

BAB I

PENDAHULUAN

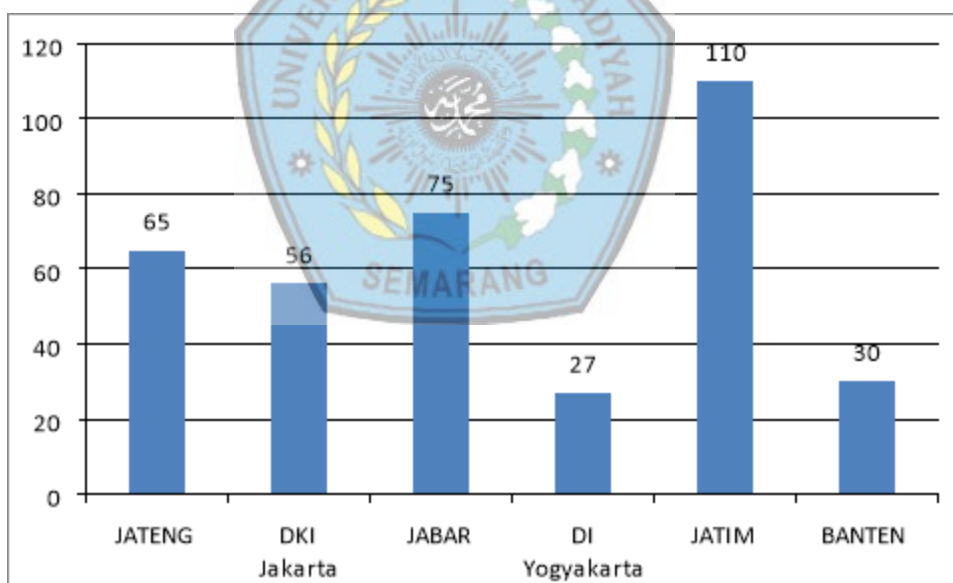
1.1 Latar Belakang

Pendidikan merupakan kebutuhan mutlak yang harus dipenuhi sepanjang hayat. Tanpa pendidikan, manusia mustahil dapat hidup dan berkembang sejalan dengan cita-cita dan tujuan hidup. Begitu juga pentingnya peranan pendidikan dalam tata kehidupan pribadi maupun masyarakat, maka dalam pengembangan watak bangsa haruslah berpegang dan bertumpu pada landasan pendidikan yang kokoh. Sebagaimana termaktub pada pembukaan Undang-Undang Dasar negara Republik Indonesia tahun 1945, yaitu “...memajukan kesejahteraan umum, mencerdaskan kehidupan bangsa, dan ikut serta dalam perdamaian dunia.” (Ansori, 2015)

Undang-Undang No. 20/2003 tentang Sistem Pendidikan Nasional menyatakan bahwa “Pembangunan nasional dalam bidang pendidikan merupakan upaya mencerdaskan kehidupan bangsa dan meningkatkan kualitas manusia Indonesia dalam mewujudkan masyarakat yang maju, adil dan makmur serta memungkinkan warga negaranya mengembangkan diri, baik berkenaan dengan aspek jasmaniah maupun rohaniah berdasarkan Pancasila dan UUD 1945”. Selanjutnya, dijelaskan dalam UU No. 20/2003, bahwa “Pendidikan nasional bertujuan mengembangkan potensi peserta didik agar manusia yang beriman dan bertaqwa terhadap Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, cakap,

kreatif, mandiri dan menjadi warga negara yang demokratis serta bertanggungjawab dalam rangka mencerdaskan kehidupan bangsa”.

Berdasarkan data buku statistik pendidikan tinggi Indonesia tercatat bahwa pulau Jawa mendominasi sebaran Perguruan Tinggi Negeri dan Perguruan Tinggi Swasta di lingkungan kemenristekdikti. Pada Pangkalan Data Pendidikan Tinggi laman <https://forlap.kemdikbud.go.id/> yang diakses pada tanggal 30 Juni 2021 menunjukkan jumlah perguruan tinggi aktif baik negeri maupun swasta di pulau Jawa ada 363 yang terbagi dalam 6 provinsi. Provinsi Jawa Tengah berjumlah 65. Provinsi DKI Jakarta berjumlah 56. Jawa Barat berjumlah 75. DI Yogyakarta berjumlah 27. Jawa Timur 110. Banten berjumlah 30.



Gambar 1.1 Jumlah Perguruan Tinggi di Jawa

Dengan banyaknya perguruan tinggi negeri maupun perguruan tinggi swasta khususnya di Jawa, persaingan untuk mendapatkan jumlah mahasiswa baru semakin ketat. Dalam melihat perguruan tinggi mana yang akan dipilih,

masyarakat sekarang sudah melihat status akreditasi perguruan tinggi. Masyarakat tentunya akan melihat perguruan tinggi yang nilai akreditasinya tinggi.

Saat ini dalam dunia pendidikan data yang berlimpah dan berkelanjutan bisa dimanfaatkan untuk data mining dalam rangka pengelolaan yang lebih baik dan pelaksanaan pembelajaran yang lebih efektif. Pada institusi pendidikan Perguruan Tinggi, data mahasiswa dan data jumlah kelulusan mahasiswa dapat menghasilkan informasi yang berlimpah berupa jumlah kelulusan setiap tahunnya, profil dan hasil akademik mahasiswa selama menempuh proses kegiatan belajar mengajar di perguruan tinggi. Adanya informasi mengenai lama studi mahasiswa tentu akan menjadi pendukung suatu pengambilan keputusan yang tepat bagi manajemen Perguruan Tinggi dalam mengambil langkah berikutnya.

Perguruan tinggi mempunyai peran strategis dalam pembangunan, khususnya pembangunan sumber daya manusia. Peran perguruan tinggi dalam pengembangan sumber daya manusia terutama sebagai penghasil agen-agen perubahan yang mampu merancang, mendorong dan memelopori perubahan. Perguruan tinggi adalah pencipta dan pendukung gagasan-gagasan baru untuk disumbangkan bagi kemajuan intelektual dan sosial masyarakat. Sejalan dengan hal itu, maka perguruan tinggi perlu melakukan perubahan, baik dalam arah serta tujuan perguruan tinggi yang menyangkut aspek kualitas, yang tercermin pada para alumninya. Lulus dengan baik dan tepat waktu merupakan salah satu parameter keberhasilan proses pembelajaran.

Kebutuhan masyarakat terhadap pelayanan kesehatan yang optimal, semakin terbukanya peluang kerja baik di dalam maupun di luar negeri dan

kebutuhan akan tenaga kesehatan yang profesional oleh penyedia layanan kesehatan, menjadi tantangan yang harus dijawab oleh penyedia tenaga kesehatan. Fakultas Kesehatan dan Keteknisian Medik (FKKM) merupakan salah satu fakultas di Universitas Widya Husada Semarang yang mencoba merealisasikan kebutuhan masyarakat tersebut. FKKM saat ini memiliki enam program studi diploma tiga yang terdiri dari Teknik Elektromedik, Teknik Rontgen, Kebidanan, Fisioterapi, Keperawatan dan Refraksi Optisi. Berbeda dengan jenjang pendidikan dasar dan menengah, pada tingkat perguruan tinggi terutama program diploma III memiliki syarat kelulusan bagi setiap mahasiswa yaitu telah menempuh minimal 108 SKS (Panduan Akademik, 2020). Lama studi mahasiswa ditetapkan dapat ditempuh dalam kurun waktu 3 tahun atau 6 semester dengan batas maksimal adalah 5 tahun atau 10 semester. Masa studi dihitung sejak pertama kali terdaftar sebagai mahasiswa.

Berbagai hal dalam dunia pendidikan mulai dari penerimaan mahasiswa baru hingga kelulusan dan evaluasi kinerja mahasiswa merupakan instrumen yang harus dilakukan agar terwujudnya pendidikan yang baik dan memenuhi standar pendidikan. Pendidikan tinggi dengan jumlah mahasiswa yang besar perlu memperhatikan kinerja mahasiswa baik dari penerimaan mahasiswa baru sebagai input maupun lulusan mahasiswa sebagai output. Pengambilan keputusan dalam evaluasi kinerja mahasiswa dapat dilakukan dengan penambangan data pendidikan (*educational data mining*) salah satunya dengan klasifikasi yang tepat untuk mempermudah mengevaluasi kinerja mahasiswa.

Pada perkembangan terbaru, teknik-teknik yang terdapat di dalam data mining mulai banyak digunakan. Khususnya teknik *decision tree* telah menjadi teknik yang populer karena *decision tree* yang dihasilkan mudah diinterpretasikan dan divisualisasikan (Chye, 2004). Metode klasifikasi telah banyak digunakan dalam berbagai bidang, seperti bidang pendidikan, pemerintahan, kesehatan, teknologi, maupun sosial. Klasifikasi sendiri didefinisikan sebagai pekerjaan mengelompokkan suatu objek ke dalam kategori tertentu. Klasifikasi dapat dilakukan pada data kategorik maupun bukan, jika data bukan kategorik maka harus diubah dalam bentuk kategorik terlebih dahulu.

Classification and Regression Tree (CART) dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen dan Charles J. Stone sekitar tahun 1980-an (Breiman et al., 1993). CART merupakan metodologi statistik nonparametrik yang dikembangkan untuk topik analisis klasifikasi, baik untuk variabel dependen kategorik maupun kontinu. CART menghasilkan suatu pohon klasifikasi jika variabel dependennya kategorik, dan menghasilkan pohon regresi jika variabel dependennya kontinu. (Komalasari, 2007)

CART dapat menyeleksi variabel-variabel dan interaksi-interaksi variabel yang paling penting dalam menentukan hasil atau variabel dependennya. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. Metode ini merupakan metode yang bisa diterapkan untuk himpunan data yang mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilahan biner.

Chi-square Automatic Interaction Detection (CHAID) merupakan salah satu analisis pohon keputusan (*decision tree*). Keunggulan dari metode *decision tree* yaitu membutuhkan waktu yang cepat untuk membentuk diagram pohon, representasi visual, dan mudah diinterpretasikan (Swain, 2016). Metode pohon keputusan (*decision tree*) mempunyai beberapa keunggulan dibandingkan metode lainnya untuk klasifikasi atau prediksi, seperti *neural network* dan analisis diskriminan (Cha, G.W *et al*, 2017). CHAID adalah suatu teknik iteratif yang menguji satu persatu variabel independen yang digunakan dalam klasifikasi dan menyusunnya berdasarkan pada tingkat signifikansi statistik *chi-square* terhadap variabel dependennya (Gallagher *et al*, 2000). Dengan kata lain, CHAID mengklasifikasikan variabel dependen kategori ke dalam kategori tertentu berdasarkan statistik *chi-square* variabel independen terhadap variabel dependen.

Metode CHAID merupakan salah satu tipe dari metode *Automatic Interaction Detection* (AID). Metode AID adalah suatu teknik untuk menganalisis kelompok data berukuran besar dengan membaginya menjadi sub-sub kelompok yang tidak saling tumpang tindih (Kass, 1982, diacu dalam Soemartojo 2002) yang diperuntukkan bagi data dengan peubah penjelas berskala ratio atau interval. Metode CHAID merupakan teknik eksplorasi nonparametrik untuk menganalisis sekumpulan data yang berukuran besar dan cukup efisien untuk menduga variabel independen yang paling signifikan terhadap peubah dependen. Interaksi antar peubah juga dapat dideteksi melalui metode ini (Du toit *et al*. 1977), sehingga diharapkan metode CHAID dapat menjadi alternatif analisis untuk menduga faktor-faktor yang menentukan suatu dependen kategorik.

Beberapa penelitian tentang metode CHAID dan CART telah dilakukan sebelumnya. Berdasarkan penelitian Siahaan, dkk (2016) dengan judul “Aplikasi Classification and Regression Tree (CART) dan Regresi Logistik Ordinal dalam Bidang Pendidikan”. Variabel yang digunakan dalam penelitian ini adalah jenis kelamin, asal daerah, program studi, status sekolah menengah, lama studi. Hasil perbandingan kedua metode menunjukkan bahwa regresi ordinal lebih baik yaitu mempunyai tingkat keakuratan klasifikasi 65% dibandingkan CART yang memiliki tingkat keakuratan klasifikasi 54,9%.

Kemudian penelitian oleh Darmawan, dkk (2017) dengan judul “Klasifikasi Lama Masa Studi Mahasiswa Menggunakan Perbandingan Metode Algoritma C.45 dan Algoritma Classification and Regression Tree”. Variabel yang digunakan dalam penelitian ini adalah IPK, jenis kelamin, asal daerah, program studi, asal sekolah dan lama studi. Hasil perbandingan kedua metode, kinerja algoritma CART lebih baik yaitu menghasilkan tingkat akurasi 60% dibandingkan algoritma C4.5 yang menghasilkan tingkat akurasi 40%.

Penelitian lain juga dilakukan Suniantara dan Muhammad Rusli (2017) dengan judul “Klasifikasi Waktu Kelulusan Mahasiswa STIKOM Bali menggunakan CHAID Regression Tree dan Regresi Logistik Biner”. Variabel yang digunakan dalam penelitian ini adalah jurusan, IPK, IPS semester 6, lama menyusun skripsi, jenis kelamin, nilai ujian masuk dan lama studi. Hasil perbandingan kedua metode menunjukkan bahwa kinerja algoritma CHAID lebih baik yaitu menghasilkan tingkat akurasi 91,2% dibandingkan regresi logistik yang menghasilkan tingkat akurasi 90,2%.

Selanjutnya penelitian oleh Wibowo, dkk (2019) dengan judul “Komparasi Algoritma *Naive Bayes* dan *Decision Tree* Untuk Memprediksi Lama Studi Mahasiswa”. Variabel yang digunakan dalam penelitian ini adalah gender, status mahasiswa, nilai, beasiswa dan lama studi. Dari komparasi dua metode menunjukkan bahwa *Decision Tree* lebih baik yaitu memiliki tingkat akurasi 55% dibandingkan *Naive Bayes* yang memiliki tingkat akurasi 30%.

Berbeda dengan penelitian sebelumnya, penelitian ini memperhatikan adanya aspek ketidakseimbangan data. *Class imbalance* atau ketidakseimbangan kelas merupakan salah satu permasalahan pada data mining. Hal ini terjadi pada saat kelas minoritas jauh lebih kecil atau lebih jarang dari kelas mayoritas (Syukron dan Subekti, 2018 dalam Bawono, Bonggo dan Rochdi Wasono. 2019). Data dikatakan tidak seimbang karena mahasiswa yang lulus tidak tepat waktu jumlahnya jauh lebih sedikit dibandingkan mahasiswa yang lulus tepat waktu. Kasus seperti ini dapat mengakibatkan salah klasifikasi pada kelas minoritas, yaitu kelas dengan jumlah amatan yang jauh lebih sedikit (Bunkhumpornpat et al, 2012). Sehingga mahasiswa yang seharusnya lulus tidak tepat waktu akan diprediksi lulus tepat waktu. Oleh karena itu penanganan pada data tidak seimbang perlu dilakukan. *Synthetic Minority Oversampling Technique* (SMOTE) merupakan salah satu metode penanganan yang dapat digunakan. SMOTE merupakan metode *oversampling* kelas minoritas dengan menciptakan data buatan (sintetik) untuk mengatasi masalah ketidakseimbangan kelas data (Chawla et al. (2002)).

Penelitian mengenai SMOTE sebelumnya telah dilakukan Sulistiyowati dan Mohamad Jajuli (2020) dengan judul Integrasi Naive Bayes dengan Teknik Sampling SMOTE untuk Menangani Data Tidak Seimbang. Dari total 878 data dengan atribut yang digunakan terbagi menjadi 813 data mayoritas (93%) dan 65 data minoritas (7%). Data tersebut menunjukkan adanya ketidakseimbangan data diantara kedua kelas. Tahapan modify melakukan proses SMOTE 500%. Berdasarkan hasil evaluasi akurasi menunjukkan bahwa akurasi Naive Bayes dengan SMOTE lebih tinggi dibandingkan Naive Bayes tanpa SMOTE dengan selisih 0,855%. Sedangkan berdasar hasil evaluasi G-Mean menunjukkan bahwa nilai G-Mean Naive Bayes dengan SMOTE lebih tinggi dibandingkan Naive Bayes tanpa SMOTE dengan selisih 0,028.

Penelitian lain juga dilakukan Franseda (2020) dengan judul Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas. Berdasarkan hasil penelitian menunjukkan bahwa model Decision Tree dan SMOTE Split Data dengan akurasi 71,12%, presisi 89,71 dan AUC 0,773. Penelitian ini masuk ke dalam kategori fair classification (cukup).

Berdasarkan uraian tersebut penelitian ini membahas mengenai klasifikasi data kelulusan mahasiswa berdasarkan lama studi menggunakan metode *Chi-Squared Automatic Interaction Detection* (CHAID) dan *Classification and Regression Tree* (CART). Pramana (2018) menyebutkan bahwa CART dan CHAID merupakan contoh klasifikasi, bagian dari algoritma data mining untuk melakukan ekstraksi pengetahuan dari suatu data. Kedua metode tersebut menarik untuk digunakan dikarenakan kondisi variabel terikat yang kategorik dengan dua

kategori atau biasa disebut biner. Sehingga topik tersebut dirasa perlu dikaji untuk menentukan metode mana yang lebih efisien dan sesuai dengan kebutuhan penelitian. Dari penelitian ini akan diketahui seberapa besar perbedaan output yang dihasilkan dari kedua metode.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana hasil tingkat akurasi metode CHAID dengan penerapan SMOTE dalam mengklasifikasi lama studi alumni Fakultas Kesehatan dan Keteknisian Medis?
2. Bagaimana hasil tingkat akurasi metode CART dengan penerapan SMOTE dalam mengklasifikasi lama studi alumni Fakultas Kesehatan dan Keteknisian Medis?
3. Bagaimana perbandingan tingkat akurasi antara metode CHAID dan CART dengan penerapan SMOTE dalam mengklasifikasi lama studi alumni Fakultas Kesehatan dan Keteknisian Medis?

1.3 Tujuan Penelitian

Tujuan penelitian yang ingin dicapai melalui penelitian ini adalah sebagai berikut.

1. Tingkat akurasi yang terbentuk dari metode CHAID dengan penerapan SMOTE dalam mengklasifikasi data lama studi alumni Fakultas Kesehatan dan Keteknisian Medis.

2. Tingkat akurasi yang terbentuk dari metode CART dengan penerapan SMOTE dalam mengklasifikasi data lama studi alumni Fakultas Kesehatan dan Keteknisian Medis.
3. Perbandingan tingkat akurasi antara metode CHAID dan CART dengan penerapan SMOTE dalam mengklasifikasi lama studi alumni Fakultas Kesehatan dan Keteknisian Medis.

1.4 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut.

1. Bagi Mahasiswa mampu mengaplikasikan ilmu statistika untuk metode *Chi-Squared Automatic Interaction Detection* (CHAID) dan *Classification and Regression Trees* (CART) dengan penerapan *Synthetic Minority Oversampling Technique* (SMOTE).
2. Bagi perpustakaan Jurusan Statistik Universitas Muhammadiyah Semarang, sebagai referensi yang dapat bermanfaat tentang metode *Chi-Squared Automatic Interaction Detection* (CHAID) dan *Classification and Regression Trees* (CART) dengan penerapan *Synthetic Minority Oversampling Technique* (SMOTE).
3. Bagi pembaca mampu menambah ilmu pengetahuan mengenai metode *Chi-Squared Automatic Interaction Detection* (CHAID) dan *Classification and Regression Trees* (CART) dengan penerapan *Synthetic Minority Oversampling Technique* (SMOTE).

1.5 Batasan Masalah

Agar pembahasan dalam penelitian ini tidak terlalu meluas, maka peneliti mencantumkan pembatasan masalah sebagai berikut.

1. Penelitian ini membandingkan tingkat akurasi dari metode *Chi-Squared Automatic Interaction Detection* (CHAID) dan *Classification and Regression Trees* (CART) dengan penerapan *Synthetic Minority Oversampling Technique* (SMOTE).
2. Studi kasus penelitian ini dibatasi pada data wisuda tahun 2018 sampai dengan 2020 Fakultas Kesehatan dan Keteknisian Medis.
3. Masalah yang akan dibahas pada skripsi ini dibatasi pada variabel-variabel yang digunakan yaitu lama studi, jenis kelamin, asal daerah, IPK, usia, program studi dan organisasi.
4. Data *training* dan *testing* yang digunakan dalam penelitian ini yaitu 70:30, 80:20 dan 90:10.