, maka kami menggunakan *tools* *RapidMiner*. Dalam *RapidMiner* sudah tersedia komponen-komponen yang diperlukan dalam proses *mining*. Untuk melakukan proses *mining* data, maka harus dilakukan pemodelan yaitu dengan menggabungkan beberapa komponen tersebut agar terjadi proses sehingga dapat mengeluarkan *output* berupa hasil *mining*.

4.3 Pengujian

Pengujian ini bertujuan untuk mengetahui akurasi kerja dengan menggunakan *Confusion Matrix* pada masing masing algoritme *decision tree* yaitu *ID3*, *Random Forest*, *CHAID* dan *Rule Induction* dalam proses mengklasifikasikan data ke dalam kelas yang telah ditentukan, sehingga didapatkan algoritme yang memiliki akurasi kerja tertinggi pada penelitian ini, lalu digunakan untuk menguji data target. Pada uji coba ini, diberikan data *training* untuk membentuk model atau aturan dari masing-masing algoritme dan data *testing* untuk menguji model atau aturan yang terbentuk.

Unjuk kerja diperoleh dengan memberikan nilai pada *confusion matrix* pada keempat algoritme tersebut untuk menghitung nilai *precision*, *recall*, dan *accuracy* dari hasil pengujian sehingga dapat menentukan algoritme terbaik yang akan digunakan pada proses pemberian rekomendasi keminatan mahasiswa.

Dalam proses pengujian, terdapat 5 percobaan yang di setiap percobaannya akan menggunakan data *testing* dan *training* dengan komposisi yang berbeda, dijabarkan pada tabel 2 berikut.

Tabel 2. Komposisi Data untuk Skenario Percobaan

Skenario	Data Training	Data Testing
P-1	60	40
P-2	65	35
P-3	70	30
P-4	75	25
P-5	80	20

Pada tabel 2, ditunjukkan bahwa pada setiap skenario percobaan secara berurutan, komposisi data *training* bertambah lima, sementara data *testing* berkurang lima. Setiap komposisi berjumlah total 100 data. Hal ini ditujukan untuk mengetahui pengaruh perubahan komposisi data *training* dan data *testing* terhadap tingkat akurasi.

Tabel 3. Confusion Matrix dengan Algoritme ID3

Percobaan	Precision	Recall	Accuration
1	71,11	69,59	72,50
2	82,04	80,85	80
3	90,77	89,22	90
4	80,85	88,89	92
5	100	100	100

Tabel 3 adalah hasil nilai *precision*, *recall*, dan *accuration* untuk pengujian algoritme ID3. Dilakukan lima percobaan berdasarkan komposisi data pada Tabel 2 skenario P-1.

Tabel 4. Confusion Matrix dengan Algoritme Random Forest

Percobaan	Precision	Recall	Accuration
1	81,40	76,24	77,50
2	74,34	75,17	74
3	84,24	84,26	83
4	84,26	84,13	84
5	84,26	84,13	84

Tabel 4 adalah hasil nilai *precision*, *recall*, dan *accuration* untuk pengujian algoritme *Random Forest*. Dilakukan lima percobaan berdasarkan komposisi data Tabel 2 skenario P-1.

Tabel 5. Confusion Matrix dengan Algoritme CHAID

Percobaan	Precision	Recall	Accuration
1	65,67	67,53	70,00
2	74,34	66,71	66
3	69,32	70,37	70
4	64,54	65,64	64
5	69,32	70,37	70

Tabel 5 adalah hasil nilai *precision*, *recall*, dan *accuration* untuk pengujian algoritme CHAID. Dilakukan lima percobaan berdasarkan komposisi data Tabel 2 skenario P-1.

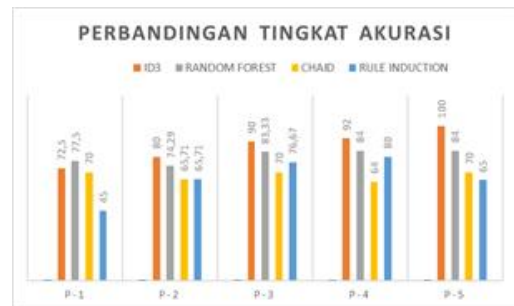
Tabel 6. Confusion Matrix dengan Algoritme Rule Induction

Percobaan	Precision	Recall	Accuration
1	47,00	43,67	45,00
2	64,78	64,74	66
3	69,32	74,07	77
4	75,66	79,37	80
5	80,33	68,33	65,00

Tabel 6 adalah hasil nilai *precision*, *recall*, dan *accuration* untuk pengujian algoritme *Rule Induction*. Dilakukan lima percobaan berdasarkan komposisi data Tabel 2 skenario P-1. Selanjutnya kami melakukan proses pada skenario 2 – 5 dengan hasil yang dibahas pada subbab selanjutnya.

4.3 Pembahasan

Berdasarkan dari pengujian masing-masing metode yang diusulkan untuk semua skenario pengujian maka dapat dirangkumkan hasilnya seperti pada grafik berikut :



Gambar 2. Perbandingan akurasi dari masing-masing algoritme pada setiap percobaan

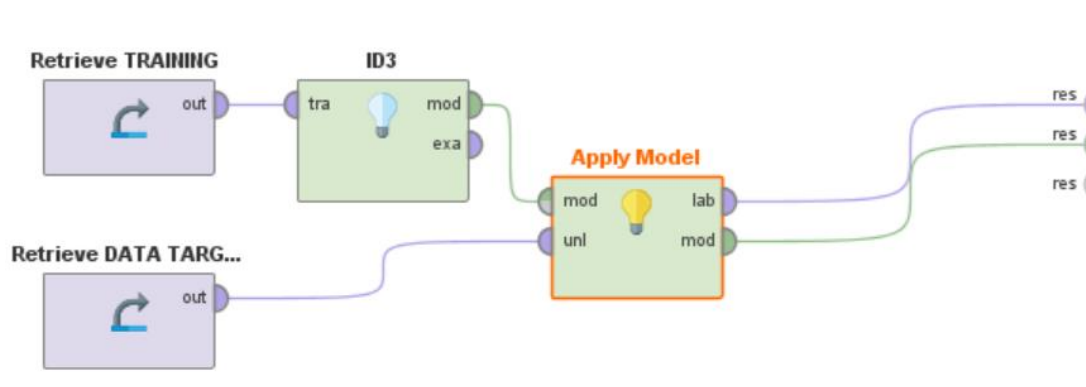
Gambar 2 adalah grafik yang merangkum hasil nilai akurasi semua algoritme per percobaan. P-1 adalah percobaan ke-1 dengan komposisi data *training* : data testing yaitu 60 : 40, algoritme yang akurasinya tertinggi adalah algoritme *Random Forest*. P-2 adalah percobaan ke-2 dengan komposisi 65 : 35, algoritme yang akurasinya tertinggi adalah algoritme ID3. P-3 adalah percobaan ke-3 dengan komposisi 70 : 30, algoritme yang akurasinya tertinggi adalah algoritme ID3. P-4 adalah percobaan ke-4 dengan komposisi 75 : 25, algoritme yang akurasinya tertinggi adalah algoritme ID3. Dan P-5 adalah percobaan ke-5 dengan komposisi 80 : 20, algoritme yang akurasinya tertinggi adalah algoritme ID3.



Gambar 3. Perbandingan rata-rata akurasi dari masing-masing algoritme pada setiap percobaan

Gambar 3 menunjukkan rata-rata akurasi dari setiap algoritme, dan terlihat bahwa algoritme ID3 memiliki rata-rata akurasi tertinggi. Melihat informasi dari kedua grafik diatas dapat disimpulkan bahwa algoritme ID3 memiliki unjuk kerja yang lebih baik dibandingkan dengan empat algoritme lainnya. Terbukti dengan akurasi tertinggi

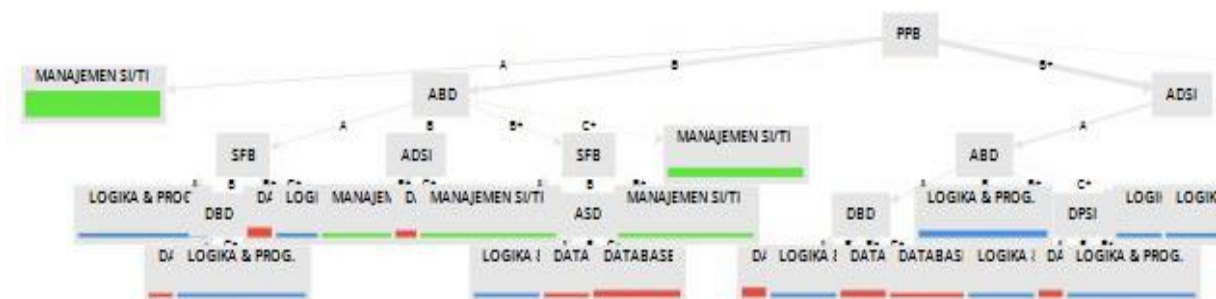
menyentuh angka 100%, tingkat akurasi terendah adalah 74,29%, dan memiliki rata-rata tingkat akurasi yang sebesar 86,9%. Sehingga dapat ditentukan bahwa algoritme yang paling sesuai untuk proses klasifikasi prediksi keminatan pada data target adalah dengan menggunakan algoritme ID3 dan menggunakan data *training* pada percobaan ke-5.



Gambar 4. Pemodelan ID3 pada RapidMiner saat pengujian

Gambar 4 menunjukkan struktur komponen yang digunakan untuk pemodelan pada *RapidMiner*. Bagian **Retrieve Training** adalah *input* data dari data *training*, sementara **Retrieve Data Target** adalah *input* dari data target yang akan diklasifikasi. Untuk

data *training* akan diujikan pada komponen algoritme **ID3**, dan hasilnya akan diterapkan / **apply** pada data target sehingga didapatkan hasil klasifikasi.

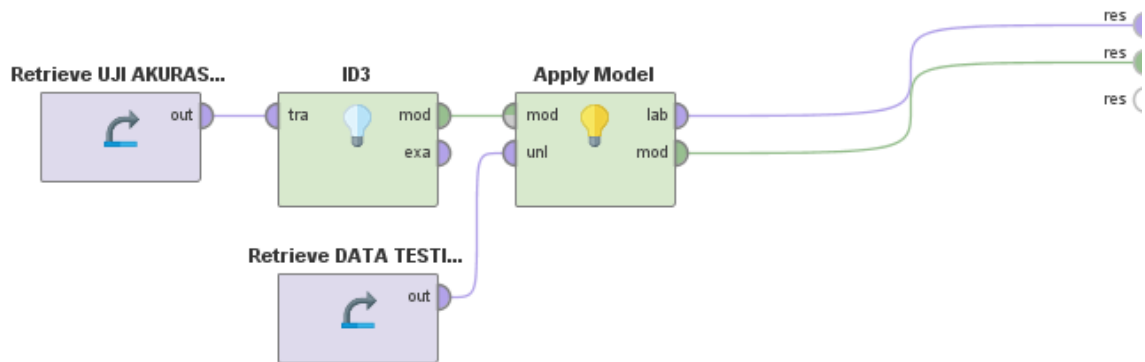


Gambar 5. Bentuk Tree yang dihasilkan algoritme ID3

Gambar 5 diperlihatkan model *tree* yang dihasilkan oleh algoritme ID3. “PPB” menjadi *root* karena mendapatkan nilai *gain* tertinggi sehingga paling berpengaruh untuk menentukan keminatan mahasiswa, sebagian besar mahasiswa yang mendapatkan nilai PPB A pasti tergolong keminatan Manajemen SI/TI. Begitu pula selanjutnya ABD menjadi komponen yang paling berpengaruh setelah

PPB, dan jika mendapatkan nilai A maka kemungkinan besar tergolong keminatan *database*.

Data target berupa sampel mahasiswa angkatan 2016 yang belum menentukan keminatan dan belum mengambil mata kuliah pilihan. Data ini belum memiliki kelas seperti pada data *training* dan data *testing*.

Gambar 6. Pemodelan ID3 pada *Rapid Miner* terhadap data target

Gambar 6 menunjukkan struktur komponen yang digunakan untuk pemodelan algoritme ID3 pada *Rapid Miner*. Input berupa data **Uji Akurasi** sebagai data *training* yang akan menjadi *rule* untuk klasifikasi, lalu **Data Testing** adalah data yang

diujikan terhadap algoritme. Untuk data *training* akan diujikan pada komponen algoritme **ID3**, dan hasilnya akan diterapkan / *apply* pada data target sehingga didapatkan hasil klasifikasi.

Row No.	NIM	prediction(M...	confidence(...	confidence(...	confidence(...
1	2016-774	DATABASE	1	0	0

Gambar 7. Klasifikasi algoritme *ID3* pada data target

Gambar 7 menunjukkan tabel hasil klasifikasi algoritme ID3. Terdapat 6 kolom. Kolom pertama dari kiri menunjukkan **nomor** dari baris data, lalu kolom **NIM** untuk identitas mahasiswa. Kolom **Prediction** menunjukkan hasil klasifikasi yang didapatkan berdasarkan nilai *confidence* tertinggi dari kelas yang ada. 3 kolom terakhir adalah nilai **confidence** dari setiap kelas yang ada (*database, manajemen SI/TI, logika&pemrograman*). Berdasarkan algoritme *ID3* mahasiswa dengan NIM 2016-774 direkomendasikan untuk mengambil keminatan "DATABASE" berdasarkan analisis pada data riwayat mata kuliah prasyarat yang telah ditempuh.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah kami laksanakan dengan menggunakan 4 algoritme decision tree yaitu *ID3*, *CHAID*, *random forest*, dan *rule induction*, kami dapat amati bahwa, semakin banyak data *training* maka cenderung nilai akurasi pada tiap-tiap pengujian semakin tinggi. Selain itu, komposisi data *training* mempengaruhi *precision*, *recall* dan *accuracy* dengan berubahnya nilai yang ditunjukkan pada hasil pengujian. Hasil pengujian pada data akademik mahasiswa angkatan 2015 dan 2016 juga menunjukkan bahwa penggunaan algoritme *ID3* menghasilkan nilai *precision* tertinggi pada percobaan ketiga dengan *recall*, dan *accuracy* masing-masing 81%, 82%, 80%, 48%, dan 80%.

DAFTAR PUSTAKA

- DENY, ARGA. 2015. *Perbandingan 3 Metode Dalam Data Mining Untuk Menentukan Strategi Penjualan Produk Makanan dan Minuman pada Toserba Lestari Baru Gelombang*. Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta
- DEWI, SARI. 2016. *Komparasi 5 Metode Algoritme Klasifikasi Data Mining pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan*. Jurnal Techno Nusa Mandiri. 8(1)
- HAN J., MICHELINE KAMBER. 2006. *Data Mining : Concepts and Techniques Second Edition*. USA: Elsevier nc.
- HANIF, TIWIK. 2016. *Perbandingan 3 Metode Data Mining Dalam Pencarian Pengetahuan Kualitas Pelayanan Kesehatan Bagi Pasien BPJS*. Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta
- LIAO. 2007. *Recent Advances in Data Mining of Enterprise Data: Algorithms and Application*. Singapore : World Scientific Publishing

- NURARYUDHA, DWIAJI. 2014. Analisis dan Implementasi Data Mining Dengan Continuous Association Rule Mining Algorithm (CARMA) Untuk Rekomendasi Mata Kuliah pada Perwalian. Seminar Nasional Ilmu Komputasi & Teknik Informatika. Fakultas Informatika Telkom University
- ROKACH, LIOR., ODED MAIMON.2005.Data Mining with Decision Tree. World Scientific Publishing Co. Pte. Ltd. Series in Machine Perception Artificial Intelligence. 69(1)
- OCHTARIO, RAYMOND,. SRI KARNILA. 2013. Sistem Data Mining Untuk Mengetahui Tingkat Kecenderungan Memilih Menu Makanan Dengan Metode Assosiation Rule Mining. Jurnal Informatika, 13(1)
- OKTANISA, I. AND SUPIANTO, A.A., 2018. Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK) Vol. 5, pp.567-576.
- PRASOJO, SATRIO., SHAUFIAH, 2015. Rekomendasi Pengambilan Mata Kuliah Pilihan Menggunakan Recursive Elimination Algorithm (Relim). E-Proceeding of Engineering, Vol. 2(1)
- RIADI, MUCHLISIN. 2017. Pengertian, Fungsi, Proses, dan Tahapan Data Mining. <https://www.kajianpustaka.com/2017/09/data-mining.html>. Diakses pada 9 Mei 2018
- SAIFUDIN, ARIES. 2017. Metode Data Mining untuk Seleksi Calon Mahasiswa pada Penerimaan Mahasiswa Baru di Universitas Pamulang. Jurnal Teknologi Universitas Muhammadiyah Jakarta. 10(1)
- SUSANTO, HERI. 2014. Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan Dan Prestasi Masa Lalu. Jurnal Pendidikan Vokasi, Vol. 4(2)
- USMAN, TAUFIQ. 2016. Perbandingan 3 Metode Data Mining Untuk Penentuan Hipertensi Di Rumah Sakit Umum Daerah Dr. Moewardi Surakarta. Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta
- WU, XINDONG,. VIPIN KUMAR.2009.The Top Ten Algorithms in Data Mining. Boca Raton: CRC Press