

BAB I PENDAHULUAN

1.1 Latar Belakang

Stres merupakan suatu kondisi disebabkan oleh transaksi antara individu dengan lingkungan sehingga menimbulkan persepsi jarak antara tuntutan yang berasal dari situasi sistem biologis, psikologis dan sosial (Anggun Saputri & Sugiharto, 2020). Stres dapat memengaruhi kesehatan fisik dan mental manusia, sebab stres merupakan bagian dari kehidupan manusia, artinya bahwa manusia tidak akan pernah luput dari pengalaman merasakan ketegangan dalam hidupnya. Stres dapat terjadi oleh semua kalangan, salah satunya ialah mahasiswa. Di kalangan mahasiswa, stres yang berkaitan dengan pembelajaran seringkali disebut sebagai stres akademik. Stres akademik adalah suatu keadaan dimana seseorang merasa cemas, khawatir, tertekan, dan tegang akibat ketidakberdayaan dalam mengatur tuntutan berupa deadline tugas dan ujian (Kurnia Putri et al., 2022).

Stressor adalah suatu peristiwa, situasi individu yang dapat menimbulkan stres dan reaksi terhadap stres (Ramadhan & Kunci, 2022). Sedangkan akademik merupakan kata sifat yang menunjukkan sesuatu yang berhubungan dengan ilmu pengetahuan, sehingga akademik stres (*stressor academic*) adalah suatu kondisi yang menimbulkan stres akibat tuntutan akademik yang melampaui kemampuan dari individu yang mengalaminya (Wilks, 2008). Stres pada mahasiswa dapat bersumber dari dua, yaitu faktor yang bersumber dari internal dan eksternal. Faktor internal terdiri dari tekanan, frustrasi, konflik, dan pemaksaan diri. Faktor eksternal terdiri dari keluarga, kampus dan lingkungan fisik (Ferderika Juniasi et al., 2023).

Statistik berperan penting dalam menganalisis faktor-faktor penyebab stres akademik dengan menggunakan berbagai metode. Oleh karena itu, diperlukan suatu alat bantu yang tepat untuk mengetahui ketepatan klasifikasi faktor-faktor yang mempengaruhi stress (Harahap et al., 2020). Tindakan yang efektif adalah dengan menggabungkan seluruh sampel ke dalam kelompok atau segmentasi yang homogen. Faktor-faktor yang mempengaruhi stres mahasiswa dalam menjalankan proses perkuliahan memerlukan metode klasifikasi dan menganalisis faktor-faktor tersebut agar efektif bila diterapkan pada data dengan pengamatan yang sangat banyak, salah satu metode yang dapat digunakan untuk melakukan klasifikasi adalah metode *Chi-squared Automatic Interaction Detection* (CHAID) dan Algoritma C5.0.

Metode CHAID merupakan salah satu analisis pohon keputusan (*decision tree*) dengan teknik *iterative* 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). Uji *Chi-square* merupakan uji nonparametrik yang tidak memerlukan uji prasyarat dan cocok untuk menguji hubungan antara variabel yang berbentuk kategori. Metode CHAID memiliki kelebihan yaitu dapat mengeksplorasi data yang berjumlah besar dengan peubah bertipe kategori dimana hasil analisis akan lebih mudah diinterpretasikan karena hasil *output* yang berupa pohon keputusan dan dapat mengetahui variabel bebas yang paling signifikan diantara variabel bebas lainnya. Pada setiap cabangnya, metode CHAID

melakukan tahap penggabungan (*merging*), tahap pemisahan (*splitting*) dan tahap pemberhentian (*stopping*). Algoritma CHAID memiliki keunggulan yaitu data akan didefinisikan ke dalam kelas yang sesuai, sehingga menjamin bahwa kelompok tersebut diidentifikasi berdasarkan pada populasi, interaksi dari semua faktor secara otomatis akan tepat pada sasaran karena pembagian dipertimbangkan berdasarkan konteks dari semua faktor, dan algoritma CHAID bersifat *iterative*, sehingga akan memberikan urutan variabel seperti yang diharapkan (Fanggidae, 2021).

Berkenaan dengan metode CHAID dalam mengklasifikasikan faktor-faktor yang mempengaruhi stres pada mahasiswa, metode klasifikasi pada Algoritma C5.0 juga mampu menghasilkan suatu pohon klasifikasi. Algoritma C5.0 banyak diterapkan dalam penelitian klasifikasi karena algoritma ini menunjukkan performa yang baik ketika digunakan dalam penelitian klasifikasi dan hasil proses klasifikasinya juga memiliki tingkat akurasi yang baik (Rahmayanti et al., 2020). Dengan menggunakan Algoritma C5.0 yang merupakan perluasan dari algoritma C4.5 dan ID 3 sebagai variasi metode untuk melihat performa dari suatu algoritma dimana pemilihan atribut yang akan diproses menggunakan perhitungan ukuran *Gain Ratio*. Ukuran *Gain Ratio* digunakan untuk memilih atribut uji pada tiap *node* di dalam pohon, sehingga *node* pada pohon terbentuk menggunakan perhitungan ukuran *Gain Ratio*. Atribut dengan nilai *Gain Ratio* tertinggi dipilih sebagai *parent node* bagi *node* selanjutnya. Algoritma C5.0 disempurnakan dengan kurasi lebih baik serta dapat menangani data yang bersifat diskrit maupun kontinyu.

Penelitian sebelumnya telah dilakukan oleh Murni Marcania pada tahun 2020 mengenai Prediksi Pengangkatan Karyawan Dengan Metode Klasifikasi Algoritma C5.0 (Studi Kasus PT. Kiyokunii Indonesia Factorys-2). Hasil yang diperoleh pada ketepatan klasifikasi dengan metode tersebut ialah 96,75% yang menunjukkan bahwa algoritma tersebut efektif dalam memprediksi menggunakan algoritma C5.0. Selanjutnya, penelitian yang dilakukan Tarigan dkk. (2024) mengenai klasifikasi kelulusan mahasiswa dengan metode CHAID dan *Exhaustive* CHAID. Hasil yang diperoleh pada ketepatan klasifikasi dengan metode CHAID sebesar 83,56%, Metode *Exhaustive* CHAID menunjukkan hasil yang serupa namun dengan keunggulan dalam memberikan informasi yang lebih rinci.

1.2 Tujuan dan Manfaat Penelitian

Adapun objek pada penelitian ini adalah mahasiswa Fakultas MIPA Universitas Hasanuddin, sehingga penelitian ini bertujuan untuk:

1. Mendapatkan ketepatan Algoritma C5.0 pada klasifikasi faktor-faktor yang mempengaruhi stres akademik pada mahasiswa.
2. Mendapatkan ketepatan metode CHAID pada klasifikasi faktor-faktor yang mempengaruhi stres akademik pada mahasiswa.
3. Mendapatkan metode yang lebih baik untuk klasifikasi faktor-faktor yang mempengaruhi stres akademik pada mahasiswa.

Penelitian ini memiliki manfaat bagi penulis juga bagi pembaca dari kalangan akademik. Penulis akan semakin memahami metode statistik khususnya Algoritma C5.0 dan CHAID yang digunakan dalam penelitian ini. Selain itu, dapat memperkaya pengetahuan penulis mengenai faktor-faktor yang menjadi penyebab stres akademik pada mahasiswa

khususnya di Fakultas MIPA Universitas Hasanuddin. Dengan penelitian ini, pembaca dapat mengetahui serta memahami metode klasifikasi statistik pada Algoritma C5.0 dan metode CHAID.

1.3 Batasan Masalah

Pada penelitian ini, metode yang digunakan adalah CHAID dan Algoritma C5.0. Data yang digunakan diperoleh dari hasil kuisioner yang diberikan kepada mahasiswa di Fakultas MIPA Universitas Hasanuddin. Ketepatan klasifikasi pada kedua metode akan dihitung menggunakan evaluasi metrik akurasi.

1.4 Landasan Teori

1.4.1 Uji Instrumen

1.4.1.1 Uji Validitas

Uji validitas digunakan untuk mengetahui valid atau tidaknya suatu kuisioner. Kuisioner dikatakan valid jika pada pertanyaan tersebut mampu untuk memberikan sesuatu yang akan diukur oleh kuisioner tersebut. Pengujian validitas penting dilakukan agar pertanyaan yang diberikan tidak menghasilkan data yang menyimpang dari gambaran variabel yang dimaksud. Pada Persamaan (1) uji validitas dapat diukur dengan menggunakan rumus *Product Moment* dengan taraf signifikan 5% atau 0,05 (Terapan et al., 2019).

$$r_{XY} = \frac{n \sum_{j=1}^n x_{ij}y_{ij} - (\sum_{j=1}^n x_{ij})(\sum_{j=1}^n y_{ij})}{\sqrt{n \sum_{j=1}^n x_{ij}^2 - (\sum_{j=1}^n x_{ij})^2} \sqrt{\sum_{j=1}^n y_{ij}^2 - (\sum_{j=1}^n y_{ij})^2}} \quad (1)$$

Apabila $r_{hitung} > r_{tabel}$, maka instrument atau item pertanyaan berkorelasi signifikan terhadap skor total (valid).

1.4.1.2 Uji Reliabilitas

Uji reliabilitas digunakan untuk mengukur suatu kuisioner yang merupakan indikator dari perubahan atau kontruk. Kuisioner dapat dikatakan reliabel atau handal jika jawaban seseorang terhadap pernyataan adalah konsisten atau stabil dari waktu ke waktu (Ayu et al., 2020). Perhitungan reliabilitas hanya bisa dilakukan jika variabel pada kuisioner tersebut sudah valid. Pada Persamaan (2) uji reliabilitas dapat diukur dengan menggunakan formula *Cronbach's alpha* (α).

$$\alpha = \left(\frac{n}{n-1} \right) \left(1 - \frac{\sum \sigma_i^2}{\sigma_{total}^2} \right) \quad (2)$$

Apabila suatu variabel menunjukkan nilai *Cronbach's Alpha* $> 0,60$ maka dapat disimpulkan bahwa variabel tersebut dapat dikatakan reliabel atau konsisten dalam mengukur (Puspasari et al., 2022).

1.4.2 Pohon Keputusan (*Decision Tree*)

Decision Tree merupakan salah satu teknik yang dapat digunakan untuk melakukan klasifikasi terhadap sekumpulan objek. Teknik ini terdiri dari kumpulan pohon keputusan, dihubungkan oleh cabang, bergerak ke bawah dari *node* sampai berakhir di *leaf node*. *Decision tree* merupakan pohon keputusan pada teknik klasifikasi untuk mempresentasikan aturan (Istiqomah et al., 2019). *Decision tree* dibentuk melalui pengklasifikasian variabel prediktor dan variabel respon. Kedua variabel tersebut dikenal dengan nama atribut.

Salah satu keuntungan utama dari pohon keputusan adalah kemampuannya untuk memberikan pemahaman yang visual tentang bagaimana suatu keputusan dibuat. Hal ini memungkinkan pengguna untuk menganalisis alur logika dibalik suatu keputusan dan memahami faktor-faktor yang paling berpengaruh dalam mengklasifikasikan atau memprediksi suatu objek (Ramadhon et al., 2024).

1.4.3 CHAID (*Chi-Square Automatic Interaction Detector*)

CHAID (Chi-Square Automatic Interaction Detector) merupakan suatu teknik *iterative* yang menguji satu persatu variabel independen yang digunakan dalam klasifikasi dan menyusunnya berdasarkan pada tingkat signifikansi statistik *Chi-Squared* terhadap variabel dependennya (Sulviana & Hamim Wigena, 2018). *Chi-Squared Automatic Interaction Detection (CHAID)* pertama kali dikenalkan oleh Dr. G. V. Kass dalam artikel dengan judul “*An Exploratory Technique for Investigating Large Quantities of Categorical Data*” pada tahun 1980.

CHAID digunakan untuk membentuk segmentasi yang membagi sebuah sampel menjadi dua atau lebih kelompok yang berbeda berdasarkan kriteria tertentu. Hal ini kemudian diteruskan dengan membagi kelompok-kelompok tersebut menjadi kelompok yang lebih kecil berdasarkan variabel-variabel independen yang lain. Prosesnya berlanjut sampai tidak ditemukan lagi variabel-variabel independen yang signifikan secara statistik (Nurul Hasana, 2021).

1.4.3.1 Uji *Chi-Square*

Teknik *Chi-Square* memungkinkan untuk mengetahui independensi antara dua variabel pada tiap level (Sondang et al., 2006).

Hipotesis:

H_0 : tidak terdapat hubungan antara variabel independen dengan variabel dependen (kedua kriteria klasifikasi bebas).

H_1 : terdapat hubungan antara variabel independen dengan variabel dependen (kedua kriteria klasifikasi tidak bebas).

Statistik uji yang ditunjukkan pada Persamaan (10):

$$\chi_{hitung}^2 = \sum_{k=1}^r \sum_{l=1}^c \left[\frac{(n_{kl} - E_{kl})^2}{E_{kl}} \right] \quad (10)$$

$$E_{kl} = \frac{n_k \cdot n_l}{n}, k = 1, 2, \dots, r \text{ dan } lj = 1, 2, \dots, c \quad (11)$$

Daerah kritis:

H_0 ditolak jika $\chi_{hitung}^2 > \chi_{\alpha:(r-1)(c-1)}^2$ atau H_0 ditolak jika nilai $p - value < \alpha$.

1.4.3.2 Algoritma CHAID

Algoritma CHAID merupakan diagram pohon CHAID yang akan digunakan untuk memprediksi nilai variabel dependen. Secara garis besar algoritma ini dapat dibagi menjadi 3 tahap, yaitu tahap penggabungan (*merging*), tahap pemisahan (*splitting*), dan tahap penghentian (*stopping*) (Hasibuan & Harahap, 2019).

1. Tahap Penggabungan (*Merging*)

Tahap penggabungan dilakukan uji signifikansi pasangan kategori pada setiap variabel independen terhadap variabel dependen. Kriteria pengujian menggunakan uji *Chi-Square*. Pasangan kategori variabel independen yang signifikan terhadap variabel dependen tidak digabung, sedangkan yang tidak signifikan akan digabung menjadi sebuah kategori gabungan.

H_0 : tidak terdapat hubungan antara variabel pertama dan variabel kedua.

H_1 : terdapat hubungan antara variabel pertama dan variabel kedua.

Tolak H_0 jika $X_{hitung}^2 > X_{\alpha,(r-1)(c-1)}^2$.

2. Tahap Pemisahan (*Splitting*)

Tahap pemisahan akan dipilih variabel independen mana yang akan digunakan sebagai *split node* (pemisah *node*) yang terbaik. Pemilihan dikerjakan dengan membandingkan $p - value$ pada setiap variabel independen. Adapun langkah-langkahnya sebagai berikut:

- a. Pilih variabel independen yang memiliki $p - value$ terkecil (paling signifikan).
- b. Jika $p - value$ kurang dari sama dengan tingkat spesifikasi α , *split node* menggunakan variabel independen ini. Jika tidak ada variabel independen dengan nilai $p - value$ yang signifikan, tidak dilakukan *split* dan *node* ditentukan sebagai *terminal node*.

3. Tahap Penghentian (*Stopping*)

Tahap penghentian dilakukan jika proses pertumbuhan pohon harus dihentikan sesuai dengan peraturan pemberhentian dibawah ini:

- a. Jika sebuah simpul menjadi murni, yaitu semua kasus dalam simpul mempunyai nilai variabel dependen yang sama.
- b. Jika semua kasus dalam sebuah simpul mempunyai nilai yang sama untuk setiap variabel independen.
- c. Jika kedalaman pohon mencapai batas kedalaman pohon maksimal yang ditentukan.
- d. Jika ukuran simpul kurang dari ukuran simpul minimal yang ditentukan.
- e. Jika pemisah dari sebuah simpul menghasilkan *child node* (*node* anak) yang ukurannya kurang dari ukuran simpul anak minimal, simpul anak yang memiliki terlalu sedikit kasus ($<$ ukuran simpul anak minimal) akan digabungkan dengan simpul anak yang paling mirip yang diukur

menggunakan $p - value$ terbesar. Namun, jika ukuran simpul anak yang dihasilkan adalah satu, pertumbuhan diagram pohon akan dihentikan.

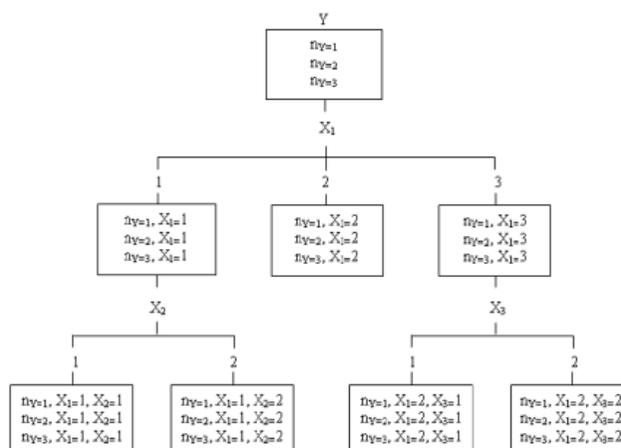
1.4.4 Variabel-Variabel CHAID

Variabel yang digunakan dalam metode CHAID untuk variabel dependen dan independen dengan kategori ordinal atau nominal. Pada kedua metode tersebut terdapat tiga jenis variabel independen sebagai berikut:

1. Variabel Monotonik
Variabel monotonik adalah variabel independen dimana kategori-kategorinya di dalamnya dapat digabungkan jika berurutan (data ordinal).
2. Variabel Bebas
Variabel bebas adalah variabel independen dimana kategori-kategori didalamnya dapat digabungkan meskipun tidak berurutan (data nominal).
3. Variabel Mengambang
Variabel mengambang adalah variabel independen yang dapat diperlakukan sebagai variabel monotonik, kecuali untuk kategori yang *missing value*, yang dapat dikombinasikan dengan kategori manapun.

1.4.5 Pohon Klasifikasi CHAID

Pada Gambar 1 CHAID akan menghasilkan sebuah diagram pohon klasifikasi yang menggambarkan pembentukan segmen. Diagram terdiri dari batang pohon (*tree trunk*) dengan membagi menjadi lebih kecil berupa cabang-cabang.



Gambar 1. Diagram Pohon Klasifikasi dan CHAID

Diagram pohon disusun mulai dari kelompok induk, berlanjut dibawahnya sub kelompok yang berturut-turut dari hasil pembagian kelompok induk berdasarkan kriteria tertentu. *Node* pada ujung pohon yang tidak terdapat percabangan lagi disebut *terminal node*. Pada pohon klasifikasi terdapat istilah kedalaman (*depth*) yang berarti banyaknya tingkatan *node-node* sub kelompok sampai ke bawah pada *node* sub kelompok yang terakhir (Eherler & Lehmann, 2001).

1. *Root node* (simpul akar) adalah sampel yang mengandung data seluruh sampel.
2. *Child node* (simpul anak) adalah simpul yang dihasilkan dari pembelahan suatu simpul yang didasarkan pada suatu variabel prediktor paling signifikan dengan kategori-kategori paling signifikan (kombinasi kategorik yang paling signifikan).
3. *Parent node* (simpul induk) adalah simpul yang dibelah berdasarkan suatu variabel prediktor, sehingga menghasilkan beberapa simpul anak.
4. *Terminal node* (simpul terminal/simpul akhir) adalah simpul yang tidak dapat dibelah lagi karena tidak ada variabel prediktor yang signifikan dalam membelah simpul tersebut. Simpul terminal mengakhiri pertumbuhan cabang pohon CHAID.

Pembelahan dimulai dari simpul akar menjadi beberapa simpul anak berdasarkan variabel independen yang paling signifikan dengan kategori-kategori yang signifikan. Masing-masing simpul anak yang dihasilkan dari pembelahan diperiksa secara terpisah untuk mengetahui apakah suatu simpul anak dapat dibelah atau tidak. Pembelahan simpul anak menggunakan salah satu variabel independen paling signifikan dengan kategori-kategori signifikan yang dipilih dari variabel sisa. Proses tersebut berlanjut dengan pembelahan berurutan untuk simpul anak yang diperoleh dari tahap sebelumnya sampai dengan tidak ada lagi simpul anak yang dapat dibelah. Pembelahan akan berhenti apabila memenuhi kriteria *stopping rule* yang ditentukan. Jika variabel dependen merupakan variabel kategorik maka variabel independen yang membelah suatu simpul akar atau simpul induk menjadi simpul anak dipilih berdasarkan uji *Chi-Square*.

1.4.6 Algoritma C5.0

Algoritma C5.0 merupakan suatu pengembangan dari algoritma C4.5, yang awalnya dikembangkan oleh Ross Quinlan pada tahun 1987 (Aidia et al., 2024). Pada algoritma C5.0, pemilihan atribut dilakukan dengan menggunakan *Gain Ratio*. Algoritma C5.0 memperlakukan variabel kontinu sama dengan yang dilakukan oleh CART, tetapi untuk variabel kategorik algoritma C5.0 memperlakukan nilai variabel kategorikal sebagai *splitter*. Subset sampel yang muncul dari percabangan akan terus dibagi lagi. Proses ini akan berlanjut hingga sampel subset tidak dapat dibagi lebih lanjut. Pada akhirnya, subset sampel yang tidak memberikan kontribusi signifikan terhadap model akan dihapus (Muningsih et al., 2024).

Langkah untuk pembentukan pohon pada algoritma C5.0 serupa dengan langkah pembentukan pohon pada algoritma C4.5. Hal yang menyerupai terdapat pada perhitungan *Entropy* dan *Information Gain*. Namun, perbedaannya terletak pada tahapan setelah perhitungan *Information Gain*, dimana algoritma ini melanjutkan perhitungannya dengan menghitung *Gain Ratio* dengan menggunakan nilai *entropy* dan *information gain* yang telah dihitung sebelumnya, sementara algoritma C4.5 berhenti pada perhitungan *Gain*. Dalam proses pembentukan pohon keputusan nilai *Information Gain* tertinggi akan terpilih sebagai *root* bagi *node* selanjutnya. Algoritma C5.0 dimulai dengan semua data yang dijadikan akar dari pohon keputusan sedangkan atribut yang dipilih akan menjadi pembagi bagi sampel berikutnya (Kantono et al., 2019). Persamaan (12) digunakan untuk mencari nilai *entropy*.

$$Entropy(S) = - \sum_{j=1}^n P(j|t_L) \log_2 P(j|t_L) \quad (12)$$

Berikutnya, Persamaan (13) digunakan untuk menghitung nilai *Information Gain* (Bijaksana et al., 2020):

$$IG(S, A) = Entropy(S) - \sum_i^m \frac{|S_i|}{|S|} \times Entropy(S_{ij}) \quad (13)$$

Dengan:

- S = nilai kasus
- A = variabel ke- l pada kategori ke- j
- m = jumlah kategori pada variabel A
- $|S_i|$ = jumlah kasus pada kategori ke- i
- $|S|$ = jumlah kasus dalam S

Setelah mendapatkan nilai *Entropy* dan *Information Gain*, tahap berikutnya ialah menghitung nilai *Gain Ratio*. Persamaan (14) digunakan untuk melakukan perhitungan *Gain Ratio*.

$$Gain\ Ratio = \frac{IG(S, A)}{\sum_{i=1}^m Entropy(S_i)} \quad (14)$$

Dengan:

- $\sum_{i=1}^m Entropy(S_i)$ = jumlah nilai *Entropy* dalam suatu variabel ke- i kategori ke- j

1.4.7 Ketepatan Klasifikasi

Salah satu alat bantu untuk menilai seberapa baik sebuah *classifier* adalah *confusion matrix* (Sulistiani & Utami, 2018). *Confusion matrix* dapat memberikan rincian pada klasifikasi, kelas yang ditampilkan di bagian atas matrix yaitu kelas yang diprediksi sedangkan kelas yang diobservasi ditampilkan di bagian kiri. *Confusion matrix* memiliki beberapa *metric* evaluasi classifier yang dapat dihitung seperti akurasi, *specifity*, dan *sensitivity*. Tabel 1 menunjukkan bentuk format umum dari *confusion matrix*.

Tabel 1. Confusion Matrix

Actual Class	Predicted Class	
	C1	-C1
C1	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
-C1	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Keterangan:

- True Positive (TP)* : Jumlah baris kelas C1 yang benar diklasifikasikan sebagai kelas C1
- False Negative (FN)* : Jumlah baris kelas C1 tetapi diklasifikasikan sebagai bukan kelas C1
- False Positive (FP)* : Jumlah baris bukan kelas C1 tetapi diklasifikasikan sebagai kelas C1

True Negative (TN) : Jumlah baris bukan kelas C1 yang benar tetapi diklasifikasikan sebagai bukan kelas C1

Sehingga perhitungan akurasi, *specificity* dan *sensitivity* secara berturut-turut ialah:

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \quad (15)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (16)$$

$$Specificity = \frac{TN}{FP + TN} \quad (17)$$

1.4.8 Stres Akademik

Dalam menjalani proses pendidikan terutama pada tingkat perguruan tinggi sering kali menghadapi masalah terkait kesehatan mental seperti stres. Stres yang dialami oleh mahasiswa adalah stres akademik (Kumar et al., 2013). Stres akademik disebabkan oleh adanya *academic stressor* yang bermula dari proses pembelajaran seperti tekanan untuk memperoleh nilai yang bagus, lamanya waktu yang dihabiskan untuk belajar, banyaknya tugas serta kecemasan dalam menghadapi ujian (Barseli et al., 2017). Stres pada mahasiswa dapat bersumber dari dua, yaitu faktor yang bersumber dari internal dan eskternal. Faktor internal terdiri dari tekanan, frustrasi, konflik, dan pemaksaan diri. Tekanan yang dimaksud merupakan hal-hal yang mencakup berat dan banyaknya tugas, *deadline* tugas yang mendesak, adanya persaingan antar teman, harapan atau keinginan dari dosen dan orang tua. Frustrasi yang dialami oleh mahasiswa seperti kekecewaan akibat belum mencapai target yang telah direncanakan, kesulitan untuk mencapai tujuan yang telah dirancang, kesulitan atau keterbatasan sumber bacaan dan kecewa terhadap keadaan. Oleh sebab itu kegagalan dan rasa frustrasi mahasiswa dalam menjalankan masa perkuliahan dapat menyebabkan stres akademik. Konflik diartikan sebagai perjuangan untuk memperoleh nilai, status dan kekuasaan dengan tujuan memperoleh keuntungan dan mengalahkan pesaing. Konflik muncul ketika individu berada dibawah tekanan untuk memilih dua atau lebih hal yang berlawanan. Pemaksaan diri merupakan hal-hal yang berkaitan dengan kecemasan terhadap segala hal, ingin bersaing dengan teman, khawatir dalam mencapai tujuan.

Faktor eksternal terdiri dari keluarga, kampus dan lingkungan fisik. Keluarga, kondisi di dalam keluarga otomatis memberikan stres pada mahasiswa. Kondisi yang dimaksud seperti konflik yang muncul, seperti orangtua yang selalu sibuk diluar rumah. Kampus, stres yang berkaitan dengan kampus mencakup tekanan akademik (tugas, mata kuliah, manajemen waktu, cara dosen mengajar) dan tekanan sebaya (persaingan, diterima atau ditolak kelompok sebaya). Lingkungan fisik, berkaitan dengan kondisi sekitar yang membuat mahasiswa tidak nyaman dan stres, misalnya cuaca panas, keramaian atau lingkungan yang padat sehingga mahasiswa tidak bisa berkonsentrasi belajar (Rohmah & Mahrus, 2024).

BAB II METODOLOGI PENELITIAN

2.1 Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data primer yang diperoleh dari hasil survei berbentuk kuisioner yang akan disebar kepada mahasiswa Fakultas MIPA Universitas Hasanuddin. Data tersebut merupakan data mahasiswa aktif dengan periode yang digunakan dalam penelitian diambil dari tahun 2021-2024. Data yang digunakan dalam penelitian ini berjumlah 96 data mahasiswa aktif yang dijadikan sebagai sampel. Penyebaran kuisioner dilakukan pada tanggal 21 November 2024 menggunakan media *Google Form* yang disebar secara online untuk mempermudah distribusi kuisioner serta memaksimalkan jangkauan responden dalam waktu yang efisien.

2.2 Variabel Penelitian

Berdasarkan Tabel 2 variabel-variabel yang digunakan dalam penelitian ini terdiri atas variabel respon dan variabel prediktor. Variabel respon yang digunakan adalah Y berupa stres. Sementara itu, variabel prediktor pada meliputi tekanan (X1), konflik (X2), frustrasi (X3), pemaksaan diri (X4), keluarga (X5), kampus (X6) dan lingkungan fisik (X7).

Tabel 2. Variabel Faktor-Faktor yang Mempengaruhi Stres Akademik

Variabel	Keterangan	Skala	Kategori
Y	Stres	Nominal	0: Stres 1: Tidak Stres
X1	Tekanan	Nominal	0: Tertekan 1: Tidak Tertekan
X2	Frustrasi	Nominal	0: Frustrasi 1: Tidak Frustrasi
X3	Konflik	Nominal	0: Berkonflik 1: Tidak Berkonflik
X4	Pemaksaan Diri	Nominal	0: Memaksakan Diri 1: Tidak Memaksakan Diri
X5	Keluarga	Nominal	0: Tidak Harmonis 1: Harmonis
X6	Kampus	Nominal	0: Terganggu 1: Tidak Terganggu
X7	Lingkungan Fisik	Nominal	0: Tidak Nyaman 1: Nyaman

2.3 Populasi dan Sampel

2.3.1 Populasi

Populasi merupakan sekumpulan objek atau subjek yang memiliki kualitas dan karakteristik tertentu yang ditentukan oleh peneliti untuk dipelajari, lalu diambil kesimpulannya (Syukri et al., 2019). Melalui pendapat tersebut menjadi salah satu acuan bagi penulis untuk menentukan populasi. Populasi yang akan digunakan adalah

mahasiswa Fakultas MIPA Universitas Hasanuddin tahun 2021-2024 terdapat 2.427 mahasiswa aktif.

2.3.2 Sampel

Sampel merupakan bagian dari jumlah dan karakteristik yang dimiliki oleh populasi, sehingga dalam pengambilan sampel harus menggunakan cara tertentu yang didasarkan oleh pertimbangan-pertimbangan yang ada (Diana et al., 2021). Pada penelitian ini, sampel akan diambil dengan jumlah kurang lebih responden, dengan pertimbangan bahwa jumlah tersebut diharapkan dapat mewakili sebagai sampel penelitian. Persamaan (17) merupakan rumus slovin yang digunakan.

$$n = \frac{N}{Nd^2 + 1} \quad (17)$$

Keterangan:

- n = jumlah sampel
 N = jumlah populasi
 d = presesi (10%)

Dari rumus diatas didapat hasil perhitungan sebagai berikut:

$$n = \frac{2.437}{2.437(0,1)^2 + 1} = \frac{2.437}{25,37} = 96$$

Berdasarkan rumus diatas besar nilai sampel adalah 96 mahasiswa.

2.4 Teknik Pengambilan Sampel

Teknik pengambilan sampel digunakan untuk menentukan sampel dalam penelitian, terdapat beberapa teknik sampling yang digunakan. Teknik sampling dibagi menjadi dua kelompok yaitu *probability sampling* dan *non probability sampling* (Firmansyah et al., 2022). Pada penelitian ini menggunakan *probability sampling* karena teknik pengambilan sampel ini memberikan peluang atau kesempatan yang sama bagi setiap anggota populasi untuk dipilih menjadi sampel. Dengan menggunakan sampel acak sederhana (*simple random sampling*) sebagai cara pengambilan sampel dari populasi yang dilakukan secara acak tanpa memperhatikan strata yang ada dalam populasi itu.

2.5 Metode Analisis

Metode analisis yang digunakan dalam penelitian ini adalah metode CHAID dan Algoritma C5.0.

1. Membuat kuisioner, kemudian disebarakan kepada mahasiswa aktif Fakultas MIPA Universitas Hasanuddin.
2. Melakukan Uji Validitas dan Uji Reliabilitas.
3. Tahapan analisis data yang dilakukan dengan metode CHAID yaitu:
 - a. Melakukan tahap Penggabungan (*Merging*). Variabel independen digabungkan berdasarkan signifikansi uji *Chi-Square* terhadap variabel dependen.

- b. Melakukan tahap Pemisahan (*Splitting*) data dengan variabel independen yang paling signifikan.
 - c. Melakukan tahap Penghentian (*Stopping*).
 - d. Menginterpretasikan pohon klasifikasi.
4. Menghitung nilai ketepatan klasifikasi dengan indikator akurasi pada metode CHAID.
5. Tahapan analisis data yang dilakukan dengan Algoritma C5.0 yaitu:
 - a. Pemilihan *root node* dimulai dengan perhitungan nilai *Entropy*
 - b. Menghitung nilai *Information Gain*
 - c. Menghitung nilai *Gain Ratio*
 - d. Menentukan kriteria menggunakan nilai *Gain Ratio* untuk mendapatkan kriteria kebaikan simpul
 - e. Mengulangi langkah a-c hingga semua kelas memiliki cabangnya masing-masing.
6. Melakukan perhitungan ketepatan klasifikasi dengan indikator akurasi pada Algoritma C5.0.
7. Melakukan interpretasi hasil
8. Membandingkan ketepatan klasifikasi pada metode CHAID dan Algoritma C5.0.