

# Prediksi Kelulusan Peserta Uji Kompetensi Profesi Menggunakan Algoritma C4.5

Anggit Wirasto <sup>1✉</sup>, Iis Setiawan Mangku Negara <sup>2)</sup>

<sup>1)</sup> *Fakultas Sains dan Teknologi, Universitas Harapan Bangsa, Banyumas, Indonesia*  
anggitwirasto@uhb.ac.id

<sup>2)</sup> *Fakultas Sains dan Teknologi, Universitas Harapan Bangsa, Banyumas, Indonesia*  
iissmn@uhb.ac.id

**Abstract**—Competence Test for Dentistry Student (Uji Kompetensi Mahasiswa Program Profesi Dokter Gigi – UKMP2DG) conducted periodically as exit exam for every dentistry student. Failed exam participants were permitted to retake the test if they have not exceeded the permitted study duration. But the longer they failed the exam may affects their performance as well. This research aimed to build a model to predict the result of UKMP2DG on each participant to be, given attributes such as GPA, origin of university, and sex. Using C4.5 algorithm, the resulted model has accuracy of 81.40%, precision of 82.84%, and recall value of 95.82%. This research result can be used by the related department in the universities to conduct special programs for their students that indicated to be failed at the exit exam.

**Keywords**—classification, machine learning, C4.5 algorithm, competence test

**Intisari**—Uji Kompetensi Mahasiswa Program Profesi Dokter Gigi (UKMP2DG) merupakan *exit exam* bagi para mahasiswa yang telah selesai menempuh pendidikan profesi dokter gigi. Meski peserta yang tidak lulus UKMP2DG dapat mengulang ujian di periode-periode berikutnya, para *retaker* ini dihadapkan pada batasan masa studi maksimal sebelum dinyatakan *drop out*. Selain itu semakin lama jeda antara penyelesaian studi dan waktu ujian dikhawatirkan berpengaruh pada penguasaan terhadap ilmu yang telah dipelajari. Penelitian ini bertujuan membangun model yang dapat digunakan untuk memprediksi kelulusan calon peserta UKMP2DG melalui data-data pendukung. Penerapan algoritma C4.5 menghasilkan model dengan tingkat akurasi 81,40%, presisi 82,84% dan recall sebesar 95,82%. Hasil penelitian ini dapat dimanfaatkan oleh Institusi Pendidikan Dokter Gigi (IPDG) untuk mengambil langkah-langkah khusus dalam mempersiapkan calon peserta yang diprediksi akan tidak lulus dalam ujian yang akan ditempuh.

**Kata kunci** — klasifikasi, machine learning, algoritma C4.5, kelulusan uji kompetensi

## I. PENDAHULUAN

Sistem pelayanan kesehatan agar bisa berfungsi dengan baik dan optimal perlu didukung oleh banyak faktor. Salah satu faktor terpenting tersebut adalah tenaga kesehatan yang kompeten dan berdedikasi. Dalam rangka meningkatkan

ketersediaan tenaga kesehatan yang berkualitas dan kompeten untuk menjalankan sistem pelayanan kesehatan dibutuhkan peningkatan kualitas pendidikan tenaga kesehatan.

Uji Kompetensi Mahasiswa Program Profesi Dokter Gigi (UKMP2DG) merupakan salah satu upaya yang dilakukan pemerintah, dalam hal ini melalui Kementerian Pendidikan dan Kebudayaan, untuk menjaga mutu, kualitas dan kompetensi tenaga kesehatan. Hal ini disebutkan dalam Undang-undang Nomor 36 Tahun 2014 Pasal 21 Ayat 7 tentang Tenaga Kesehatan. UKMP2DG diperuntukkan bagi mahasiswa program profesi dokter gigi yang telah menempuh pendidikan profesi dan menyelesaikan seluruh pekerjaan klinik beserta ujiannya. UKMP2DG berlaku sebagai salah satu syarat kelulusan atau *exit exam*.

UKMP2DG di Indonesia telah dilaksanakan sejak tahun 2014 secara periodik, sebanyak empat atau lima kali dalam satu tahun. Ujian yang diselenggarakan adalah Uji Teori dalam bentuk *Computer Based Test* (CBT) dan Uji Praktik dalam bentuk *Objective Structured Clinical Examination* (OSCE). Peserta harus lulus dalam kedua jenis ujian tersebut agar dapat dinyatakan kompeten. Peserta yang tidak lulus pada satu atau kedua jenis ujian dapat mengulang jenis ujian tersebut pada periode-periode berikutnya selama masih dalam batas toleransi studi.

Persentase kelulusan peserta dalam uji teori di periode yang sudah berlangsung berkisar antara 70% sampai dengan 85% setiap periodenya. Meskipun peserta yang tidak lulus memiliki kesempatan untuk mengulang pada periode berikutnya, namun hal ini perlu diperhatikan sebab semakin lama peserta mengikuti UKMP2DG maka semakin panjang pula jeda waktu antara peserta tersebut belajar di program studi profesi dengan waktu ujian. Dikhawatirkan ingatan dan pemahaman terhadap teori-teori yang didapatkan selama masa studi semakin lama semakin berkurang dan mempengaruhi performa peserta dalam mengerjakan soal uji teori UKMP2DG.

Peserta yang tidak lulus setelah berkali-kali mengikuti ujian juga terancam tidak bisa lulus akibat telah habis toleransi masa studinya. Tentunya hal ini berakibat tidak baik bagi mahasiswa itu sendiri yang akhirnya dianggap tidak mampu

menyelesaikan studi, serta bagi IPDG akan mempengaruhi tingkat kelulusan dan *drop out* mahasiswanya.

Penelitian ini bertujuan untuk mengolah data kepesertaan UKMP2DG selama beberapa periode dalam rangka menghasilkan model *machine learning* menggunakan algoritma C4.5 yang dapat memprediksi kelulusan calon peserta berdasarkan data-data pendukung. Prediksi ini dapat digunakan untuk memutuskan upaya bimbingan khusus terhadap para calon peserta yang diprediksi tidak lulus. Dengan demikian persentase kelulusan UKMP2DG di tiap periodenya diharapkan akan naik.

Beberapa penelitian terkait menggunakan algoritma C4.5 telah dilakukan antara lain untuk memprediksi keberhasilan mahasiswa dalam studi berdasarkan beberapa faktor seperti jumlah buku yang dipinjam di perpustakaan, lamanya waktu yang dihabiskan di perpustakaan, lamanya waktu yang dihabiskan di asrama, rerata waktu yang dihabiskan untuk perjalanan ke kampus, dan beberapa faktor lain [1]. Hasil penelitian tersebut memperlihatkan akurasi sebesar 80,84% dan disimpulkan dapat digunakan oleh bagian kemahasiswaan untuk membantu membimbing mahasiswa.

Penelitian lainnya dilakukan oleh Kusdaryono dan Nugroho [2] untuk memprediksi masa studi mahasiswa Akuntansi pada STIE Muhammadiyah Jakarta. Penelitian ini dilakukan untuk membuat sebuah aplikasi yang dapat melakukan prediksi terhadap lama studi dan predikat kelulusan mahasiswa dengan menerapkan teknik data mining dengan menggunakan Algoritma C4.5. Langkah-langkah yang dilakukan pada penelitian ini yaitu dengan mengambil data induk kelulusan mahasiswa dan data data kelulusan sehingga dapat diketahui informasi tentang lama studi dan predikat kelulusan mahasiswa. Pada penelitian ini mahasiswa akan diklasifikasikan dengan menggunakan algoritma klasifikasi C4.5. Sehingga hasil kesimpulan dari penelitian tersebut adalah pohon klasifikasi yang dijadikan sebagai informasi yang strategis yang terkait dengan lama studi mahasiswa dan predikat kelulusan.

Penelitian berikutnya dilakukan oleh Hartanto Kamagi David [3] dari program studi Teknik Informatika Universitas Multimedia Nusantara yang menerapkan Algoritma C4.5 kedalam implementasi program dekstop untuk memprediksi masa studi mahasiswa. Penelitian ini menggunakan klasifikasi dengan data IPK, IPS, dan jenis kelamin mahasiswa. Pada penelitian ini menggunakan Algoritma C4.5 dalam menentukan prediksi kelulusan berdasarkan atribut jenis kelamin, asal sekolah, dan IPS semester 1 sampai dengan 6. Hasil dari prediksi kelulusan dari prediksi penelitian ini dapat membantu program studi untuk mengetahui kelulusan mahasiswa menggunakan aplikasi Desktop dengan persentase 87.5% dari 60 data training dan 40 data testing.

## II. METODOLOGI PENELITIAN

### A. Data Mining

*Data Mining* merupakan salah satu bidang keilmuan yang berkembang dengan sangat cepat dikarenakan tingginya kebutuhan untuk mendapatkan nilai tambah dari basis data berskala besar, dalam rangka membantu pengambilan keputusan. Tujuan dari data mining adalah untuk mengetahui pola universal dari data yang ada [4].

### B. Decision Tree

*Decision tree* atau sering disebut dengan pohon keputusan merupakan pohon yang biasanya digunakan untuk mendapatkan jawaban keputusan dari suatu masalah. Dalam pohon keputusan dibagi menjadi dua bentuk yaitu berupa pohon biner dan tidak biner. Pohon berbentuk biner jika semua fitur dalam suatu data set tersebut menggunakan dua macam kategorikal, sedangkan yang termasuk bukan biner yaitu jika dalam fitur memiliki lebih dari dua macam kategorikal. *Decision tree* banyak digunakan untuk menyelesaikan masalah seperti untuk menentukan suatu keputusan dalam berbagai bidang contohnya dalam bidang diagnosis, struktur data, dan lain-lain. Keuntungan menggunakan pohon keputusan ini karena prosedurnya dapat diprediksi [4].

Algoritma *decision tree* yang banyak digunakan dalam teknik *data mining* antara lain ID3, C4.5, dan *Random Forest* [5]. Aplikasi penggunaan algoritma *decision tree* antara lain dilakukan oleh Robert Burduk [6] dalam dunia medis untuk diagnosa penyakit perut bagian atas. Moon, Sung Seek [7] menggunakan algoritma *decision tree* dalam melakukan klasifikasi untuk menemukan relasi antara jumlah rokok yang dikonsumsi per hari dengan umur mulai merokok, tingkat pendidikan, dan stress psikologis.

### C. Algoritma C4.5

Algoritma C4.5 pertama kali diperkenalkan oleh Quinlan untuk menghasilkan model klasifikasi [8]. Klasifikasi adalah teknik *data mining* yang dapat digunakan untuk memprediksi keanggotaan suatu data ke dalam kelas-kelas [9]. Algoritma C4.5 adalah algoritma klasifikasi *supervised machine learning* paling mendasar yang sangat luas digunakan dan umumnya memberikan hasil yang bagus [10].

Algoritma C4.5 merupakan perbaikan dari algoritma ID3 [11] dengan peningkatan pada kecepatan dan efisiensi penggunaan memori. Perbedaannya yaitu pada ID3 hanya bisa dilakukan pada fitur yang mempunyai tipe kategorikal (nominal atau ordinal) dan tipe numerik (interval atau rasio) tidak bisa digunakan. Berdasarkan masalah tersebut maka perbaikan dilakukan pada Algoritma C4.5 yang dapat menangani fitur bertipe numerik, melakukan pemotongan pohon keputusan, dan penurunan rule set pada data yang digunakan. Pada Algoritma C4.5 perhitungan yang dilakukan yaitu menggunakan nilai *gain* dalam menentukan fitur menjadi pemecah node atau akar pada pohon keputusan yang diinduksi [4].

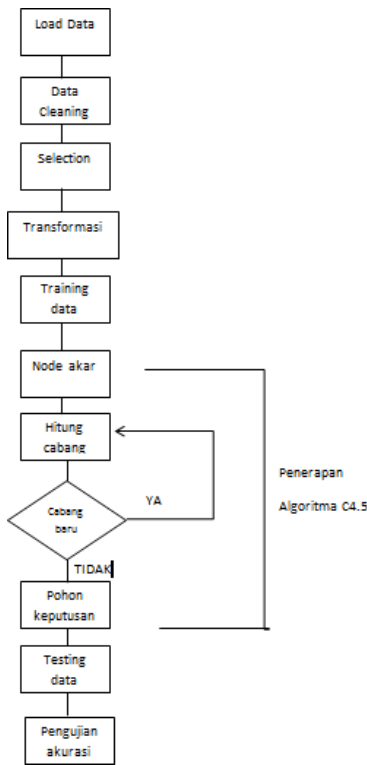
### D. Obyek Penelitian

Obyek penelitian ini adalah riwayat kepesertaan UKMP2DG dari berbagai Institusi Pendidikan Dokter Gigi (IPDG) dari tahun 2014 sampai dengan 2021. Data yang tersedia adalah nama, tanggal lahir, jenis kelamin, asal IPDG, IPK, ujian keberapa, dan status kelulusan Uji Teori.

### E. Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari database pendaftaran UKMP2DG, yaitu tabel data kepesertaan yang memuat informasi peserta, ujian yang diikuti, serta kelulusan ujian tersebut. Data yang ada tersedia mulai dari periode pertama pelaksanaan UKMP2DG periode 1 tahun 2014 sampai tahun 2021.

F. Tahapan Data Mining



Gambar 1. Tahapan Data Mining

1) Load data

Proses load data merupakan proses awal pada tahapan penelitian. Data yang telah dikumpulkan berupa file dalam format MS Excel kemudian di-load ke dalam program untuk diolah.

2) Pembersihan data

Data yang telah dikumpulkan cenderung memiliki beberapa data yang tidak sesuai format dan tidak konsisten. Contohnya data dengan nilai IPK kosong lebih dari 4,00. Data-data tersebut dibuang dari dataset karena akan mempengaruhi model yang dihasilkan.

3) Seleksi data

Tidak semua atribut pada data yang dikumpulkan akan digunakan. Dilakukan seleksi pada atribut-atribut yang ada agar hasil yang lebih akurat bisa didapatkan. Atribut yang digunakan adalah asal IPDG, IPK, dan ujian keberapa sebagai atribut bebas, serta kelulusan Uji Teori sebagai atribut tak bebas.

4) Transformasi data

Data atribut dengan tipe string atau alfabet perlu diubah ke dalam bentuk numerik (angka).

5) Training data

Pemodelan menggunakan algoritma C4.5 dilakukan pada data-data yang dikelompokkan sebagai training, dan mencari node atau akar sampai cabang terakhir tidak bisa dihitung lagi.

6) Node akar

Tahap ini merupakan tahap terpenting karena data yang sebelumnya telah dipersiapkan akan dimining menggunakan algoritma C4.5. Ada tiga macam *node* pada *decision tree*, yaitu *root node* (akar), *internal*

*node* (cabang), dan *leaf node* (daun). Tahap pertama adalah dengan menentukan *node* akar yang akan digunakan untuk percabangan. *Node* akar adalah *node* tertinggi, tidak memiliki input dan tidak dapat memiliki lebih dari satu output. Atribut yang dipilih sebagai akar didasarkan pada perhitungan nilai *gain* tertinggi dari setiap atribut yang ada. Formulasi perhitungan *gain* adalah sebagai berikut:

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

Dengan keterangan:

- S: sampel
- A: atribut
- n: jumlah partisi himpunan atribut
- |S<sub>i</sub>|: jumlah sampel dalam S

Sedangkan nilai *entropy* dapat dihitung melalui formulasi berikut:

$$Entropy(S) = \sum_{i=1}^n n - p_i \log_2 p_i \tag{2}$$

Dengan keterangan:

- S: sampel
- n: jumlah partisi
- p<sub>i</sub>: jumlah yang bersolusi positif pada kriteria tertentu

7) Menghitung cabang

*Node* kedua adalah *node* cabang, *node* ini hanya boleh memiliki satu input dan sedikitnya dua output. Aturan dalam membuat cabang baru atau *node* dilakukan jika pada cabang sebelumnya masih bisa dilakukan perhitungan lebih lanjut. Cabang atau *node* baru tersebut dilakukan perhitungan sampai tahap akhir menemukan keputusan, *node* terakhir ini disebut sebagai *node* daun yang hanya memiliki satu input dan tidak memiliki output apapun.

8) Pohon keputusan

Pohon keputusan merupakan hasil analisis pemecahan masalah pengambilan keputusan mengenai kemungkinan atau probabilitas yang akan mempengaruhi keputusan tersebut. Hasil dari pohon keputusan diperoleh dari perhitungan cabang atau *node* dan setiap simpul daun menandai label kelas. Proses dalam pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon (tree) kemudian mengubah model pohon tersebut menjadi aturan (rule).

9) Pengujian akurasi

Dalam melakukan pengujian tingkat akurasi menggunakan tabel ConfusionMatrix, ada pengukuran kinerja menggunakan confusion matrix, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah True Positive (TP), TrueNegative (TN), False Positive (FP) dan False Negative (FN). Nilai True Negative (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan False Positive (FP)

merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, True Positive (TP) merupakan data positif yang terdeteksi benar. False Negative (FN) merupakan kebalikan dari True Positive, sehingga data positif, namun terdeteksi sebagai data negatif.

Berdasarkan nilai True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP) dapat diperoleh nilai akurasi, presisi dan recall. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan Persamaan 3. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan Persamaan 4. Sementara itu, recall menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. Nilai recall diperoleh dengan Persamaan 5.

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

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

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

- a. TP adalah True Positive, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- b. TN adalah True Negative, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- c. FN adalah False Negative, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
- d. FP adalah False Positive, yaitu jumlah data positif namun terklasifikasi salah oleh sistem

### III. HASIL DAN PEMBAHASAN

#### A. Hasil Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data kepesertaan UKMP2DG mahasiswa program profesi dari berbagai IPDG mulai tahun 2014 sampai dengan 2021. Data mentah yang berhasil dikumpulkan sebanyak 19.895 data peserta Uji Teori. Hasil load data diperlihatkan pada Gambar 2. Data tersebut memiliki atribut-atribut sebagai berikut:

##### 1. Asal IPDG

Asal IPDG menunjukkan dari Universitas mana peserta ujian tersebut berasal. Terdapat 31 IPDG yang berbeda. Atribut asal IPDG ini berupa kode Perguruan Tinggi dan Program Studi sesuai kodefikasi yang ada pada Pangkalan Data Perguruan Tinggi.

##### 2. Jenis kelamin

Jenis kelamin memiliki dua kategori, yaitu Laki-laki dan Perempuan.

	kompetensi_kode_pt_ps	retaker	ipk	jenis_kelamin	kelulusan
5396	100312901	0	3.08	P	Pass
9324	101112901	0	3.31	L	Pass
1420	10100912901	1	2.73	L	Failed
5720	10100912901	0	3.10	P	Pass
6572	101112901	2	3.16	P	Failed

Gambar 2. Hasil load data

##### 3. IPK

Atribut IPK berisi data IPK terakhir peserta selama menempuh studi di program profesi dokter gigi. Data IPK memiliki tipe data kontinyu mulai dari IPK 0 sampai dengan IPK tertinggi yaitu 4,00.

##### 4. Retaker

Retaker menyatakan apakah pada periode tersebut peserta mengikuti ujian untuk pertama kalinya (*first taker*, atau *retaker* ke-0) atau sudah mengulang selama beberapa kali.

##### 5. Kelulusan

Atribut kelulusan menyatakan hasil dari Uji Teori yang ditempuh peserta pada periode tersebut, apakah lulus (*pass*) atau gagal (*failed*).

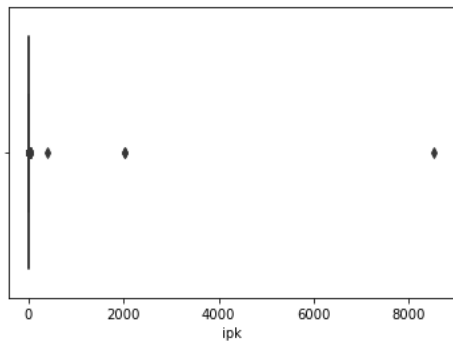
#### B. Pre-processing (Data Selection, Cleansing, Transformation)

Proses selection dilakukan untuk mengambil komponen apa saja yang akan dipakai dalam penelitian ini. Data yang diputuskan untuk dipakai adalah asal IPDG, IPK, *retaker*, jenis kelamin, dan kelulusan. Asal IPDG dipilih sebagai penyeimbang untuk nilai IPK karena tidak semua IPDG memiliki standar yang sama dalam pemberian nilai. Selanjutnya dilakukan cleansing terhadap data untuk membuang data yang tidak valid dan tidak konsisten, seperti diperlihatkan pada Gambar 3. Data-data yang dibersihkan antara lain:

- Data IPK bernilai 0 dan yang lebih besar dari 4,00
- Data-data yang memiliki nilai kosong
- Data yang memiliki nilai kelulusan selain *pass* dan *failed*

Dari pembersihan data ini diperoleh dataset berisi 17.597 data. Selanjutnya data tersebut dibagi menjadi set data train dan test dengan proporsi 3:1 sehingga didapatkan set data train sebanyak 13.467 data dan set data tes sebanyak 4.130 data.

Tahap berikutnya adalah transformasi data untuk menyesuaikan atribut agar sesuai dengan proses data mining. Seperti terlihat pada Gambar 2 atribut yang menyatakan asal IPDG yaitu "kompetensi\_kode\_pt\_ps" sebenarnya adalah nilai yang menunjukkan kategori sehingga perlu disesuaikan. Hasil transformasi dapat dilihat di Gambar 4, data asal IPDG dikodekan ulang menjadi data kategori.



Gambar 3. Distribusi nilai pada atribut "ipk"

	ipdg_asal	retaker	ipk	jenis_kelamin	kelulusan
5396	IPDG-14	0	3.08	P	Pass
9324	IPDG-8	0	3.31	L	Pass
1420	IPDG-4	1	2.73	L	Failed
5720	IPDG-4	0	3.10	P	Pass
6572	IPDG-8	2	3.16	P	Failed

Gambar 4. Hasil transformasi data

C. Data Training (Penerapan C4.5)

Data training adalah proses pelatihan data yang digunakan untuk membuat pemodelan algoritma C4.5 atau membentuk sebuah model classifier. Proses data training dilakukan pada sebanyak 13.467 data.

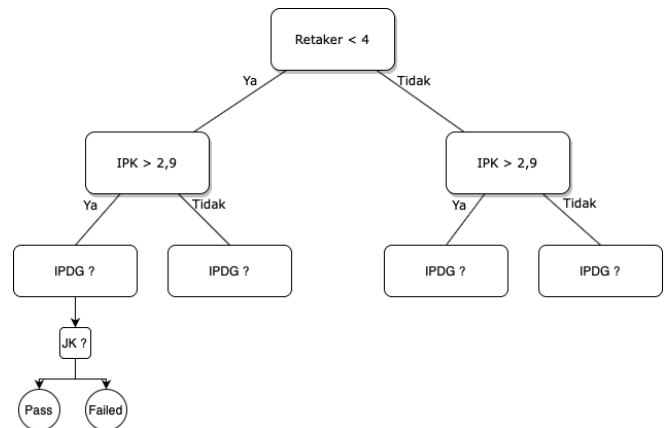
D. Decision Tree

Pemodelan menggunakan algoritma C4.5 menghasilkan decision tree sebagaimana diilustrasikan pada Gambar 5. Terlihat sebagai node akar adalah atribut retaker, kemudian node cabang berikutnya berturut-turut adalah IPK, asal IPDG, dan jenis kelamin.

Sedangkan feature importance dari masing-masing atribut yang digunakan adalah seperti ditunjukkan oleh Tabel 1. Atribut atau feature yang diketahui memiliki importance terbesar adalah "ipdg\_asal", diikuti oleh "ipk", "retaker", dan terakhir "jenis\_kelamin". Atribut "ipdg\_asal" memiliki nilai importance lebih besar dibandingkan atribut "ipk" diperkirakan karena setiap IPDG memiliki kualitas pembelajaran dan input mahasiswa yang berbeda. IPDG tertentu terutama yang berstatus sebagai PTN sebagian besar mahasiswanya didapatkan dari tes masuk yang sangat ketat dan persaingan tinggi sehingga bisa dipastikan memiliki mahasiswa dengan kualitas akademik yang baik.

feature	importance
ipdg_asal	0,991301
ipk	0,377562
retaker	-0,021998
jenis_kelamin	-0,346865

Tabel 1. Feature Importance



Gambar 5. Decision Tree dari Proses Algoritma C4.5

Atribut berikutnya dengan nilai importance positif adalah "ipk". Hal ini dikarenakan IPK merupakan parameter capaian pembelajaran seorang mahasiswa, sehingga pada umumnya mahasiswa dengan IPK tinggi memiliki kualitas akademik yang baik juga. Namun nilai importance atribut ini tidak terlalu tinggi sebab ada kemungkinan IPK yang cukup tinggi didapatkan setelah mahasiswa mengulang beberapa mata kuliah. Disamping itu, standar IPK antara satu IPDG dan IPDG lain dapat saja berbeda.

E. Data Testing

Data testing digunakan untuk menguji performa dan kebenaran terhadap korelasi dalam model yang didapatkan. Pengujian dilakukan terhadap 4.130 data dan menghasilkan Confusion Matrix seperti terlihat pada Tabel 2 sebagai berikut:

True Positive	3.062
False Positive	635
True Negative	299
False Negative	134

Tabel 2. Confusion Matrix Hasil Pengujian

Maka dari Confusion Matrix tersebut dapat dihitung hasil akurasi, presisi, dan recall sebagai berikut:

$$\text{Akurasi} = (3.062 + 299) / 13.467 \times 100\% = 81,40\%$$

$$\text{Presisi} = 3.062 / (3.062 + 635) \times 100\% = 82,84\%$$

$$\text{Recall} = 3.062 / (3.062 + 134) \times 100\% = 95,82\%$$

IV. KESIMPULAN

Akurasi sebesar 81,40% dapat disimpulkan sebagai hasil akurasi yang cukup baik karena berdasarkan pengujian confusion matrix dapat dikatakan baik jika tingkat akurasinya lebih dari 80%. Model yang dihasilkan dapat digunakan dalam memprediksi kelulusan mahasiswa calon peserta UKMP2DG. Hasil prediksi dapat dijadikan acuan bagi IPDG untuk memberikan bimbingan khusus kepada para calon peserta yang diprediksi akan tidak lulus.

Sebagai topik penelitian selanjutnya adalah mencoba mengklasifikasikan dataset yang digunakan dalam penelitian ini dengan menambahkan data nilai uji teori yang diperoleh setiap peserta. Dengan penambahan data tersebut model yang dihasilkan diharapkan dapat memberikan hasil yang lebih akurat. Selain itu perlu juga dilakukan data training menggunakan algoritma klasifikasi yang lain dan kemudian

hasilnya dibandingkan dengan hasil yang didapatkan menggunakan algoritma C4.5 pada penelitian ini.

#### REFERENSI

- [1] L. Li, S. M. Yao, Z. Ou, and Q. J. Chen, "Forecast of Student Achievement Variation Trend Based on C4.5 Decision Tree," in *Proceedings of the 2015 International Conference on Artificial Intelligence and Industrial Engineering*, 2015, pp. 383–386.
- [2] A. Kusdaryono and D. Nugroho, "Penerapan algoritma C4.5 untuk memprediksi masa studi kelulusan S1 Akuntansi pada STIE Mummadiyah Jakarta," *Tlicom*, vol. 5, no. 3, pp. 201–211, 2017.
- [3] H. K. David, "Implementasi Data Mining dengan Algoritma C4 . 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *ULTIMATICS*, vol. VI, no. 1, 2016.
- [4] E. Prasetyo, *Data mining - Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: ANDI OFFSET, 2014.
- [5] S. Sathyadevan and R. R. Nair, "Comparative Analysis of Decision Tree Algorithms: ID3, C4.5 and Random Forest," *Comput. Intell. Data Min.*, vol. 1, pp. 549–562, 2014.
- [6] R. Burduk and M. Wozniak, "Different Decision Tree Induction Strategies For a Medical Decision Problem," *Cent. Eur. J. Med.*, vol. 7, no. 1, pp. 193–193, 2012.
- [7] S. S. Moon, S.-Y. Kang, and W. Jitpitaklert, "Decision Tree Models for Characterizing Smoking Patterns of Older Adults," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 445–451, 2012.
- [8] D. P. I. M. Meena and D. V. Perumal, "Performance of C4.5 and Naive Bayes Algorithm to Predict Stomach Cancer - An analysis," *IJARCCCE*, vol. 5, no. 11, 2016, doi: 10.17148/ijarccce.2016.511103.
- [9] M. A. Muslim, A. J. Herowati, E. Sugiharti, and B. Prasetyo, "Application of the pessimistic pruning to increase the accuracy of C4.5 algorithm in diagnosing chronic kidney disease," in *Journal of Physics: Conference Series*, 2018, vol. 983, no. 1, doi: 10.1088/1742-6596/983/1/012062.
- [10] R. Jothikumar and R. V. Siva Balan, "C4.5 classification algorithm with back-track pruning for accurate prediction of heart disease," *Biomed. Res.*, vol. 2016, no. Special Issue 2, 2016.
- [11] Y. Kustiyahningsih and E. Rahmanita, "Aplikasi Sistem Pendukung Keputusan Menggunakan Algoritma C4.5. untuk Penjurusan SMA," *J. Simantec*, vol. 5, no. 2, 2016.