

Kinerja Algoritma *Classification and Regression Tree (Cart)* dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta

Nurul Indah Prabawati¹, Widodo², M. Ficky Duskarnaen³

^{1,2,3} Pendidikan Teknik Informatika dan Komputer Fakultas Teknik
Universitas Negeri Jakarta

¹nurul.indahp94@gmail.com, ²widodo@unj.ac.id, ³duskarnaen@unj.ac.id

Abstrak

Organisasi kemahasiswaan adalah fasilitas yang disediakan oleh perguruan tinggi sebagai wadah untuk mengembangkan kemampuan non akademis, minat dan bakat mahasiswa. Namun, dalam kenyataannya banyak mahasiswa yang mengikuti organisasi mengalami penurunan prestasi hingga tidak dapat lulus tepat waktu. Di Universitas Negeri Jakarta belum adanya sistem yang dapat mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi. Sebelum membangun sistem pengambilan keputusan, diperlukan penelitian mengenai akurasi suatu algoritma agar sistem keputusan yang dibuat memiliki tingkat akurasi yang tinggi. Penelitian ini menggunakan algoritma *data mining* yaitu algoritma *Classification and Regression Tree (CART)*. *CART* merupakan metode pohon keputusan biner. *CART* dikembangkan untuk melakukan analisis klasifikasi pada peubah respon baik yang nominal, ordinal, maupun kontinu. Metode klasifikasi *CART* terdiri dari dua metode yaitu metode pohon regresi dan pohon klasifikasi. Data mahasiswa yang mengikuti organisasi yang lulus tepat waktu dan tidak lulus tepat waktu akan diolah menggunakan algoritma *CART*. Setelah diklasifikasikan data tersebut akan dihitung hasil akurasinya menggunakan *K-fold Cross Validation* dengan nilai $K = 5$, $k = 10$, dan $K = 20$. Berdasarkan hasil contoh data mahasiswa yang mengikuti organisasi menunjukkan bahwa hasil perhitungan akurasi algoritma *CART* terbaik diperoleh ketika nilai $K = 20$. Algoritma *CART* telah mampu mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta. Algoritma *CART* menghasilkan rata-rata akurasi 80%.

Kata kunci : Kinerja, Data Mining, Organisasi, Akurasi, *CART*.

1. Pendahuluan

Perguruan tinggi merupakan kelanjutan pendidikan menengah yang diselenggarakan untuk mempersiapkan peserta didik untuk menjadi anggota masyarakat yang memiliki kemampuan akademis dan profesional yang dapat menerapkan, mengembangkan dan menciptakan ilmu pengetahuan, teknologi dan kesenian (UU No.2 tahun 1989, Pasal 16 ayat (1)).

Tujuan dari pendidikan tinggi adalah mempersiapkan generasi penerus bangsa agar dapat memajukan bangsa menjadi lebih baik. Peserta didik di perguruan tinggi disebut mahasiswa. Kata mahasiswa sendiri diambil dari kata "maha" yang berarti sesuatu yang lebih tinggi tingkatannya sedangkan siswa adalah pelajar atau orang yang menuntut ilmu. Jadi mahasiswa dapat diartikan sebagai pelajar atau siswa yang tingkatannya lebih tinggi dibandingkan dengan siswa biasa.

Selain kemampuan akademis, mahasiswa juga dituntut untuk memiliki kemampuan-kemampuan non akademis seperti kemampuan bekerjasama, kemampuan dalam memimpin, kemampuan berpikir kritis, rasa tanggung jawab dan lain-lain. Kemampuan non akademis ini tidak dapat dimiliki apabila mahasiswa hanya mengandalkan

materi yang diberikan oleh dosen. Oleh karena itu perguruan tinggi menyediakan organisasi kemahasiswaan.

Organisasi kemahasiswaan adalah fasilitas yang disediakan oleh perguruan tinggi sebagai wadah untuk mengembangkan kemampuan non akademis, minat dan bakat mahasiswa. Dengan mengikuti organisasi, mahasiswa diharapkan dapat meningkatkan kemampuan yang dimilikinya sehingga dapat menjalankan perannya dengan lebih baik.

Di balik semua manfaat positif yang bisa didapatkan oleh mahasiswa, kegiatan kuliah sambil mengikuti organisasi merupakan hal yang sulit dilakukan. Mahasiswa harus dapat membagi waktu dan konsentrasinya serta bertanggung jawab terhadap komitmen dari kedua aktivitas tersebut.

Apabila mahasiswa tidak dapat membagi waktu dan konsentrasinya dengan baik, maka resiko mengalami penurunan prestasi akademik dapat meningkat. Tak jarang pula mahasiswa yang aktif berorganisasi mengalami kegagalan dalam matakuliah dan harus mengulang matakuliah tersebut di semester berikutnya sehingga tidak dapat lulus tepat waktu.

Untuk mengetahui lama masa studi mahasiswa yang mengikuti organisasi, data mahasiswa dapat diklasifikasikan berdasarkan waktu kelulusan. Namun mengklasifikasikan data mahasiswa cukup sulit jika dilakukan secara manual. Universitas Negeri Jakarta belum memiliki sistem yang dapat mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi. Sebelum membangun sistem pengambilan keputusan, diperlukan penghitungan algoritmik untuk mengetahui kinerja suatu algoritma agar sistem keputusan yang dibuat memiliki tingkat akurasi yang tinggi.

Data mining merupakan serangkaian proses untuk mendapatkan informasi yang berguna dari data yang besar. Data mining juga dapat diartikan sebagai pernekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Dalam data mining terdapat banyak jenis teknik, diantaranya yaitu algoritma CART.

Classification and Regression Trees (CART) dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard S. Olshen dan Charles J pada tahun 1984. CART dikembangkan untuk melakukan analisis klasifikasi pada peubah respon baik yang nominal, ordinal, maupun kontinu. Metode klasifikasi CART terdiri dari dua metode yaitu metode pohon regresi dan pohon klasifikasi. Jika variable dependen yang dimiliki bertipe kategorik maka CART akan menghasilkan pohon klasifikasi (*classification tree*). Sedangkan jika variable dependen, yang dimiliki bertipe kontinu atau numerik maka CART akan menghasilkan pohon regresi (*regression tree*).

Data mahasiswa yang mengikuti organisasi kemahasiswaan memiliki banyak atribut yang bertipe kategorik baik nominal maupun ordinal. Sehingga untuk mengolah data tersebut dapat menggunakan algoritma CART. Untuk itu peneliti ingin menguji kinerja algoritma CART dalam mengklasifikasikan lama masa studi mahasiswa yang aktif berorganisasi di Universitas Negeri Jakarta.

1.2 Identifikasi Masalah

Berdasarkan latar belakang di atas, dapat diidentifikasi beberapa masalah sebagai berikut:

1. Universitas Negeri Jakarta belum memiliki aplikasi sistem pembuat keputusan yang dapat memprediksi lama masa studi mahasiswa yang mengikuti organisasi.
2. Sulitnya mengklasifikasikan data mahasiswa secara manual.
3. Diperlukan penghitungan algoritmik untuk mengetahui kinerja suatu algoritma agar sistem keputusan yang dibuat memiliki tingkat akurasi yang tinggi.

1.3 Pembatasan Masalah

Pembatasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian dilakukan untuk mengetahui kinerja algoritma CART dalam mengklasifikasikan data dan tidak membangun sistem.
2. Data yang diproses adalah data mahasiswa angkatan 2011 dan 2012 yang mengikuti organisasi kemahasiswaan pada tahun 2015.

1.4 Rumusan Masalah

Berdasarkan latar belakang, identifikasi dan pembatasan masalah di atas maka permasalahan pada penelitian ini adalah “Bagaimana kinerja algoritma CART untuk mengklasifikasikan lama masa studi mahasiswa yang aktif berorganisasi di Universitas Negeri Jakarta?”

1.5 Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengetahui kinerja algoritma CART dalam mengklasifikasikan lama masa studi mahasiswa yang aktif berorganisasi di Universitas Negeri Jakarta.

1.6 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Bagi mahasiswa, dapat menjadi acuan untuk penelitian selanjutnya dan mengetahui bagaimana penerapan algoritma CART dalam mengklasifikasikan data.
2. Bagi Dosen, dapat mengetahui kinerja algoritma *CART* dan sebagai bahan referensi untuk mengajar *data mining*.
3. Bagi Universitas, dapat mengetahui kinerja algoritma *CART* dalam mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi dan menjadi bahan pertimbangan dalam menentukan algoritma pengklasifikasian data untuk masalah lama masa studi mahasiswa.

2. Tinjauan Pustaka

2.1. Teori

2.1.1 Pengertian Kinerja

Menurut Bernadin dan Russel, kinerja adalah suatu catatan tentang hasil yang diperoleh dari fungsi-fungsi pekerjaan atau aktivitas selama kurun waktu tertentu (Ruky, 2002:15). Sedangkan menurut Payaman, kinerja adalah tingkat pencapaian atas pelaksanaan tugas tertentu (Simanjuntak, 2005:1).

Kinerja algoritma adalah nilai yang dihasilkan sebuah algoritma dalam melaksanakan pekerjaan tertentu. Pengukuran kinerja algoritma dilakukan dengan membandingkan antara hasil prediksi algoritma dengan nilai target variabel data *testing* sebagai data sebenarnya (Faisal, 2017: 69). Secara logika sederhana, kinerja algoritma dapat disimpulkan sebagai berikut:

$$\text{kinerja} = \frac{\text{jumlah instance yang diprediksi benar}}{\text{jumlah instance}} \quad (1)$$

2.1.2 Data Mining

Menurut Han dan Kamber (2012:8), *data mining* adalah proses pencarian pola yang menarik dan pengetahuan dari data yang besar. Sumber data

yang termasuk diantaranya *database*, *data warehouses*, *theWeb*, repositori informasi lainnya, atau data yang mengalir ke dalam sistem secara dinamis.

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan dalam *database*. *Data mining* merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan yang berguna dari *database* yang besar. Menurut Gartner Group, data mining adalah proses rekayasa pola matematis dari kumpulan data yang besar. Pola ini dapat berupa aturan, persamaan, hubungan, kecenderungan, atau model prediksi (Turban dkk., 2005: 263).

Data Mining dibagi menjadi beberapa kelompok menurut fungsi dan tugas masing-masing, yaitu:

1. Prediksi (*Prediction*)
Proses untuk menemukan pola dari data dengan menggunakan beberapa variabel untuk memprediksikan variabel lain yang tidak diketahui jenis atau nilainya.
2. Deskripsi (*Description*)
Proses untuk menemukan suatu karakteristik penting dari data dalam suatu *database*.
3. Klasifikasi (*Classification*)
Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan *class* atau konsep dari suatu data. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan.
4. Asosiasi (*Association*)
Proses ini digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data.

2.1.3 Klasifikasi

Klasifikasi adalah proses mencari sebuah model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep. (Han dan Kamber, 2012: 18). Klasifikasi data memiliki dua langkah proses yang terdiri dari *learning step* di mana model klasifikasi dibangun dan *classification step* di mana model digunakan untuk memprediksi label kelas dari data yang diberikan. (Han dan Kamber, 2012: 328)

Proses klasifikasi memiliki empat komponen yaitu (Gorunescu, 2011: 15) :

1. Kelas
Kelas merupakan variabel terikat yang merepresentasikan label yang terdapat pada objek.
2. *Predictor*
Predictor adalah variable bebas yang direpresentasikan oleh atribut data.
3. *Training Dataset*

Training Dataset adalah set data yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.

4. *Testing Dataset*

Testing Dataset adalah set data baru yang akan diklasifikasikan oleh model yang telah dibuat.

2.1.4 Decision Tree

Menurut Han, dkk (2012:330), *decision tree* adalah struktur pohon seperti diagram alir, di mana masing-masing simpul internal (simpul non daun) menunjukkan sebuah tes pada atribut, masing-masing cabang mewakili hasil pengujian, dan setiap *node* daun (atau *node* terminal) memegang label kelas. Simpul paling atas di pohon adalah simpul akar (*root*).

Pada *decision tree* terdapat 3 jenis node, yaitu:

- a. *Root Node*, merupakan *node* awal yang tidak memiliki input dan bisa tidak mempunyai output atau mempunyai *output* lebih dari satu.
- b. *Internal Node*, merupakan *node* percabangan. *Node* ini hanya memiliki satu input dan memiliki minimal dua *output*.
- c. *Leaf node* atau *terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu input dan tidak mempunyai output.

Beberapa model *decision tree* yang sudah dikembangkan antara lain IDS, C4.5, dan *CART*. *CART* adalah singkatan dari *classification and regression trees* (Santosa, 2007: 95).

2.1.5 Classification And Regression Trees (CART)

2.1.5.1. Pengertian Classification And Regression Trees

CART (Classification and Regression Trees) adalah salah satu metode atau algoritma dari salah satu teknik eksplorasi data decision tree. Metode ini dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard S. Olshen dan Charles J pada tahun 1984 dalam makalah mereka "*Classification and Regression Trees*". (Timofevv, 2005 :7)

Metode klasifikasi *CART* (Classification And Regression Trees) merupakan metode nonparametrik yang berguna untuk mendapatkan suatu kelompok data yang akurat sebagai pencari dari suatu pengklasifikasian. Metode klasifikasi *CART* terdiri dari dua metode yaitu metode pohon regresi dan pohon klasifikasi. Jika variabel dependen yang dimiliki bertipe kategorik maka *CART* menghasilkan pohon klasifikasi (*classification trees*). Sedangkan jika variabel dependen yang dimiliki bertipe kontinu atau numerik maka *CART* menghasilkan pohon regresi (*regression trees*).

Menurut Breiman, dkk ada lima ukuran *impurity* yang sering digunakan . tiga diantaranya digunakan untuk pohon klasifikasi dan dua untuk pohon regresi (De' ath dan Fabricius, 2000:3182). Untuk pohon klasifikasi, *impurity* didefinisikan dalam proporsi, *c*, dari tanggapan pada masing-

masing kategori. Tiga kriteria (indeks) umum adalah:

1. Indeks *Entropy*. Indeks ini identik dengan indeks keragaman Shannon-Weiner, dan membentuk kelompok dengan meminimalkan keragaman dalam kelompok.
2. Indeks Gini. Indeks ini memisahkan kategori besar menjadi grup terpisah, sedangkan indeks informasi cenderung membentuk kelompok yang terdiri lebih dari satu kategori pada awal pemecahan.
3. *Twoing Index*. Indeks ini dapat digunakan untuk lebih dari dua kategori. Indeks ini mendefinisikan dua “kategori super” di setiap perpecahan yang ditetapkan oleh indeks gini. Sedangkan untuk pohon regresi, terdapat dua kriteria (indeks) umum yaitu:
 1. Jumlah kuadrat grup rata-rata (mean). Ini setara dengan model kuadrat terkecil.
 2. Jumlah penyimpangan mutlak terhadap median. Ini akan memberikan pohon yang kuat. Namun, untuk data yang didominasi dengan nilai nol (0), kriteria ini bisa jadi tidak efektif, terutama bila variabel penjelas bersifat kategoris.

2.1.5.2. Indeks Gini

Indeks gini adalah kriteria berbasis ketidakmurnian yang mengukur perbedaan antara distribusi probabilitas dari nilai atribut target. (Rokach, 2015; 62) Indeks gini merupakan aturan pemecahan data yang paling banyak digunakan.

Indeks gini pada simpul t didefinisikan sebagai:

$$i(t) = \sum_{i,j} C(i|j)p(i|t)p(j|t) \quad (2)$$

2.1.5.3. Tahap-Tahap Pembentukan Algoritma CART

Algoritma *CART* melalui tiga tahapan, yaitu pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi dan penentuan pohon klasifikasi optimum (Sumartini dan Purnami, 2015:2338).

1. Pembentukan pohon klasifikasi

Pada tahap ini, hal pertama yang harus dilakukan adalah menentukan variabel dan *threshold* untuk dijadikan pemilah tiap simpul. Tahapan pembentukan pohon klasifikasi terdiri dari:

- a. Penentuan Pemilah

Pada tahap ini sampel data *learning* akan dipilah untuk mencari himpunan data yang lebih homogen. Data akan dipilah menggunakan indeks gini karena indeks gini akan selalu memisahkan kelas dengan anggota paling besar/ kelas terpenting dalam simpul terlebih dahulu. persamaan indeks gini adalah:

$$i(t) = \sum_{i,j-1} p(j|t)p(i|t), i \neq j \quad (3)$$

Dengan $p(j|t)$ adalah proporsi kelas j pada simpul t . Pemilahan yang terpilih akan membentuk suatu himpunan kelas yang disebut dengan simpul. Simpul tersebut akan melakukan pemilahan secara rekursif sampai diperoleh *terminal nodes*. Tahap

berikutnya adalah menentukan kriteria *goodness of split* untuk mengevaluasi pemilah dari pemilah s pada simpul t dengan persamaan:

$$\phi(s, t) = \Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (4)$$

t_L = cabang kiri dari nokhtah keputusan t

t_R = cabang kanan dari nokhtah keputusan t

Pemilah yang menghasilkan $\phi(s, t)$ lebih tinggi merupakan pemilah terbaik.

- b. Penentuan Simpul Terminal

Pengembangan pohon akan berhenti ketika jumlah pengamatan pada simpul kurang dari standar yang telah ditentukan (N_{min}) atau ketika sudah mencapai batasan tingkat kedalaman (*depth*) dalam pohon maksimal.

- c. Penandaan Label Kelas

Label kelas pada simpul terminal ditentukan berdasarkan aturan jumlah terbanyak, yaitu jika:

$$p(j_0|t) = \max_j \frac{N_j(t)}{N(t)} \quad (5)$$

Label kelas untuk simpul terminal t adalah j_0 yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul t yang paling kecil sebesar:

$$r(t) = 1 - \max_j p(j|t) \quad (6)$$

2. Pemangkasan pohon klasifikasi

Pemangkasan pohon dilakukan untuk mencegah *overfitting*. *Overfitting* adalah situasi di mana algoritma menghasilkan pengklasifikasian yang sangat sesuai dengan data *training* namun kehilangan kemampuan untuk menggeneralisasikan kejadian yang tidak direpresentasikan selama pelatihan (Rokach, 2007: 57). Pemangkasan pohon dapat dilakukan dengan pengukuran *cost complexity minimum* dengan rumus:

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \quad (7)$$

$R_\alpha(t)$ merupakan kombinasi linear biaya dan kompleksitas pohon yang dibentuk dengan menambahkan *cost penalty* bagi kompleksitas terhadap biaya kesalahan klasifikasi pohon. Selanjutnya, dilakukan pencarian pohon bagian $T(a) < T_{max}$ yang meminimumkan $R_\alpha(t)$ yaitu:

$$R_\alpha(T(a)) = \min_{T < T_{max}} R_\alpha(t) \quad (8)$$

3. Penentuan pohon klasifikasi optimum

Pohon klasifikasi yang terbentuk mungkin akan memiliki tingkat kompleksitas yang tinggi. Oleh karena itu perlu dioptimalkan terlebih dahulu sebelum digunakan untuk mengklasifikasikan data baru. Optimasi pohon akan memilih ukuran pohon yang tepat dan memotong *nodes* yang tidak signifikan. Salah satu algoritma pemangkasan yang dapat digunakan adalah *cross validation*.

2.1.5.4. Kelebihan dan Kekurangan CART

Algoritma *CART* memiliki kelebihan sebagai berikut (Timofeev, 2005: 19-24):

1. *CART* adalah metode non parametrik. Oleh karena itu *CART* tidak membutuhkan specification of any functional form.

2. *CART* tidak harus menentukan variabel di awal proses. Algoritma *CART* akan memilih sendiri variabel yang akan paling signifikan dan membuang variabel yang tidak signifikan.
3. *CART* dapat dengan mudah menangani *outliers*. *Outliers* adalah data yang memiliki karakteristik unik yang terlihat berbeda jauh dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim baik untuk sebuah variabel tunggal atau variabel kombinasi.
4. *CART* tidak memiliki asumsi dan memiliki komputasi yang cepat.
5. *CART* fleksibel dan memiliki kemampuan untuk menyesuaikan diri sesuai kebutuhan.

Sedangkan kekurangan Algoritma *CART* adalah:

1. *CART* mungkin akan menghasilkan pohon keputusan yang tidak stabil.
2. *CART* membagi data hanya berdasarkan satu variabel.

2.1.6 K-fold Cross Validation

K-fold Cross Validation adalah salah satu cara menemukan parameter terbaik dari satu model dengan cara menguji besarnya *error* pada data *test*. Dalam *cross validation* data dibagi menjadi *k* - subset dengan ukuran yang sama. Kemudian subset tersebut akan dibagi menjadi *data train* dan *data test* (Santosa, 2007: 220).

Tujuan dari *cross validation* adalah untuk mendefinisikan dataset untuk "menguji" model dalam tahap pelatihan (yaitu, *validasi data*), dalam rangka untuk membatasi masalah seperti terjadinya *overfitting*, memberikan wawasan tentang bagaimana model akan menggeneralisasi independen dataset dan lain-lain.

2.1.7 Organisasi

Pengertian organisasi adalah perkumpulan atau wadah bagi sekelompok orang untuk bekerjasama, terkendali dan dipimpin untuk tujuan tertentu (Hidayah, 2014). Organisasi biasanya memanfaatkan suatu sumber daya tertentu misalnya lingkungan, cara atau metode, material, mesin, uang, dan beberapa sumberdaya lain dalam rangka mencapai tujuan organisasi tersebut.

Ciri-ciri organisasi menurut Siwanto (2007: 73) yaitu :

1. Suatu organisasi adalah adanya sekelompok orang yang menggabungkan diri dengan suatu ikatan norma, peraturan, ketentuan dan kebijakan yang telah dirumuskan dan masing-masing pihak siap untuk menjalankannya dengan penuh tanggung jawab.
2. Dalam suatu organisasi yang terdiri atas sekelompok orang tersebut saling mengadakan hubungan timbal balik, saling memberi dan menerima dan juga saling bekerjasama untuk melahirkan dan merealisasikan maksud (*purpose*), sasaran (*objective*) dan tujuan (*goal*).
3. Dalam suatu organisasi yang terdiri atas sekelompok orang yang saling berinteraksi dan

bekerjasama tersebut diarahkan pada suatu titik tertentu., yaitu tujuan bersama dan ingin direalisasikan.

2.1.8 Organisasi Kemahasiswaan

Organisasi kemahasiswaan merupakan bentuk kegiatan di perguruan tinggi yang diselenggarakan dengan prinsip dari, oleh dan untuk mahasiswa (Sukirman, 2004:72, diacu dalam Pratiwi, 2017: 56). Organisasi tersebut merupakan wahana dan sarana pengembangan diri mahasiswa ke arah perluasan wawasan peningkatan ilmu dan pengetahuan, serta integritas kepribadian mahasiswa. Organisasi kemahasiswaan juga sebagai wadah pengembangan kegiatan ekstrakurikuler mahasiswa diperguruan tinggi yang meliputi pengembangan penalaran, keilmuan, minat, bakat dan kegemaran mahasiswa itu sendiri (Paryati Sudarman, 2004:34-35).

2.1.9 Lama Masa Studi

Lama masa studi adalah lamanya waktu yang dibutuhkan mahasiswa dalam menyelesaikan beban studi yang telah ditentukan.

2.2. Prosedur Penelitian

Prosedur yang dilakukan pada penelitian ini tersusun sebagai berikut: (1) Mengajukan surat permohonan izin mengadakan penelitian kepada Wakil Rektor III Universitas Negeri Jakarta, (2) Mengolah data mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta, (3) Membuat algoritma *CART* menggunakan *Python ver 2.7.2*. (4) Menguji kinerja algoritma *CART*, (5) Analisis hasil kinerja algoritma *CART*, dan (6) Penarikan kesimpulan.

3. Metodologi Penelitian

3.1. Tempat dan Waktu Penelitian

Penelitian dilakukan di program studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta. Waktu penelitian dilaksanakan pada bulan Februari 2017 – Juli 2017.

3.2. Teknik dan Prosedur Pengumpulan Data

Peneliti mengumpulkan data pada penelitian ini menggunakan pendekatan data primer. Prosedur yang dilakukan oleh penulis pada penelitian ini adalah penulis menyerahkan surat penelitian skripsi kepada bidang kemahasiswaan Universitas Negeri Jakarta. Kemudian data tersebut diberikan langsung kepada peneliti dalam bentuk *soft copy* yang berisi data mahasiswa mengikuti organisasi kemahasiswaan pada tahun 2015. Jumlah data yang digunakan dalam penelitian ini adalah 200 data yang terdiri dari 159 data mahasiswa yang lulus tepat waktu dan 41 data mahasiswa yang lulus tepat waktu. Parameter prediktor yang digunakan dalam penelitian ini adalah nama organisasi, jenis kelamin, fakultas, dan IPK. Sedangkan parameter yang akan dijadikan target adalah parameter kelulusan. Parameter kelulusan akan diberi label angka 1 (satu)

untuk mahasiswa yang lulus tepat waktu dan nilai 0 (nol) untuk mahasiswa yang tidak lulus tepat waktu.

3.3. Teknik Analisis Data

K-fold Cross Validation adalah salah satu cara menemukan parameter terbaik dari satu model dengan cara menguji besarnya *error* pada data *test*. Dalam *cross validation* data dibagi menjadi *k* - subset dengan ukuran yang sama. Kemudian subset tersebut akan dibagi menjadi *data train* dan *data test*. (Santosa, 2007: 220).

Pada penelitian ini akan dilakukan *K-fold Cross Validation* sebanyak 3 kali dengan nilai *k* = 5, 10, dan 20. Untuk nilai *k* = 5 data akan dibagi menjadi 5 subset, 4 subset digunakan untuk *training* dan 1 subset digunakan sebagai data *testing*. Sedangkan untuk nilai *k* = 10, data akan dibagi menjadi 10 subset dengan 9 *data training* dan 1 *data test*. Begitu pula dengan nilai *k* = 20, data akan dibagi menjadi 20 *subset* dengan 19 *data training* dan 1 *data test*.

4. Hasil Penelitian dan Pembahasan

4.1. Deskripsi Hasil Penelitian

Data yang digunakan dalam penelitian ini adalah data mahasiswa angkatan 2011 dan 2012 yang mengikuti organisasi kemahasiswaan di Universitas Negeri Jakarta pada tahun 2015. Data ini diperoleh dari Wakil Rektor III bagian Kemahasiswaan Universitas Negeri Jakarta.

4.1.1. Pre-processing Data

Sebelum data diproses menggunakan algoritma *CART*, data yang telah dikumpulkan harus melalui tahap *pre-processing data* terlebih dahulu. *Pre-processing data* dilakukan agar data sesuai dengan atribut yang diperlukan dalam penelitian sehingga mengurangi kesalahan pemrosesan data pada saat dilaksanakannya eksperimen.

4.1.1.1. Proses Penyortiran Data

Data awal yang didapatkan oleh penulis masih berupa data mentah dari bidang Kemahasiswaan Universitas Negeri Jakarta. Oleh karena itu data tersebut harus disortir terlebih agar sesuai dengan kebutuhan penelitian. Data yang diberikan bidang kemahasiswaan Universitas Negeri Jakarta adalah data mahasiswa angkatan 2011 – 2014. Sedangkan data yang dibutuhkan pada penelitian ini adalah data mahasiswa angkatan 2011 dan 2012. Dari hasil penyortiran tersebut diperoleh 200 data mahasiswa angkatan 2011 dan 2012 dengan 159 data mahasiswa yang lulus tidak tepat waktu dan 41 data mahasiswa yang lulus tepat waktu.

4.1.1.2. Proses Pemangkasan Data

Setelah melalui tahap penyortiran, tahap selanjutnya adalah proses pemangkasan data. Pada tahap penyortiran, data masih memiliki banyak unsur yang tidak diperlukan dalam penelitian, maka data perlu dipersempit dengan menghapus unsur yang tidak diperlukan tersebut.

4.1.1.3. Pemberian Label Pada Data

Setelah data dipersempit, maka selanjutnya data diberikan label untuk mempermudah proses pembelajaran data (*learning step*).

Data mahasiswa yang lulus tepat waktu diberi label angka 1 (satu), sedangkan untuk data mahasiswa yang tidak lulus tepat waktu diberi label angka 0 (nol). Setelah itu data disimpan dengan format .csv untuk memudahkan pengolahan data menggunakan *python*.

4.2. Analisis Data Penelitian

Setelah data selesai melalui tahap *pre-processing*, tahap selanjutnya adalah tahap pengolahan data. Dataset yang sudah diberi label tersebut akan diolah menggunakan algoritma *CART*. Alat yang peneliti gunakan pada penelitian ini adalah *Jetbrains Pycharm Community Edition 2017.1.4* dengan dasar *Python 2.7.2*.

4.3. Pembahasan

Berdasarkan hasil penelitian di atas, peneliti akan membahas secara detail mengenai hasil akurasi dari algoritma *CART*. Kinerja algoritma dapat dilihat dari nilai rata – rata akurasi algoritma.

Hasil Akurasi Algoritma CART		
K = 5	K = 10	K = 20
0.800	0.817	0.818

Tabel di atas adalah perbandingan hasil rata-rata akurasi algoritma *CART* menggunakan metode *K-fold Cross Validation*. Pada nilai *K* = 5 diperoleh nilai rata-rata akurasi sebesar 0.800. Sedangkan pada nilai *K* = 10, nilai rata-rata akurasi bertambah menjadi 0.817. Terakhir, ketika nilai *K* = 20, nilai rata-rata akurasi bertambah menjadi 0.818. Berdasarkan hasil tersebut, kinerja algoritma *CART* telah mencapai nilai akurasi 80%. Sedangkan nilai rata-rata akurasi tertinggi adalah ketika nilai *K* = 20 yaitu 0.818.

5. Kesimpulan dan Saran

5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat diambil kesimpulan sebagai berikut:

1. Kinerja algoritma *CART* menggunakan metode uji *K-fold Cross Validation* dengan nilai *K* = 5, 10, dan 20 diperoleh nilai akurasi tertinggi pada nilai *K* = 20 dengan nilai rata – rata akurasi 0.818.
2. Algoritma *CART* telah mampu mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta.

5.2. Saran

Peneliti memiliki beberapa masukan untuk penelitian selanjutnya yang berhubungan dengan klasifikasi data mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta, dengan harapan hasil pengklasifikasian lebih akurat.

- Menambahkan *data train*, semakin banyak *data train* maka akan semakin akurat hasil dari klasifikasi.

- Mencoba menggunakan algoritma yang lain agar lebih akurat.
- Menambah parameter klasifikasi.

6. Daftar Pustaka

- [FT] Fakultas Teknik. (2012). *Pedoman Akademik*. Jakarta: Fakultas Teknik, Universitas Negeri Jakarta.
- [FT] Fakultas Teknik. (2015). *Buku Panduan Penyusunan Skripsi dan Non Skripsi*. Fakultas Teknik, Universitas Negeri Jakarta.
- Anonim. (2010). Teori Kinerja. Tersedia: <https://teorionline.wordpress.com/2010/01/25/teori-kinerja/>. Diakses 9 Agustus 2017.
- De'ath, G. & Fabricius, K. E. (2000). Classification and regression trees: a powerful yet technique for ecological data analysis. *Ecology*, 81(11):3178-3192.
- Faisal, M. R. (2017). *Seri Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R*. [tempat tidak diketahui]: M Reza Faisal.
- Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques*. Berlin: Springer. Tersedia: <https://books.google.com/books?isbn=3642197213>. Diakses 15 Mei 2017.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. Ed ke-3. Amerika Serikat: Elsevier.
- Hidayah, dkk. (2014). Organisasi Manajemen. Tersedia: <http://myrealblo.blogspot.co.id/2015/11/dasar-dasar-manajemen-organisasi.html>. Diakses 20 Juli 2017.
- Nawawi, H. H. (2005). *Manajemen Sumber Daya Manusia untuk Bisnis yang Kompetitif*. Yogyakarta: Universitas Gajah Mada.
- Pratiwi, S. S. (2017). Pengaruh keaktifan mahasiswa dalam organisasi dan motivasi belajar terhadap prestasi belajar mahasiswa Fakultas Ekonomi Universitas Negeri Yogyakarta. *Jurnal Pendidikan dan Ekonomi*, 6(1): 54-64.
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. *Encyclopedia of Database Systems*. 532-538.
- Rokach, L. & Maimon, O. (2015). *Data Mining with Decision Trees – Theory and Applications*. Ed ke-2 Singapura: World Scientific Publishing Co. Pte. Ltd.
- Ruky, A. S. (2006). *Sistem Manajemen Kinerja*. Jakarta: Gramedia Pustaka Utama.
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Simanjuntak, P. J. (2005). *Manajemen dan Evaluasi Kerja*. Jakarta: Fakultas Ekonomi Universitas Indonesia.
- Siswanto. (2007). *Pengantar Manajemen*. Jakarta : PT. Bumi Aksara.
- Sumartini, S. H. & Purnami, S. H. (2015). Penggunaan metode classification and regression trees (CART) untuk klasifikasi rekurensi pasien kanker serviks di RSUD Dr. Soetomo Surabaya. *Jurnal Sains dan Seni ITS*, 4(2):2337-3520.
- Timofeev, R. (2005). Classification and Regression Trees (CART) Theory and Applications. [tesis]. Berlin: Humboldt University.
- Turban, E., Aronson, J. E., & Liang, T. P. (2005). *Decision Support Systems and Intelligent Systems*. Ed ke-7. New Jersey: Prentice-Hall, Inc.