

Aplikasi Data Mining Dengan Metode CHAID Dalam Menentukan Status Kredit

Muhammad Rizki¹, Muhammad Isnaini Hadiyul Umam², Muhammad Luthfi Hamzah³

^{1,2} Program Studi Teknik Industri, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293

Email: muhammad.rizki@uin-suska.ac.id, muhammad.isnaini@uin-suska.ac.id

³Sistem Informasi, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293

Email : muhammad.luthfi@uin-suska.ac.id

ABSTRAK

Seiring dengan digalakkannya Industrial 4.0, data mining menjadi topik yang hangat untuk dibahas dikalangan peneliti. Perkembangan teknologi yang begitu cepat memaksa kita untuk dapat mengambil keputusan dengan cepat pula. Kredit macet menjadi salah satu resiko terbesar lembaga keuangan. Resiko kredit macet ini wajib diminimalisir dengan menganalisa faktor status nasabah berdasarkan data personalnya, sehingga dapat dilakukan klasifikasi berdasarkan hubungan antar faktor tersebut. Salah satu kunci utama memenangkan persaingan pasar yaitu dengan menentukan target pasar. Data mining menyediakan banyak alat bantu untuk klasifikasi, salah satunya dengan menggunakan metode analisis CHAID (*Chi-square Automatic Interaction Detection Analysis*). Diagram pohon keputusan yang dihasilkan dari Analisis CHAID dapat memberikan informasi tentang derajat hubungan antara *variable independent* dan *dependent*, serta informasi tentang karakteristik masing-masing kategori. Dalam hal ini, analisis CHAID digunakan untuk menentukan klasifikasi nasabah berdasarkan status kredit nasabah sebagai *variable* terikat dan data pribadi nasabah sebagai *variable* bebas. Dengan menggunakan uji *Chi-square*, dari total 7 *variables independent*, hanya 5 *variable* yang signifikan dengan *variable dependent*. *Variable-variable* tersebut adalah *variable independent* usia, pekerjaan, pendidikan, jangka waktu dan jumlah pinjaman. Berdasarkan hasil analisis CHAID didapatkan empat kelas. Kelas nasabah dengan pekerjaan sebagai (Aparatur Sipil Negara) ASN merupakan kelas yang memiliki resiko kredit macet yang paling minimal.

Kata Kunci: Data mining, Kredit, CHAID, Klasifikasi

ABSTRACT

Industrial 4.0, data mining has become a hot topic to discuss. The rapid development of technology forces us to make decisions quickly too. Bad Debt is one of the biggest risks in the banking area. The risk of bad debt must be minimized by analyzing status customer based on personal data so we can classify based on the correlation between these factors. One of the strategy to overcome in market is determining market kelompokation. Data mining provides many tools for classification, one of which is using the CHAID (Chi-square Automatic Interaction Detection Analysis) method. The decision tree diagram generated from the CHAID analysis can provide information about the degree of relationship between the independent variable and the dependent variable and information on the characteristics of each class. In this study, CHAID analysis is used to determine customer classification based on the customer's credit status as the dependent variable and customer personal data as the independent variable. By using the Chi-square test, out of a total of 7 independent variables, only 5 variables are significant with the dependent variable. These variables are the independent variables of age, occupation, education, period and loan amount. Based on the results of the CHAID analysis, four classes were obtained. The class of customers who work as ASN is the class that has the least risk of bad credit.

Keywords: Data Mining, Credit, CHAID, Classification

Pendahuluan

Selama ini dalam menjalankan kegiatan perkreditan banyak kredit yang belum dikembalikan sesuai jadwal yang telah ditentukan. Adanya kredit macet tersebut telah menimbulkan kerugian bagi

lembaga keuangan terkait. Dalam hal pembiayaan, tingkat pembiayaan yang lebih tinggi disertai dengan risiko kredit macet yang sangat besar. Risiko kredit ini harus diminimalisir agar lembaga keuangan dapat menjaga kelangsungan usahanya agar dapat

menyediakan produk pembiayaan kepada nasabah sasaran

Menghadapi kenyataan tersebut di atas, maka perlu dilakukan prediksi terhadap status kredit macet lembaga keuangan. Perkiraan status kredit merupakan upaya yang dilakukan oleh lembaga perkreditan untuk menghindari kredit macet atau tunggakan pinjaman yang jatuh tempo. Oleh karena itu, prediksi status kredit macet sangat berguna untuk mengurangi risiko kredit macet dan menempatkan kelompok nasabah yang tepat untuk meminimalkan risiko kredit macet.

Beberapa penelitian terdahulu yang terkait dengan prediksi status kredit yaitu Yohanes dan Siti [1], mempelajari pembentukan kelompokkasi dengan aplikasi analisis CHAID dengan permasalahan penentuan kelompokkasi nasabah potensial, dimana kelompokkasi nasabah potensial akan meminimumkan resiko kredit macet. Kemudian Lehmann dan Eherler [2], mempelajari profil responder dengan membandingkan CHAID dan analisis *dependensi* dalam proses pengambilan keputusan. Yohanes dan Ingrid [3], mengetahui kelompokkasi user berdasarkan kepuasa dengan Analisis Chaid. Kemudian Hoare [4], menggunakan metode CHAID untuk membuat klasifikasi dalam dataset. Algoritma CHAID telah terbukti menjadi pendekatan yang efektif untuk memperoleh segmentasi yang cepat tetapi bermakna di mana segmen didefinisikan dalam istilah demografis atau variabel lain yang memprediksi tunggal variabel kriteria kategori (dependen) [5][8].

Untuk meminimalkan risiko kredit dengan memfokuskan pada faktor status nasabah berdasarkan status kredit, kelompokkasi pasar akan didasarkan pada hubungan antara faktor-faktor tersebut (*variable* terikat) dan *variable* prediktor (*variable* terikat). Dengan demikian, metode *dependensi* dapat dijalankan dalam proses menghasilkan kelompok berdasarkan *variable independent* yang memiliki pengaruh besar terhadap *variable independent*. Namun, sebagai teknik eksplorasi, atau sebagai teknik pilihan terakhir ketika metode tradisional gagal, pohon klasifikasi, menurut pendapat banyak peneliti, tak tertandingi [6][7].

Penelitian sebelumnya yang dilakukan adalah mengidentifikasi kelompok nasabah potensial, dengan harapan dapat meminimalkan risiko kredit macet dengan menggunakan CHAID. Dalam penelitian ini *variable* bebas yang digunakan sebagai unsur penyusun adalah jenis kelamin, usia, tingkat pendidikan, pekerjaan, sektor kredit dan masa kredit[9][10].

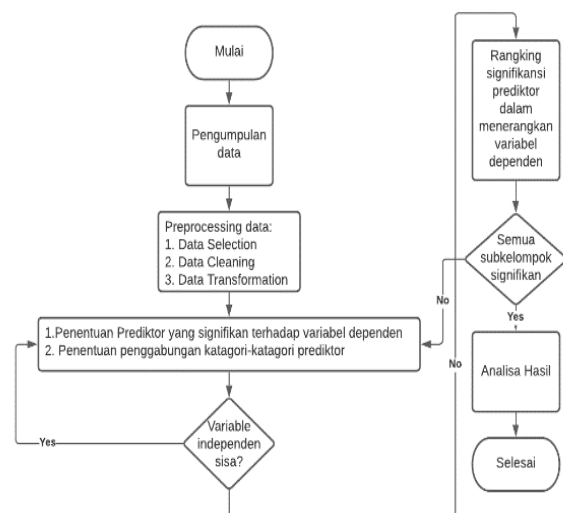
Dalam penelitian ini, analisis CHAID akan digunakan untuk menentukan kelompokkasi pelanggan. Kelompokkasi ini bertujuan untuk mengidentifikasi kelompok nasabah potensial dengan harapan dapat meminimalkan risiko kredit macet.

Metode Penelitian

Data yang diperlukan untuk penelitian (data nasabah pengguna pembiayaan pemohon kredit. Data ini kemudian dijadikan *variable* dalam analisis CHAID[11]. *Variable* dalam penelitian ini meliputi *variable dependent* dan *variable independent* yaitu:

1. *Variable Dependent*
 - a. Nasabah dengan status pinjaman macet (M).
 - b. Nasabah dengan status pinjaman tidak macet (TM).
2. *Variable Independent*
 - a. Jenis Kelamin (Laki-laki dan Perempuan)
 - b. Usia (< 23 tahun, 24- 30 tahun, 31 – 40 tahun, 41 – 50 tahun, > 51 tahun)
 - c. Pekerjaan (ASN, Pegawai Swasta, Wiraswasta, Pedagang, Petani)
 - d. Tingkat pendidikan (SD, SMP, SMA, Diploma, S1/S2/S3)
 - e. Sektor Kredit (konsumtif, jasa perdagangan pertanian, industri)
 - f. Tenor pembiayaan (5 hingga 100 bulan)
 - g. Total kredit (<Rp. 500.000, Rp. 500.000 hingga Rp. 1.000.000, Rp. 1.000.001 hingga Rp. 2.000.000, >Rp. 2000.000)

Preprocessing data meliputi pemilihan data, pembersihan data dan konversi data. Melalui penggabungan, pemisahan, dan penghilangan langkah, analisis *chi-square* digunakan untuk menentukan prediktor yang signifikan dari *variable dependent*. Hasil klasifikasi di pohon CHAID akan ditampilkan dalam diagram pohon. Penentuan target pasar didasarkan pada hasil kalsifikasi karakteristik pelanggan saat menganalisis bagan CHAID yang terbentuk[12]



Gambar 1. Diagram alir penelitian

Hasil dan Pembahasan

Dengan adanya uji *Chi-square* dimaksudkan untuk mengetahui hubungan antara status pinjaman kredit yang terbentuk dengan karakteristik dari nasabahnya. Penentuan dan penyusunan deskriptor tiap kelas dilakukan dengan menggunakan metode tabulasi silang menggunakan CROSSTAB SPSS[13].

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	45.240 ^a	4	.000
Likelihood Ratio	35.026	4	.000
Linear-by-Linear Association	29.645	1	.000
N of Valid Cases	2421		

a. 1 cells (10.0%) have expected count less than 5. The minimum expected count is .54.

Gambar 2. Hasil *Chi-square* test untuk usia

Dari gambar 2. diketahui bahwa *Chi-square* hitung (hasil output SPSS) adalah sebesar 45,240 sedangkan *Chi-square* tabel (hasil tabel *Chi-square*) sebesar 9,488. Karena *Chi-square* hitung lebih besar dari *Chi-square* tabel ($45,361 \geq 9,488$), maka H_0 ditolak berarti terdapat hubungan yang signifikan antara status dengan usia. Kesimpulan yang sama terdapat pada *variable-variable* pekerjaan dan tingkat Pendidikan.

Namun kesimpulan yang berbeda terdapat pada *variable* jenis kelamin dimana *Chi-square* hitung lebih kecil dari *Chi-square* tabel yaitu $0,007 \leq 3,841$, maka H_0 diterima berarti tidak terdapat hubungan yang signifikan antara status dengan jenis kelamin nasabah. Demikian pula *variable* sektor kredit dimana *Chi-square* hitung (hasil output SPSS) adalah sebesar 3,700 sedangkan *Chi-square* tabel (hasil tabel *Chi-square*) sebesar 9,488, maka H_0 diterima berarti tidak terdapat hubungan yang signifikan antara status dengan sektor kredit.

Variable jangka waktu diketahui bahwa *Chi-square* hitung (hasil output SPSS) adalah sebesar 28,316 sedangkan *Chi-square* tabel (hasil tabel *Chi-square*) sebesar 19,675. Karena *Chi-square* hitung lebih besar dari *Chi-square* table, maka H_0 ditolak berarti terdapat hubungan yang signifikan antara status dengan jangka waktu. Pada *variable* jumlah pinjaman diketahui bahwa *Chi-square* hitung (hasil output SPSS) adalah sebesar 7,977 sedangkan *Chi-square* tabel (hasil tabel *Chi-square*) sebesar 7,815. Karena *Chi-square* hitung lebih besar dari *Chi-square* tabel ($7,977 > 7,815$), maka ditolak berarti terdapat hubungan yang signifikan antara status dengan jumlah pinjaman. Maka dalam analisis diagram pohon CHAID *variable-variable* yang tidak signifikan tidak diikut sertakan.

Analisis CHAID

Tahap awal dalam analisis CHAID adalah tahap penggabungan[14]. Dalam kasus ini *variable* pekerjaan dibagi menjadi 5 kategori:

1. ASN
2. Pegawai Swasta
3. Wiraswasta
4. Pedagang
5. Petani

Setelah tahap analisis CHAID, *variable-variable* tersebut kemudian dikelompokkan menjadi 3 kategori, yaitu :

1. Kategori Wiraswasta
2. Kategori Pegawai Swasta
3. Kategori gabungan dari kategori petani, wiraswasta dan kategori pedagang

Hal ini berarti bahwa *variable* pekerjaan kategori petani, wiraswasta, dan pedagang memenuhi syarat kesignifikanan *chi-square* untuk bisa digabung menjadi satu kategori tunggal berdasarkan analisis CHAID yang ada.

Penggabungan kategori juga terjadi pada *variable* lainnya, yaitu Jumlah Pinjaman. Dalam penelitian ini *variable* jumlah pinjaman dibagi menjadi 4 kategori yaitu :

1. < Rp. 500.000,-
2. Rp. 500.000,- hingga Rp. 1.000.000,-
3. Rp. 1.000.001,- hingga Rp. 2.000.000,-
4. Rp. 2.000.000,-

Dimana *variable* ini oleh CHAID dikelompokkan menjadi 2 kategori yaitu:

1. < Rp. 2.000.000,-
2. > Rp. 2.000.000,-

Hal ini berarti bahwa *variable* jumlah Pinjaman kategori 1, 2, dan 3 memenuhi syarat untuk bisa digabung menjadi satu kategori berdasarkan analisis CHAID yang ada.

Dari hasil penentuan skala *variable dependent* dan *variable independent* dapat kita ketahui bahwa pada penelitian ini sebenarnya terdapat 5 *independent*. Dari hasil *Chi-square* dan diagram pohon CHAID menunjukkan bahwa ada 2 *variable independent* yang signifikan terhadap *variable dependent*, yaitu *variable* pekerjaan, dan jumlah pinjaman. Hal ini juga dapat dilihat dari hasil diagram pohon CHAID, bahwa pohon klasifikasi tersebut mempunyai 2 kedalaman, di mana *variable* pekerjaan membagi status kredit pada kedalaman ke-1, kemudian *variable* jumlah pendidikan pada kedalaman ke-2. Sehingga ada tiga *variable independent* lainnya yang tidak mempunyai hubungan *significant* dengan *variable dependent*, yaitu *variable* usia, *variable* pendidikan, *variable* jenis kelamin, *variable* sektor kredit, dan *variable* masa kredit.

Tahap Pemisahan

Pada diagram klasifikasi CHAID, *variable-variable independent* yang berperan dalam pemisahan adalah pekerjaan pada kedalaman pertama, dan jumlah pinjaman pada kedalaman kedua. Hasil klasifikasi CHAID memperlihatkan *variable independent* pertama yang terpilih dalam memisahkan nasabah berdasarkan status nasabah adalah pekerjaan. Pekerjaan merupakan *variable independent* yang mempunyai nilai *p-value* terkecil

dengan nilai 0.000 dan nilai *chi-square* terbesar yaitu 340.681. Kemudian pekerjaan membagi status nasabah dengan menggunakan kategori-kategori yang telah digabung secara optimal menjadi sub kelompok yang baru[15].

Pada sub kelompok yang baru tersebut, dilakukan tahap penggabungan ulang berdasarkan *variable independent* yang belum terpilih dengan proses yang sama seperti yang dijelaskan pada sub bahasan sebelumnya. Kemudian *variable* jumlah pinjaman merupakan *variable independent* selanjutnya yang terpilih dalam memisahkan nasabah berdasarkan status nasabah menjadi sub kelompok baru lagi.

Kesignifikanan *variable* pekerjaan dan *variable* jumlah pinjaman dalam membagi status nasabah dapat ditunjukkan pada nilai *p-value* dan *chi-square*. Apabila nilai *p-value* dari kedua *variable independent* tersebut kurang dari nilai signifikansi (α) 0.05 yang berarti terdapat hubungan antara kedua *variable independent* tersebut dengan status nasabah.

Berdasarkan pemisahan tersebut, sebanyak 1084 nasabah mempunyai pekerjaan kategori 1 (ASN), sebanyak 138 nasabah mempunyai pekerjaan kategori 2 (pegawai swasta), dan sebanyak 1199 nasabah mempunyai pekerjaan kategori 3 (Petani, wiraswasta, Pedagang). Nasabah yang mempunyai pekerjaan kategori 1 yaitu nasabah dengan pinjaman yang berstatus tidak lancar (MACET) berjumlah 4 nasabah (0.4%), sedangkan 1080 nasabah (99.6%) merupakan nasabah dengan pinjaman yang berstatus lancar (TIDAK MACET). Selanjutnya nasabah yang mempunyai pekerjaan kategori 1 dibagi menjadi 2 kelompok berdasarkan penggabungan optimal *variable* jumlah pinjaman yaitu nasabah dengan jumlah pinjaman kurang dari Rp. 2.000.000 berjumlah 808 nasabah (0% nasabah dengan kredit macet dan 100% nasabah dengan kredit tidak macet) dan jumlah pinjaman lebih dari Rp. 2.000.000 (1.4% untuk nasabah dengan pinjaman yang berstatus macet dan 98.6% untuk nasabah dengan pinjaman yang berstatus tidak macet).

Tahap Pemberhentian

Pemberhentian dilakukan ketika semua sub kelompok telah dianalisis dan juga telah berisi pengamatan-pengamatan dengan jumlah yang sedikit. Dan pada analisis CHAID, pertumbuhan pohon terhenti dengan dua kedalaman yang merupakan kedalaman maksimum untuk pertumbuhan pohon pada SPSS[16].

Kelompok Nasabah

Kelompok pada nasabah pembiayaan di lembaga keuangan yang dihasilkan dengan metode CHAID telah diterangkan pada diagram pohon klasifikasi CHAID. Berdasarkan diagram tersebut nasabah dibagi menjadi 4 kelompok dengan karakteristik seperti yang nampak pada tabel 1.

Tabel 1. Tabel jumlah nasabah dengan pinjaman yang berstatus lancar dan tidak lancar

Kelompok	Jumlah Status	Jumlah Status
	<u>Kredit Macet</u>	<u>Kredit Tidak Macet</u>
1	0	808
2	4	272
3	50	88
4	64	1135
Total	118	2303

Target Pasar

Dari empat kelompok yang terbentuk, dapat dihitung jumlah masing-masing nasabah dengan pinjaman yang berstatus macet dan tidak macet dalam tabel 2. Dari tabel tersebut diketahui bahwa kelompok yang memenuhi untuk dijadikan acuan dalam sasaran nasabah adalah pada kelompok 1. Hal ini berdasarkan acuan bahwa untuk menentukan nasabah baru yang dapat memperkecil kredit macet, maka akan dipilih kelompok-kelompok dengan jumlah nasabah dengan status kredit macetnya terkecil.

Nasabah pada kelompok 1 mempunyai karakteristik sebagai nasabah dengan pekerjaan ASN dan jumlah pinjaman kurang dari Rp. 2.000.000 lebih bisa dipercaya untuk dijadikan sebagai nasabah karena mereka mempunyai penghasilan tetap perbulannya. Disisi lain, jumlah pinjaman kurang dari Rp. 2.000.000 lebih dapat dilunasi tepat waktu daripada jumlah pinjaman yang lebih dari Rp.2.000.000.

Jadi untuk selanjutnya, lembaga keuangan dapat memberikan prioritas pada calon nasabah dengan karakteristik sesuai pada kelompok 1 supaya jumlah pembiayaan dengan status pinjaman tidak lancar dapat diminimalisir.

Dari data perusahaan diketahui jumlah nasabah dengan status kredit tidak macet terbanyak adalah nasabah dengan pekerjaan ASN. Kelompok terbaik sebagai prioritas strategi pasar adalah kelompok 1, yaitu nasabah dengan pekerjaan ASN dan jumlah pinjaman kurang dari Rp. 2.000.000. Hal ini memperlihatkan bahwa keadaan yang terjadi di dalam perusahaan telah sesuai dengan analisis yang dihasilkan

Kesimpulan

Berdasarkan hasil pengolahan data dan analisis yang telah dilakukan, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Pada kasus ini, analisis CHAID mendapatkan empat kelompok yang berbeda, yaitu:

- a. Karakteristik nasabah yang memiliki pekerjaan sebagai ASN dengan jumlah pinjaman \leq Rp. 2.000.000.
 - b. Karakteristik nasabah yang memiliki pekerjaan sebagai ASN dengan jumlah pinjaman $>$ Rp. 2.000.000 .
 - c. Karakteristik nasabah dengan pekerjaan sebagai Pegawai Swasta.
 - d. Karakteristik nasabah dengan pekerjaan sebagai Petani, Wiraswasta, dan Pedagang.
2. Untuk meminimumkan jumlah pinjaman dengan status kredit macet di masa yang akan datang maka Lembaga keuangan sebaiknya memprioritaskan calon nasabah dengan karakteristik nasabah yang memiliki pekerjaan sebagai ASN yang lebih pasti pendapatannya tiap bulan, karena dalam database lembaga keuangan sebagian besar anggota mempunyai pekerjaan sebagai ASN. Hal ini terlihat dari hasil analisis CHAID, kelompok ke-1 (node ke-1 pada diagram pohon CHAID) memiliki resiko kredit macet paling rendah sebesar 0,4% (4/1084) dan besar kelompok 44,8% (1084/2421) dari total nasabah.

Daftar Pustaka

- [1] Siti Nur, S. (2006). Metode CHAID Untuk Klasifikasi Data Mining (Studi Kasus: Bursa Efek Kuala Lumpur)
- [2] Eherler, D., & Lehmann, T. (2001, September). Responder profiling with CHAID and dependency analysis. In *Data Mining for Marketing Applications Workshop, University of Freiburg, Germany*. Retrieved on August (Vol. 30, p. 2012).
- [3] Hasana, S. N., & Kunto, Y. S. (2006). Analisis Chaid Sebagai Alat Bantu Statistika Untuk Segmentasi Pasar (Studi Kasus pada Koperasi Syarif Al-Hidayah). *Jurnal Manajemen Pemasaran*, 1(2).
- [4] Hoare, R. (2004, July). Using CHAID for classification problems. In *New Zealand Statistical Association Conference*, Wellington, New Zealand.
- [5] Magidson J. & Vermunt JK. (2006). An Extension of CHAID Treebased Segmentation Algorithm to Multiple Dependent variables. Departement of methodology and Statistic, Tilburg University, Netherland.
- [6] Hoare R. (2004). Using CHAID for Classification Problems. Wellington: Paper presented at New Zealand Statistical Association 2004 Conference.
- [7] Santosa, B. (2007). Data mining teknik pemanfaatan data untuk keperluan bisnis. *Yogyakarta: Graha Ilmu*, 978(979), 756.
- [8] Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12). Springer Science & Business Media.
- [9] Wirania, Y., Mara, M. N., & Kusnandar, D. (2013). Pembentukan pohon klasifikasi dengan metode CHAID. *BIMASTER*, 2(1).
- [10] Muhajir, M. (2016). Metode Improved CHAID (chi-squared automatic interaction detection) Pada Analisis Kredit Macet BMT (Baitul Mal wa Tamwil). *EKSAKTA: Journal of Sciences and Data Analysis*, 16(1), 55-63.
- [11] Pratiwi, Y. D. (2016). Klasifikasi Alumni FMIPA UNS Menggunakan Metode Chi-Squared automatic Interaction Detection (CHAID) dan Exhaustivechaid.
- [12] Atieh, M. A., Pang, J. K., Lian, K., Wong, S., Tawse-Smith, A., Ma, S., & Duncan, W. J. (2019). Predicting peri-implant disease: Chi-square automatic interaction detection (CHAID) decision tree analysis of risk indicators. *Journal of periodontology*, 90(8), 834-846.
- [13] Pebrianti, W. (2020). Demografi Pengguna Produk Digital Asli Kota Pontianak Metode Crosstabs. *Cakrawala Management Business Journal*, 3(1), 571-582.
- [14] Kurnia, L. (2017). Analisis Daya Tahan Mahasiswa dalam Melanjutkan Studi dengan Chi-Squared Automatic Interaction Detection. *Sainstek: Jurnal Sains dan Teknologi*, 8(1), 56-63.
- [15] Atti, A., Ginting, K. B., & Ndi, M. Z. (2019). Analisis Karakter Dosen Dalam Melakukan Penelitian Dengan Metode Chi-Square Automatic Interaction Detection (Chaid). *Jurnal Diferensial*, 1(1), 1-6.
- [16] Suhendra, M. A., Ispriyanti, D., & Sudarno, S. (2020). Ketepatan Klasifikasi Pemberian Kartu Keluarga Sejahtera Di Kota Semarang Menggunakan Metode Regresi Logistik Biner Dan Metode Chaid. *Jurnal Gaussian*, 9(1), 64-74.