Penerapan Algoritma *Cart* Untuk Memprediksi Status Kelulusan Mahasiswa

Muhammad Faisal Amin
Program Studi Teknik Informatika, STMIK Banjarbaru
Jl. A. Yani Km. 33,3 Banjarbaru
Faisal.indonesia@gmail.com

Abstrak

Lulus tepat waktu merupakan salah satu elemen penilaian akreditasi perguruan tinggi. Dalam jenjang S1, mahasiswa dinyatakan lulus tepat waktu jika bisa menyelesaikan studinya <= delapan semester atau empat tahun. BAN-PT menetapkan standar kelulusan tepat waktu sebesar >= 50%. Jika standar tersebut tidak terpenuhi, maka akan mengurangi nilai akreditasi. STMIK Banjarbaru sebagai salah satu sekolah tinggi komputer terbesar di Kalimantan Selatan mengalami masalah tersebut. Tingkat kelulusan prodi teknik informatika (TI) STMIK Banjarbaru sangat rendah, yaitu hanya 18,83%. Masalah tersebut mendorong pihak prodi TI untuk melakukan evaluasi dan langkah strategis dalam upaya meningkatkan tingkat kelulusan mahasiswa agar target dari BAN-PT bisa tercapai. Untuk itu perlu diketahui secara dini pola dari mahasiswa yang cenderung tidak lulus tepat waktu. Pada penelitian ini diusulkan algoritma cart untuk memprediksi kelulusan mahasiswa. Algoritma ini akan mengolah dataset profil mahasiswa yang berjumlah 238 data. Dataset ini memiliki label status kelulusan. Nilai dari label bernilai kategorikal, yaitu tepat dan terlambat. Adapun fitur atau atribut yang digunakan, yaitu jenis kelamin, asal sekolah, nilai uan, nilai sekolah, dan umur. Hasil dari algoritma cart ini berupa model pohon keputusan yang sangat mudah untuk dianalisis. Bahkan, oleh orang awam sekalipun. Dari model ini akan dipetakan pola mahasiswa yang berpotensi lulus tepat waktu dan terlambat. Berdasarkan hasil eksperimen, akurasi dari model algoritma ini sebesar 64,37%.

Kata Kunci: Prediksi kelulusan mahasiswa, decision tree, algoritma cart

Abstract

Graduate on time is one element of college accreditation. In the bachelor degree, students graduated on time if it could complete his studies <= eight semesters or four years. BAN-PT sets the standard on time graduation of >= 50%. If these standards are not met, it will reduce the value of accreditation. STMIK Banjarbaru as one of the largest computing college in South Kalimantan experiencing the issue. Informatics engineering study program graduation rate (TI) STMIK Banjarbaru very low at only 18,83%. The issue prompted the study program IT to conduct evaluations and strategic step in efforts to increase the graduation rate of students for the target of BAN-PT can be achieved. It is necessary to know at an early pattern of students who are less likely to graduate on time. In this study, the proposed cart algorithm for predict students graduation. This algorithm will process the student profile datasets totaling 238 data. This dataset has a graduation status label. The value of label-value categorically, that is appropriate and overdue. The inputs are used, namely gender, origin of the school, uan grades, school grades, and age. The results of this cart algorithms such as decision tree models that are very easy to analyze. Moreever, for novices. From this model will be mapped patterns of potential students who graduate on time and late. Based on the experimental results, the accuracy of this algorithm model by 64,37%

Keywords-prediction, graduation, decision tree, cart algorithm

1. PENDAHULUAN

Lulus tepat waktu merupakan salah satu elemen penilaian akreditasi perguruan tinggi [1]. Dalam jenjang S1, mahasiswa dinyatakan lulus tepat waktu jika bisa menyelesaikan studinya <= delapan semester atau empat tahun. BAN-PT menetapkan standar kelulusan tepat waktu sebesar >= 50% [2]. Jika standar tersebut tidak terpenuhi, maka akan mengurangi nilai

akreditasi. Oleh karena itu, perguruan tinggi perlu memprediksi sedini mungkin performance mahasiswa yang berpotensi terlambat lulus.

Perguruan tinggi umumnya mempunyai database besar yang berisi data profil mahasiswa. Data-data tersebut dapat diolah sehingga menghasilkan pengetahuan dengan metode-metode dalam data mining. Untuk memprediksi ketepatan status keluluhan mahasiswa, fungsi data mining yang dipilih adalah klasifikasi. Hal ini karena data yang digunakan mempunyai label bertipe kategorikal.

Banyak penelitian mengusulkan metode klasifikasi data mining untuk memprediksi performace belajar siswa. Metode yang digunakan untuk prediksi performance mahasiswa adalah decision tree [3], naïve bayes [4], neural network [5]. Masing-masing metode menawarkan keunggulan masing-masing.

Algoritma pohon keputusan mempunyai keunggukan dibandingkan algoritma lain. Algoritma ini menghasilkan rule dalam bentuk pohon yang sangat mudah untuk dianalisis. Algoritma ini juga mendapatkan popularitas yang tinggi karena akurasinya yang tinggi [6]. Algoritma yang paling banyak digunakan untuk membangun pohon keputusan adalah ID3, C4.5 dan CART.

STMIK Banjarbaru merupakan salah satu perguruan tinggi komputer terbesar di Kalimantan Selatan. Jumlah mahasiswa pada tahun 2015 sebesar 1690 orang. Namun, tingkat kelulusan tepat waktu prodi teknik informatika(TI) sangat rendah. Berdasarkan laporan data kelulusan [7] mahasiswa yang lulus tepat waktu hanya 18,83%. Masalah tersebut mendorong pihak prodi TI untuk melakukan evaluasi dan langkah strategis dalam upaya meningkatkan tingkat kelulusan mahasiswa agar target dari BAN-PT bisa tercapai. Untuk itu perlu diketahui secara dini pola dari mahasiswa yang cenderung tidak lulus tepat waktu.

2. METODOLOGI

2.1 Perancangan Penelitian

2.1.1 Desain Riset

Penelitian ini menggunakan desain riset metode eksperimen. Metode ini menguji kebenaran sebuah hipotesis dengan statistik dan menghubungkannya dengan masalah penelitian [8]. Misalnya, menguji akurasi algoritma cart untuk memprediksi waktu kelulusan mahasiswa. Dalam penelitian ini, akan dilakukan eksperimen menggunakan algoritma cart terhadap data profil mahasiswa untuk memprediksi ketepatan waktu kelulusan mahasiswa. Hasil dari eksperimen ini adalah pengetahuan berupa pohon keputusan dan akurasi dari algoritma cart.

2.1.2 Pengumpulan Data

Pengumpulan data terbagi menjadi dua, yaitu pengumpulan data primer dan data sekunder. Data primer adalah data yang dikumpulkan pertama kali dan untuk melihat apa yang sesungguhnya terjadi. Data sekunder adalah data yang sebelumnya pernah dibuat oleh seseorang baik di terbitkan atau tidak. Pengumpulan data primer pada penelitian ini menggunakan metode observasi dan interview. Data yang dikumpulkan adalah data kelulusan mahasiswa dan data profil mahasiswa. Sedangkan dalam pengumpulan data sekunder menggunakan buku, jurnal, publikasi dan lain-lain.

2.2 Eksperimen

Eksperimen pada penelitian ini mengacu pada model *Cross-Standard Industry for Data Mining* (CRISPDM) [9]. CRISPDM terdiri dari 6 tahap, yaitu:

1. *Business Understanding*

   Pihak prodi TI STMIK Banjarbaru ingin mengetahui pola dari mahasiswa yang berpotensi lulus tepat waktu dan terlambat. Dengan diketahui pola tersebut secara lebih cepat, maka akan memudahkan pihak prodi menyiapkan strategi antisipasinya. Data yang akan diolah telah memiliki label berupa kategorikal, yaitu lulus tepat waktu dan lulus terlambat. Karena label berupa kategorikal, maka algoritma data mining yang dipilih adalah

jenis klasifikasi. Pada penelitian ini diusulkan untuk menerapkan algoritma klasifikasi *decision tree*. Algoritma *decision tree* yang dipilih adalah cart.

2. Data *Understanding*

Data kelulusan diambil dari STMIK Banjarbaru jurusan teknik informatika mulai tahun 2008 sampai 2015. Data ini terdiri dari 749 record. Terdiri dari 11 field, yaitu nama, nim, ipk, predikat, lama studi (bulan), lama studi (semester), lama tugas akhir (bulan), no.skep.yudisium, tgl yudisium, tahun akademik.

Di data ini belum tersedia label status ketepatan lulus. Label ini akan kita buat berdasarkan lama studi (semester). Jika lama studi < 8,5 semester maka tepat. Jika > 8.5, maka terlambat. Memang seharusnya lulus tepat waktu itu <= 8 semester. Ini karena ada beberapa mahasiswa yang lulus <= 8 semester tetapi yudisiumya di semester > 8. Jumlah record untuk label tepat dan terlambat masing-masing 141 (18,83%) dan 608 (81,17%).

3. Data *Preparation*

Tidak semua data kelulusan diolah menjadi data training. Hanya beberapa field yang diperlukan. *Field* yang tidak diperlukan, seperti nama, IPK, predikat, lama studi (bulan), lama tugas akhir (bulan), lama tugas akhir (semester), no.skep.yudisium, tanggal yudisium, dan tahun akademik. Dalam data kelulusan tidak ada informasi mengenai nilai uan dan nilai sekolah. Data nilai uan dan uas ini diambil dari data registrasi mahasiswa. Namun, banyak mahasiswa yang nilai uas dan uasnya kosong. Setelah dilakukan seleksi dan penggabungan data, diperoleh sekita 238 data yang akan menjadi data training.. Field yang digunakan sebagai atribut, yaitu nim, tanggal lahir, jenis kelamin, asal sekolah, nilai uan dan nilai sekolah. Field yang dijadikan label adalah status kelulusan.

4. *Modeling*

Tahap ini akan diterapkan algoritma cart pada data training untuk menghasilkan model berupa pohon keputusan. Adapun langkah-langkah untuk menghasilkan pohon keputusan dengan algoritma cart, sebagai berikut [10]:

1. Menyusun calon cabang (*candidate split*)

Penyusunan ini dilakukan terhadap seluruh atribut secara lengkap. Daftar yang berisi calon cabang disebut daftar calon cabang mutakhir

2. Menilai kinerja seluruh calon cabang yang ada pada calon cabang mutakhir

Nilai kinerja ditentukan dengan cara menghitung nilai besaran kesesuaian $\Phi(s|t)$ dan didefinisikan sebagai:

$$\Phi(s|t) = 2P_L P_R \sum_{j=1}^{jmlkategori} |P(j|t_L) - P(j|t_R)| \quad (1)$$

Keterangan:

t_L=calon cabang kiri dari noktah keputusan t

t_R=calon cabang kanan dari noktah keputusan t

P_L=Rasio jumlah record pada calon cabang kiri $t_L$dengan jumlah record training

P_R=Rasio jumlah record pada calon cabang kanan $t_R$dengan jumlah record training

$P(j|t_L)$=Rasio jumlah record berkategori j pada calon cabang kiri t_L

$P(j|t_R)$=Rasio jumlah record berkategori j pada calon cabang kanan t_R

Bila didefinisikan besaran dari persamaan tersebut akan didapat

$$\Phi(s|t) = \sum_{j=1}^{jumlah\ kategori} |P(j|t_L) - P(j|t_R)| \qquad (2)$$

Maka berdasarkan persamaan (1) akan didapatkan persamaan berikut

$$\Phi(s|t) = \sum_{j=1}^{jumlah\ kategori} 2P_L P_R \Phi(s|t) \qquad (3)$$

3. Menentukan calon cabang yang benar-benar akan menjadi cabang

Ini dipilih dengan cara memilih nilai Φ(s|t) terbesar. Setelah itu percabangan tersebut digambar. Jika tidak ada lagi noktah keputusan, eksekusi algoritma cart akan dihentikan. Namun, jika masih terdapat noktah keputusan, ulangi langkah kedua dengan terlebih dahulu membuang calon cabang yang telah menjadi cabang. Ini berarti telah diperoleh calon cabang mutakhir yang baru.

5. *Evaluation*

Pada tahap ini model yang dihasilkan dari algoritma cart akan diukur akurasinya. Metode yang digunakan adalah *cross validation*.

6. *Deployment*

Pada tahap ini model yang dihasilkan dari algoritma cart akan diterapkan untuk memprediksi status kelulusan mahasiswa.

3. HASIL DAN PEMBAHASAN

Untuk mendapatkan model algoritma cart dan akurasinya, maka data yang sudah valid akan diolah dengan metode *cross-validation*. Metode ini menghasilkan model dan akurasinya sekaligus. Terlebih dahulu hasil prediksi akan diringkas dalam suatu tabel. Tabel ini disebut dengan *confusion matrix*. Pada tabel 1ditampilkan *confusion matrix*. Berdasarkan eksperimen yang ditampilkan melalui confusion matrix, hasil akurasi dari model yang diusulkan sebesar 64,67%.

Tabel 1 *Confusion matrix*

	True tepat	True terlambat
Pred. tepat	31	34
Pred. terlambat	50	123

Dari model pohon keputusan ditemukan atribut-atribut yang menentukan kelulusan mahasiswa. Atribut yang menentukan seperti nilai sekolah, nilai uan, umur, dan jenis kelamin. Atribut yang menjadi root adalah nilai sekolah. Namun, terdapat atribut yang tidak digunakan, yaitu atribut asal sekolah.

Selain itu, dari model pohon keputusan ini ditemukan beberapa aturan/pengetahuan penting. Ada beberapa aturan yang dapat mengklasifikasikan secara penuh *class* kelulusan. Namun, ada juga beberapa aturan yang tidak dapat mengklasifikasikan secara penuh *class* kelulusan. Meskipun demikian, kecenderungan *class* mengarah ke mana tetap dapat terlihat. Pada gambar 1 ditampilkan model pohon keputusan dari algoritma cart. Berikut ini akan diuraikan aturan-aturan hasil ekstraksi pohon keputusan.

Nilai UAN > 49.075
| Nilai UAN > 60.250
| | Nilai Sekolah > 78.960
| | | Nilai UAN > 65.075
| | | | Nilai UAN > 88.350: Terlambat {Tepat=0, Terlambat=2}
| | | | Nilai UAN ≤ 88.350
| | | | | Jenis Kelamin = L
| | | | | | Umur > 17.500
| | | | | | | Nilai UAN > 77.715: Tepat {Tepat=7, Terlambat=0}
| | | | | | | Nilai UAN ≤ 77.715
| | | | | | | | Umur > 18.500: Tepat {Tepat=2, Terlambat=1}
| | | | | | | | Umur ≤ 18.500: Terlambat {Tepat=1, Terlambat=2}
| | | | | | Umur ≤ 17.500: Terlambat {Tepat=0, Terlambat=2}
| | | | | Jenis Kelamin = P: Tepat {Tepat=8, Terlambat=0}
| | | Nilai UAN ≤ 65.075: Terlambat {Tepat=0, Terlambat=3}
| | Nilai Sekolah ≤ 78.960
| | | Jenis Kelamin = L
| | | | Umur > 19.500
| | | | | Nilai UAN > 64
| | | | | | Nilai Sekolah > 66.325
| | | | | | | Nilai UAN > 75.325
| | | | | | | | Nilai Sekolah > 76.240: Tepat {Tepat=2, Terlambat=1}
| | | | | | | | Nilai Sekolah ≤ 76.240: Tepat {Tepat=2, Terlambat=0}
| | | | | | | Nilai UAN ≤ 75.325
| | | | | | | | Nilai Sekolah > 74.035: Terlambat {Tepat=0, Terlambat=3}
| | | | | | | | Nilai Sekolah ≤ 74.035
| | | | | | | | | Nilai Sekolah > 69.425
| | | | | | | | | | Nilai UAN > 67.140
| | | | | | | | | | | Nilai UAN > 69.825
| | | | | | | | | | | | Umur > 28: Terlambat {Tepat=0, Terlambat=2}
| | | | | | | | | | | | Umur ≤ 28: Tepat {Tepat=2, Terlambat=0}
| | | | | | | | | | | Nilai UAN ≤ 69.825: Terlambat {Tepat=0, Terlambat=2}
| | | | | | | | | | Nilai UAN ≤ 67.140: Tepat {Tepat=2, Terlambat=0}
| | | | | | | | | Nilai Sekolah ≤ 69.425: Terlambat {Tepat=0, Terlambat=2}
| | | | | | Nilai Sekolah ≤ 66.325: Tepat {Tepat=3, Terlambat=0}
| | | | | Nilai UAN ≤ 64: Terlambat {Tepat=0, Terlambat=4}
| | | | Umur ≤ 19.500
| | | | | Nilai UAN > 68.165

| | | | | | Nilai Sekolah > 74.155
| | | | | | | Nilai Sekolah > 75.140: Terlambat {Tepat=0, Terlambat=13}
| | | | | | | Nilai Sekolah ≤ 75.140
| | | | | | | | Nilai Sekolah > 74.565: Tepat {Tepat=3, Terlambat=0}
| | | | | | | | Nilai Sekolah ≤ 74.565: Terlambat {Tepat=1, Terlambat=1}
| | | | | | Nilai Sekolah ≤ 74.155
| | | | | | | Nilai UAN > 72.135: Terlambat {Tepat=0, Terlambat=22}
| | | | | | | Nilai UAN ≤ 72.135
| | | | | | | | Nilai UAN > 71.935: Terlambat {Tepat=1, Terlambat=1}
| | | | | | | | Nilai UAN ≤ 71.935: Terlambat {Tepat=0, Terlambat=8}
| | | | | Nilai UAN ≤ 68.165
| | | | | | Nilai Sekolah > 74.700: Tepat {Tepat=3, Terlambat=0}
| | | | | | Nilai Sekolah ≤ 74.700
| | | | | | | Nilai Sekolah > 67.300: Terlambat {Tepat=0, Terlambat=7}
| | | | | | | Nilai Sekolah ≤ 67.300: Tepat {Tepat=2, Terlambat=0}
| | | Jenis Kelamin = P
| | | | Nilai UAN > 69.940
| | | | | Nilai UAN > 80.500: Terlambat {Tepat=0, Terlambat=2}
| | | | | Nilai UAN ≤ 80.500
| | | | | | Nilai UAN > 71.205
| | | | | | | Nilai UAN > 76.235
| | | | | | | | Umur > 19.500: Terlambat {Tepat=1, Terlambat=2}
| | | | | | | | Umur ≤ 19.500: Tepat {Tepat=6, Terlambat=0}
| | | | | | | Nilai UAN ≤ 76.235
| | | | | | | | Nilai Sekolah > 76.045: Terlambat {Tepat=0, Terlambat=2}
| | | | | | | | Nilai Sekolah ≤ 76.045
| | | | | | | | | Umur > 18.500
| | | | | | | | | | Nilai UAN > 74.665: Terlambat {Tepat=0, Terlambat=2}
| | | | | | | | | | Nilai UAN ≤ 74.665: Tepat {Tepat=2, Terlambat=1}
| | | | | | | | | Umur ≤ 18.500: Tepat {Tepat=2, Terlambat=0}
| | | | | | Nilai UAN ≤ 71.205: Tepat {Tepat=4, Terlambat=0}
| | | | Nilai UAN ≤ 69.940
| | | | | Nilai Sekolah > 65.085

| | | | | | Nilai UAN > 68.750: Terlambat {Tepat=1, Terlambat=2}
| | | | | | Nilai UAN ≤ 68.750: Terlambat {Tepat=0, Terlambat=8}
| | | | | Nilai Sekolah ≤ 65.085: Tepat {Tepat=2, Terlambat=1}
| Nilai UAN ≤ 60.250

| | Nilai Sekolah > 73.815: Tepat {Tepat=7, Terlambat=0}

| | Nilai Sekolah ≤ 73.815

| | | Nilai UAN > 59.735: Tepat {Tepat=3, Terlambat=0}

| | | Nilai UAN ≤ 59.735

| | | | Nilai UAN > 54.490

| | | | | Nilai Sekolah > 68: Terlambat {Tepat=0, Terlambat=3}

| | | | | Nilai Sekolah ≤ 68: Terlambat {Tepat=1, Terlambat=1}

| | | | Nilai UAN ≤ 54.490

| | | | | Umur > 18.500: Tepat {Tepat=2, Terlambat=0}

| | | | | Umur ≤ 18.500: Tepat {Tepat=1, Terlambat=1}

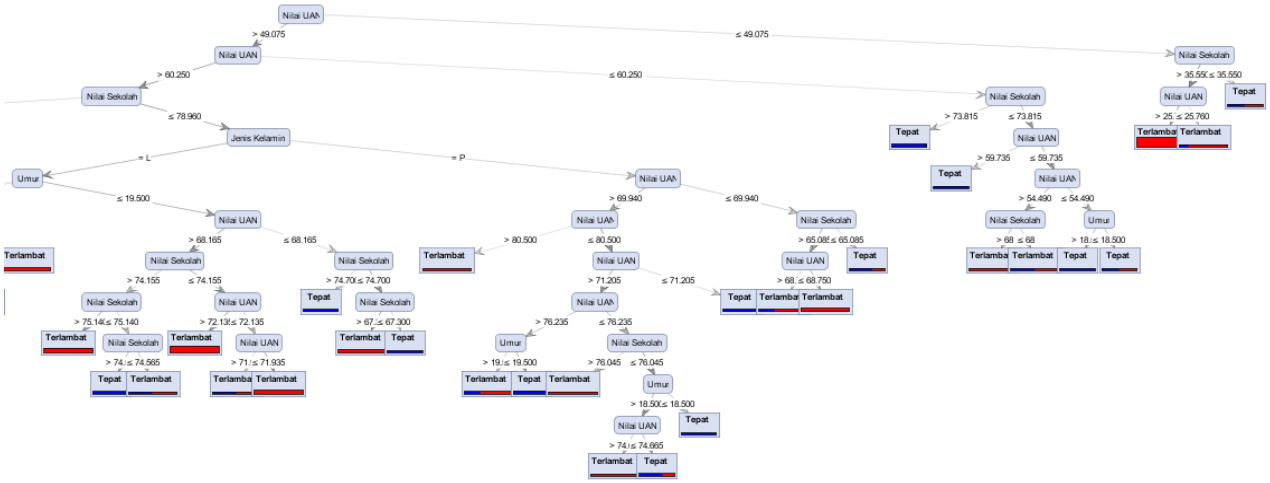
Nilai UAN ≤ 49.075

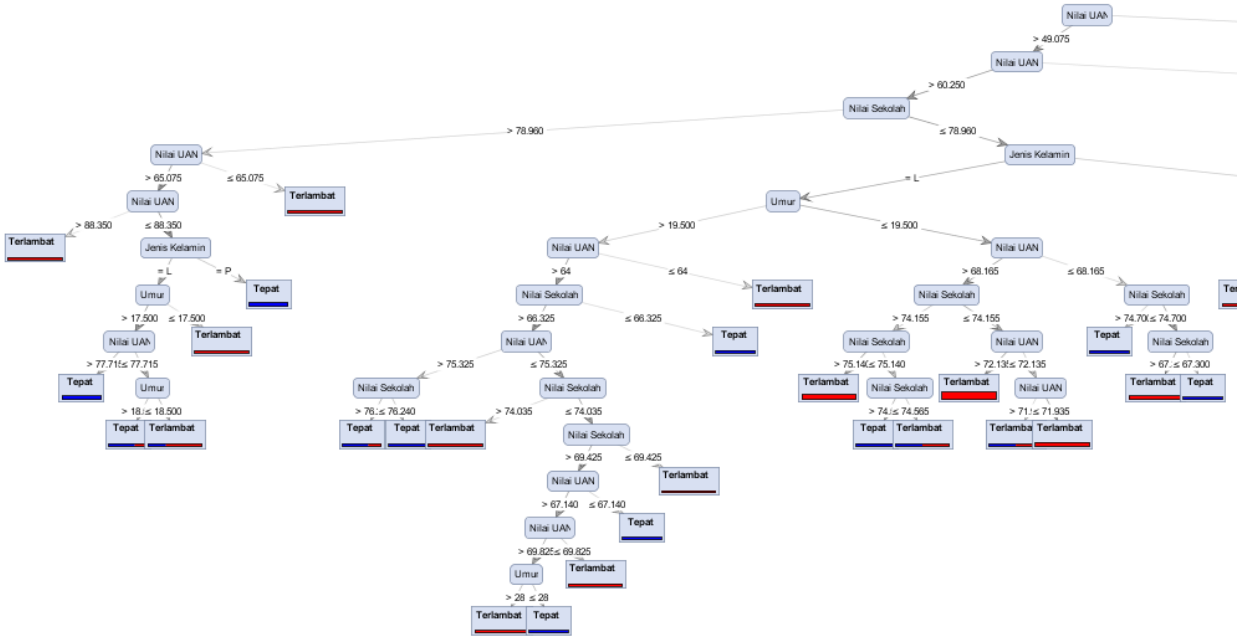
| Nilai Sekolah > 35.550

| | Nilai UAN > 25.760: Terlambat {Tepat=0, Terlambat=35}

| | Nilai UAN ≤ 25.760: Terlambat {Tepat=1, Terlambat=4}

| Nilai Sekolah ≤ 35.550: Tepat {Tepat=1, Terlambat=1}





Gambar 1 Model Pohon Keputusan algoritma cart

4. Kesimpulan

Berdasarkan hasil eksperimen algoritma cart terbukti cukup akurat dalam memprediksi kelulusan mahasiswa. Hasil akurasinya sebesar 64,37%. Diharapkan model pohon keputusan ini akan membantu pihak prodi TI mengambil keputusan yang lebih tepat dan cepat.

Hasil akurasi dari model algoritma cart masih perlu ditingkatkan. Beberapa cara yang bisa dilakukan, seperti menambah data training, menambah field atau atribut, mencoba algoritma klsifikasi yang lain, menggabungkan dengan algoritma lain, dsb.

Daftar Pustaka

[1] Anonim, *Buku 5 Pedoman Penilaian Instrumen Akreditasi Sarjana (Versi 23 Juli 2010).* BAN-PT. 2010.
[2] Anonim, *Buku-6-Matriks-Penilaian-Akreditasi-Sarjana.* BAN-PT. 2010.
[3] Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R., & Honrao, V. *Predicting Student's Performance Using ID3 And C4.5 Classification Algorithms*. International Journal of Data Mining & Knowledge Management Process *(IJDKP)*, 2013.
[4] Shaziya, H., Zaheer, R., & Kavitha, G. *Prediction of Students Performance in Semester Exams using a Naïve bayes Classifier*. International Journal of Innovative Research in Science, Engineering and Technology, 2015.
[5] Naser, S. A., Zaqout, I., Ghosh, M. A., Rasha, A., & Alajrami, E. *Predicting Student Performance Using Artificial Neural Network: in the Faculty of Engineering and Information Technology.* International Journal of Hybrid Information Technology, 221-228, 2015.
[6] Nithyasri, B., Nandhini, K., & Chandra, E*. Classification Techniques in Education Domain*. International Journal of Computer, Mathematical Sciences and Applications, 15-23, 2011.
[7] Anonim, *Laporan Data Kelulusan 2015.* Prodi TI Stmik Banjarbaru, Banjarbaru. 2015.
[8] Kothari, *Research Methology Methods and Techniques.* Research Methology Methods and Techniques, 2004.
[9] Larose, D. T., *Discovering Knowledge In Data.* New Jersey: Wiley, 2005.
[10] Susanto, S., & Suryadi, D., *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data.* Yogyakarta: Andi Publisher, 2010.