

Received: 19 September 2022
Revised: 20 December 2022
Accepted: 29 December 2022
Published: 31 December 2022

Imrpoved Chi-Square Automatic Interaction Detection Pada Kasus Literasi Informasi

Zian Bula^{1, a)}, Resmawan^{2, b)}, La Ode Nashar^{1, c)}, Salmun K. Nasib^{1, d)}

¹ Prodi Statistika, Universitas Negeri Gorontalo, Bone Bolango 96119, Indonesia

² Prodi Matematika, Universitas Negeri Gorontalo, Bone Bolango 96119, Indonesia

E-mail: ^{a)}zianbula.statistik@gmail.com, ^{b)}resmawan@ung.ac.id, ^{c)}laode.nashar@ung.ac.id,
^{d)}salmun@ung.ac.id

Abstract

Information Literacy skills are needed to find quality sources and manage and sort information so that it can be used to improve the quality of life and community empowerment. The number of factors that affect information literacy causes the need for classification. The method used is Improved Chi-Square Automatic Interaction Detection (Improved CHAID), which aims to classify influencing factors with Information Literacy abilities. This study uses primary data, namely 237 Mananggu Young Generation (15-24 years), with Information Literacy as the dependent variable. The independent variables consist of reading interest, reading habits, gender, digital literacy, information needs, critical thinking, and information-seeking behavior. Based on the Improved CHAID analysis, the factors that significantly affect information literacy are Reading Habits (83%), Information Needs (89%), and Critical Thinking (94%). The classification performance of Testing Data is 40%, with a classification accuracy of 77% or from 95 samples, there are 73 samples that are properly classified. The sensitivity of 78% shows that the classification results are able to predict samples that have information literacy, 74% specificity indicates that the classification results are able to predict samples that do not have information literacy, and Press's Q 27.38 indicates a stable classification or statistically significant.

Keywords: Classification, Improved CHAID, Classification Performance, Information Literacy

Abstrak

Kemampuan Literasi Informasi dibutuhkan untuk mencari sumber yang bermutu, mengelola dan memilah informasi sehingga dapat dimanfaatkan dalam meningkatkan kualitas kehidupan dan pemberdayaan masyarakat. Banyaknya faktor-faktor yang mempengaruhi literasi informasi menyebabkan perlunya dilakukan klasifikasi. Metode yang digunakan adalah *Improved Chi-Square Automatic Interaction Detection (Improved CHAID)* yang bertujuan untuk mengklasifikasikan faktor-faktor pengaruh dengan kemampuan Literasi Informasi. Penelitian menggunakan data primer

yaitu 237 Generasi Muda Mananggu (15-24 tahun) dengan variabel dependen Literasi Informasi. Adapun variabel Independen yang terdiri dari Minat Baca, Kebiasaan Membaca, Jenis Kelamin, Literasi Digital, Kebutuhan Informasi, Berpikir Kritis, dan Perilaku Pencarian Informasi. Berdasarkan Analisis *Improved CHAID* diperoleh faktor-faktor yang signifikan mempengaruhi Literasi informasi yaitu Kebiasaan Membaca (83%), Kebutuhan Informasi (89%), dan Berpikir Kritis (94%). Kinerja klasifikasi dari *Testing Data* 40% dengan akurasi klasifikasi yaitu 77% atau dari 95 sampel terdapat 73 sampel tepat diklasifikasi. Adapun sensitifitas 78% menunjukkan hasil klasifikasi mampu memprediksi sampel yang memiliki Literasi informasi, spesifisitas 74% menunjukkan hasil klasifikasi mampu memprediksi sampel yang tidak memiliki Literasi informasi dan *Press's Q* 27,38 menunjukkan stabilnya klasifikasi atau signifikan secara statistik.

Kata-kata kunci: Klasifikasi, *Improved CHAID*, Kinerja Klasifikasi, Literasi Informasi

PENDAHULUAN

Indonesia berada dalam zaman informasi yang identik dengan literasi sehingga menggambarkan kemampuan dalam interaksi, komunikasi, dan aktualisasi yang dinyatakan melalui lisan maupun tertulis. Literasi mempunyai arti melek huruf atau kemampuan membaca dan menulis untuk memahami suatu informasi maupun menganalisa fenomena yang terjadi. Kemampuan literasi penting dalam keberhasilan generasi muda, karena memiliki keterampilan literasi yang baik dapat memahami informasi melalui lisan maupun tertulis. Penguasaan literasi yang baik juga dapat mendukung kompetensi-kompetensi yang dimiliki generasi muda. Kompetensi saling mendukung jika generasi muda menguasai literasi atau melek dan dapat memilih informasi sehingga mendukung keberhasilan hidup (Irianto & Febrianti, 2017).

Informasi semakin mudah diakses karena perkembangan teknologi informasi yang pesat. Sehingga dibutuhkan kemampuan dalam mencari sumber yang berkualitas, mengolah dan memilah informasi untuk dijadikan sarana dalam peningkatan kualitas hidup atau dalam pemberdayaan masyarakat. Karena tidak semua informasi yang disediakan pada media cetak maupun digital dapat digunakan untuk memenuhi kebutuhan. Sehingga diperlukan identifikasi terlebih dahulu sebagai dasar dalam pengembangan kemampuan Literasi Informasi pada kehidupan masyarakat. Literasi Informasi sering berkaitan dengan media digital tetapi kenyataannya sumber informasi bukan hanya melalui media digital tetapi media cetak seperti buku, modul, koran, majalah, dan lainnya. Media cetak sering juga dijadikan referensi informasi dalam membangun dan meningkatkan kualitas hidup (Mulyono 2020).

Berdasarkan hasil observasi pada 100 (seratus) generasi muda (15-24 tahun) dengan menggunakan kuisioner pada tanggal 7 Maret 2022 di Kecamatan Mananggu, diperoleh informasi bahwa 41% generasi muda yang sering mengikuti perkembangan informasi pada setiap hari, 44% jarang mengikuti dan 15% tidak pernah mengikuti perkembangan informasi. Banyak generasi muda Mananggu yang menggunakan media sosial yaitu 76% dari 100 responden. Media yang sering digunakan untuk mendapatkan informasi yaitu media digital karena mudah untuk mengaksesnya. Informasi diperoleh dari media digital yaitu *Google* (52%), *Facebook* (51%), *WhatsApp* (34%), *Youtube* (26%), dan *Instagram* (22%). Sebagian besar responden tidak kesulitan dalam mengakses informasi (66%) tetapi kemudahan penelusuran belum seimbang dengan melakukan penelusuran kebenaran dari informasi atau berita yang diperoleh, dikarenakan responden jarang melakukan penelusuran terhadap kenyataan informasi (50%) dan tidak pernah (18%). Responden mengutamakan membagikan informasi secara langsung melalui media sosial atau kepada kerabat yaitu 87% dari 100 responden. Hal tersebut mempengaruhi literasi informasi generasi muda di Mananggu, yang kemungkinan disebabkan oleh beberapa faktor yaitu minat baca, kebiasaan, jenis kelamin, literasi digital, kebutuhan informasi, berpikir kritis, dan perilaku pencarian informasi.

Banyaknya faktor-faktor yang mempengaruhi literasi informasi menyebabkan perlunya klasifikasi dari faktor-faktor tersebut terhadap literasi informasi. Salah satu metode yang dapat digunakan untuk pengklasifikasian faktor-faktor yang mempengaruhi literasi informasi generasi muda adalah metode *Improved Chi-Square Automatic Interaction Detection (Improved CHAID)*. Metode *Improved CHAID*

merupakan pengembangan yang digunakan pada CHAID yang memerlukan perbaikan karena memiliki kelemahan menghasilkan bias setelah proses pemilihan variabel independen (uji *Chi-Square*) yang dapat mengubah hasil interpretasi dari data pohon klasifikasi (Damayanti, 2018). Metode *Improved CHAID* menghasilkan output grafis dari pohon klasifikasi sehingga mempermudah interpretasinya karena dapat dilihat secara langsung pola dari pemisahan dan penggabungan variabel. Tahap analisis pada *Improved CHAID* dilakukan dengan cara yang hampir sama dengan analisis pada metode CHAID. Pada analisis *Improved CHAID* tahap kedua setelah pengujian Chi-Square dilanjutkan dengan uji *Tschuprow's*, setelah itu dilanjutkan dengan koreksi bonferroni kemudian pembentukan klasifikasi dari hasil segmentasi menjadi lebih baik (Muhajir, 2016). Dalam hal ini *Improved CHAID* dikembangkan untuk mengurangi bias sehingga hasil klasifikasi lebih baik.

Improved CHAID telah dilakukan untuk beberapa pengklasifikasian seperti kasus peningkatan rantai pasokan (Balasubramaniam, C.P. and Gunasundari, 2017), dan rumah tangga miskin (Yuliasari, 2021). Adapun pengklasifikasian metode CHAID digunakan pada kasus pencabutan layanan oleh pelanggan Indihome (Khodijatunnuriyah & Pratiwi, 2019), masa penyelesaian skripsi mahasiswa (Fitri 2020), penentuan target pasar (Rizki, Isnaini, Umam & Hamzah, 2020), penentuan segmentasi pariwisata (Díaz-p, García-gonz & Fyall, 2020), kelanjutan siswa mengikuti bimbingan belajar (Islamy 2021), termasuk pengaruh prestasi akademik mahasiswa (Fanggidae, Ekowati, Nenohai & Udil 2021). Metode CHAID juga dikombinasikan dengan metode lain, seperti pengklasifikasian pada kasus pemberian kartu keluarga sejahtera digunakan CHAID dan Regresi *Logistik Biner* (Suhendra 2020), status angkatan kerja digunakan CHAID dan *Regresi Logistik* (Juwita, 2021). Berbagai penelitian yang disebutkan sebelumnya menunjukkan bahwa metode *Improved CHAID* cukup baik atau akurat untuk pengklasifikasian.

Penelitian ini menggunakan *Improved CHAID* dengan uji *Chi-Square Koreksi Yates* dan *Tschuprow's* untuk klasifikasi faktor-faktor yang mempengaruhi Literasi Informasi pada generasi muda (15-24 tahun) di Kecamatan Mananggu. Faktor-faktor yang mempengaruhi literasi informasi yaitu minat baca, kebiasaan membaca, jenis kelamin, literasi digital, kebutuhan informasi, berpikir kritis, dan perilaku pencarian informasi. Pada penelitian ini dilakukan pengklasifikasian pada faktor-faktor yang signifikan mempengaruhi literasi informasi pada generasi muda Mananggu dan hasil klasifikasi faktor-faktor yang signifikan mempengaruhi literasi informasi pada generasi muda Mananggu menggunakan *Improved CHAID*, serta mengetahui kinerja klasifikasi menggunakan *Improved CHAID* berdasarkan *Testing Data*.

METODOLOGI

Bahan dan Data

Penelitian ini menggunakan data primer yang terdiri dari 237 (dua ratus tiga puluh tujuh) sampel atau generasi muda Mananggu (15-24 tahun). Variabel yang digunakan adalah variabel dependen literasi informasi (Y) dan variabel independen terdiri dari faktor-faktor yang mempengaruhi literasi informasi yaitu minat baca (X_1), kebiasaan (X_2), jenis kelamin (X_3), literasi digital (X_4), kebutuhan informasi (X_5), berpikir kritis (X_6), dan perilaku pencarian informasi (X_7). Pengambilan sampel menggunakan *sampling insidental* dengan alat bantu Instrumen Penelitian (kuisisioner). Sebelumnya peneliti telah melakukan Validasi Ahli instrumen penelitian pada tiga dosen ahli, dan telah dilakukan uji *Validitas* dan *Relibilitas* dengan 30 sampel yang menunjukkan instrumen *valid* dan *reliabel*.

Metode Penelitian

Penelitian ini dilaksanakan bulan Januari – September 2022. Pengolahan data menggunakan analisis *Improved CHAID*. Adapun tahapan pada penelitian ini yaitu sebagai berikut:

1. *Input* data penelitian.

2. *Splitting Data* menjadi *Training Data* digunakan untuk pengklasifikasian data menggunakan *Improved CHAID*, sedangkan *Testing Data* untuk menguji hasil klasifikasi yang telah dibentuk dari *Training Data*.
3. Menghitung uji *Chi-Square Koreksi Yates* untuk mengetahui ada tidaknya hubungan faktor-faktor pengaruh terhadap Literasi Informasi dan uji *Tschuprow's* untuk mengetahui tingkat hubungan pada hasil uji *Chi-Square Koreksi Yates*.
4. Tahap pemisahan yaitu menentukan *Split Node* pada *Training Data* berdasarkan nilai uji *Tschuprow's*. Adapun ketentuan dalam pemilihan *Split Node* yaitu:
 - a) Memilih variabel independen terbaik dengan nilai *p-value* terkecil, nilai *Chi-Square Koreksi Yates* terbesar atau yang paling signifikan, dan nilai *Tschuprow's* terbesar atau hubungan variabel independen dan variabel dependen lebih kuat yang digunakan sebagai pemisah *node*.
 - b) Jika tidak terdapat variabel independen dengan nilai *p-value* terkecil, *Chi-Square Koreksi Yates* terbesar atau yang paling signifikan, dan nilai *Tschuprow's* terbesar atau hubungan variabel independen dan variabel dependen lebih kuat, maka tidak dilakukan pemisahan dan *node* digunakan sebagai terminal *node* atau akhir dari *node* dan melanjutkan pada tahap *stopping*.
5. Tahap Penghentian yaitu pembentukan pohon klasifikasi akan dihentikan jika telah memenuhi aturan yaitu:
 - a) Jika semua sub grup telah berisi pengamatan yang jumlahnya telah dianalisis dan terlalu sedikit.
 - b) Jika nilai *p-value* variabel independen lebih besar dari nilai alpha maka tidak dapat melakukan pemisahan.
 - c) Setelah pohon mencapai batas maksimum spesifikasi maka proses pertumbuhan akan berhenti. Contoh peneliti menentukan untuk batas maksimal kedalaman pohon klasifikasi adalah sampai pada Level 3 (tiga) maka pertumbuhan pohon tersebut akan berhenti berdasarkan ketentuan tersebut.
6. Melakukan pengujian kinerja klasifikasi menggunakan *confusion matrix* (akurasi, sensitifitas, dan spesifisitas) dan *Press's Q*.
7. Menyimpulkan hasil dari klasifikasi.

HASIL DAN PEMBAHASAN

Splitting Data

Splitting Data bertujuan membagi data penelitian menjadi dua bagian yaitu *Training Data* dan *Testing Data*. *Training Data* digunakan untuk pengujian menggunakan analisis *Improved CHAID* sedangkan *Testing Data* digunakan untuk menguji kinerja klasifikasi yang telah dibentuk pada *Training Data*. Penelitian sebelumnya oleh Yuliasari (2021) yang melakukan pembagian data menggunakan 5 rasio (50%:50%, 60%:40%, 70%:30%, 80%:20%) menghasilkan akurasi klasifikasi terbaik yaitu rasio 60%:40%, maka penelitian ini menggunakan *Training Data* 60% dan *Testing Data* 40% dari data penelitian yang berjumlah 237 sampel.

Uji *Chi-Square* dan Uji *Tschuprow's*

Uji *Chi-Square* bertujuan untuk mengetahui independensi antara pasangan variabel. Diketahui data penelitian menggunakan variabel independen yang memiliki dua kategori sehingga tidak melakukan penggabungan kategori, sehingga lanjut pengujian independensi yang menggunakan uji *Chi-Square* (X^2) dengan *Koreksi Yates* karena tabel kontingensi berukuran 2×2 . Pasangan variabel dinyatakan signifikan ketika memiliki nilai *P-value* kurang dari tingkat signifikan (α) 0,05 atau 5% atau nilai dari X^2_{hitung} lebih besar dari X^2_{tabel} . Adapun pasangan variabel yang paling signifikan ditentukan dengan melihat nilai *P-value* terkecil atau X^2_{hitung} terbesar. Persamaan *koreksi Yates* adalah sebagai berikut (Aritonang, 2016):

$$X^2_{hitung} = \frac{N(n_{11}n_{22} - n_{12}n_{21})^2}{(n_{11} + n_{12})(n_{11} + n_{21})(n_{12} + n_{22})(n_{21} + n_{22})} \tag{1}$$

Keterangan:

- n_{ij} : Banyak pengamatan baris ke-i dan kolom ke-j
- N : Jumlah data pengamatan (*Training Data*)

Dilanjutkan uji *Tschuprow's* yang bertujuan untuk melihat tingkat hubungan pada masing-masing pasangan variabel yang dinyatakan signifikan. Menurut Tschuprow (1939) Secara umum, diperoleh persamaan sebagai berikut (Balasubramaniam, C.P. and Gunasundari, 2017):

$$T = \frac{X^2}{W\sqrt{(V-1) \times (U-1)}} \tag{2}$$

Keterangan:

- V: Kategori variabel independen
- U: Kategori variabel dependen
- W: Jumlah Sampel

Menurut Cahyono (2017) dalam (Yuliasari, 2021) bahwa pengkategorian hubungan antar variabel secara umum dapat mengikuti pengelompokan pada TABEL 1.

TABEL 1. Pengkategorian Hubungan antar Variabel

Besarnya Nilai Hubungan (Koefisien Korelasi)	Interpretasi Hubungan
$0,50 \leq KK < 1,00$	Kuat
$0,20 \leq KK < 0,50$	Sedang
$0,00 \leq KK < 0,20$	Lemah

Hasil uji *Chi-Square* dengan Koreksi Yates pada *Training Data* 60% ditunjukkan pada TABEL 2.

TABEL 2. Hasil *Chi-Square* Koreksi Yates

Pasangan Variabel	X^2 Hitung	X^2 Tabel	<i>P-value</i>	Keterangan
Minat Baca dan Literasi Informasi	12,7600	3,841	1,179E-04	Signifikan
Kebiasaan Membaca dan Literasi Informasi	21,9242	3,841	9,026E-07	Signifikan
Jenis Kelamin dan Literasi Informasi	1,9845	3,841	0,2282	Tidak Signifikan
Literasi Digital dan Literasi Informasi	3,7223	3,841	0,0258	Signifikan
Kebutuhan Informasi dan Literasi Informasi	6,7910	3,841	0,0041	Signifikan
Berpikir Kritis dan Literasi Informasi	17,0478	3,841	0,0000	Signifikan
Perilaku Pencarian Informasi dan Literasi Informasi	6,3949	3,841	0,0047	Signifikan

TABEL 2 menunjukkan bahwa variabel independen yang memiliki hubungan dengan variabel Literasi Informasi yaitu Minat Baca, Kebiasaan Membaca, Literasi Digital, Kebutuhan Informasi,

Berpikir Kritis, dan Perilaku pencarian informasi. Sedangkan variabel Jenis Kelamin tidak memiliki hubungan variabel Literasi Informasi. Pasangan variabel Kebiasaan Membaca dan Literasi Informasi memiliki nilai X^2_{hitung} terbesar dan nilai P -value terkecil. Hasil dari uji *Chi-Square koreksi Yates* diperoleh pasangan variabel signifikan untuk dilanjutkan pada uji *Tschuprow's* yang dapat dilihat pada TABEL 3.

TABEL 3. Hasil Pengujian uji *Tschuprow's*

Pasangan Variabel	<i>Tschuprow's</i>	Keterangan
Minat Baca dan Literasi Informasi	0,0899	Lemah
Kebiasaan Membaca dan Literasi Informasi	0,1544	Lemah
Literasi Digital dan Literasi Informasi	0,0262	Lemah
Kebutuhan Informasi dan Literasi Informasi	0,0478	Lemah
Berpikir Kritis dan Literasi Informasi	0,1202	Lemah
Perilaku Pencarian Informasi dan Literasi Informasi	0,0450	Lemah

Diketahui dari TABEL 3 bahwa seluruh pasangan variabel independen (minat baca, kebiasaan membaca, literasi digital, kebutuhan informasi, berpikir kritis, dan perilaku pencarian informasi) memiliki hubungan yang lemah terhadap variabel Literasi Informasi, tetapi yang memiliki nilai *Tschuprow's* terbesar yaitu pasangan variabel Kebiasaan Membaca dan Literasi Informasi yaitu 0,1544.

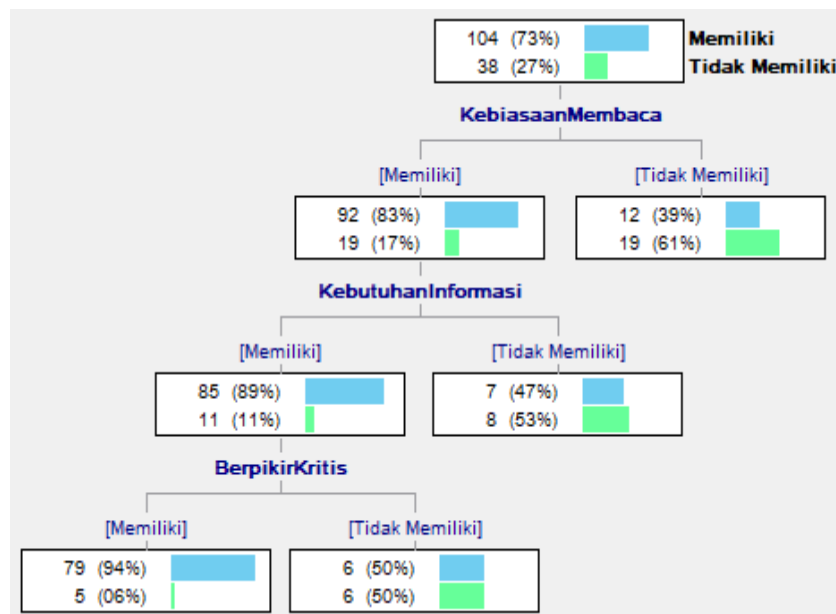
Pemisahan *Node* Pohon Klasifikasi

Tahap pemisahan bertujuan untuk membentuk pohon klasifikasi dengan memilih *Split Node*. Tahap pemisahan pada umumnya berdasarkan nilai *Koreksi Bonferroni* tetapi data penelitian yang digunakan variabel independennya hanya memiliki dua kategori, maka untuk penentuan *Split Node* dilihat dari nilai hitung uji *Tschuprow's*. Uji *Tschuprow's* merupakan pengujian tingkat hubungan pasangan variabel nominal dan koefisien estimasi sederhana (Bergsma, 2013) Adapun pasangan variabel signifikan yang memiliki nilai uji *Tschuprow's* terbesar pada *Training Data* 60% yaitu Kebiasaan Membaca dan Literasi Informasi (0,1544) yang berarti Kebiasaan Membaca dengan dua kategori menjadi pemisah dari variabel Literasi Informasi. Selanjutnya menentukan *Internal Node* yang menjadi pemisah dari variabel Kebiasaan Membaca dengan perhitungan yang sama seperti pemilihan *Split Node*.

Pemisah untuk variabel Kebiasaan Membaca adalah Kebutuhan Informasi dengan nilai uji *Tschuprow's* sebesar 0,1191 yang ditentukan berdasarkan kategori memiliki Kebiasaan Membaca, karena pada kategori tidak memiliki Kebiasaan Membaca variabel Kebutuhan informasi pada kategori kedua tidak memiliki data sehingga *Chi-Square Koreksi Yates* tidak dapat dihitung. Sedangkan Kebutuhan Informasi akan dipisah dengan variabel Berpikir Kritis dengan nilai uji *Tschuprow's* sebesar 0,1664 yang ditentukan berdasarkan kategori yang memiliki Kebutuhan Informasi, karena pada kategori tidak memiliki Kebutuhan Informasi variabel Berpikir Kritis menjadi tidak signifikan. Untuk variabel Minat Baca, Literasi Digital, dan Perilaku Pencarian Informasi tidak dapat dijadikan pemisah untuk variabel Berpikir Kritis karena menjadi tidak signifikan.

Penghentian Pohon Klasifikasi

Penghentian merupakan proses penentuan *Leaf Node* atau batas akhir pembentukan pohon klasifikasi. Penghentian dilakukan ketika pertumbuhan pohon sudah memenuhi peraturan atau sudah tidak ada lagi *Internal Node*. Hasil klasifikasi *Improved Chaid* dengan *training data* 60% dapat dilihat pada gambar pohon klasifikasi **GAMBAR 1**.



GAMBAR 1. Pohon Klasifikasi Improve CHAID

Dilihat pada **GAMBAR 1** bahwa dengan *Training Data* 60% menghasilkan jumlah data 142 responden. *Node* pertama pohon klasifikasi terdapat variabel dependen (literasi informasi) yang terbagi menjadi dua kategori yaitu memiliki literasi informasi sebesar 104 (73%) responden dan tidak memiliki literasi informasi 38 (27%) responden. Gambar di atas menunjukkan bahwa terdapat tiga variabel independen yang signifikan mempengaruhi Literasi Informasi yaitu Kebiasaan Membaca, Kebutuhan Informasi, dan Berpikir Kritis. Hasil pembentukan pohon klasifikasi *Improved CHAID Training Data* 60% menghasilkan 4 *segmen* yaitu sebagai berikut:

1. Dari 142 sampel terdapat 31 sampel yang tidak memiliki Kebiasaan Membaca dengan dibagi berdasarkan dua kategori Literasi Informasi yaitu yang tidak memiliki Literasi Informasi yaitu 61% atau 19 sampel, sedangkan yang memiliki Literasi Informasi 39% atau 12 sampel.
2. Dari 142 sampel terdapat 111 sampel yang memiliki Kebiasaan Membaca dengan dibagi berdasarkan dua kategori Literasi Informasi yaitu yang memiliki Literasi Informasi 83% atau 92 sampel, sedangkan 17% atau 19 sampel tidak memiliki Literasi Informasi, tetapi dari 111 sampel tersebut terdapat 15 sampel yang tidak memiliki Kebutuhan Informasi dengan dibagi berdasarkan dua kategori Literasi Informasi yaitu 53% tidak memiliki Literasi Informasi sedangkan 47% memiliki Literasi informasi.
3. Dari 142 sampel terdapat 111 sampel yang memiliki Kebiasaan Membaca dengan 96 sampel yang memiliki Kebutuhan Informasi. Adapun 96 sampel tersebut terdapat 84 sampel yang Berpikir Kritis dengan dibagi berdasarkan dua kategori Literasi Informasi yaitu 94% memiliki Literasi Informasi sedangkan sisanya 6% atau 5 responden tidak memiliki Literasi Informasi.
4. Dari 142 sampel terdapat 111 sampel yang memiliki Kebiasaan Membaca dengan 96 sampel yang memiliki Kebutuhan Informasi. Adapun 96 sampel tersebut terdapat 12 sampel yang tidak Berpikir Kritis dengan dibagi berdasarkan dua kategori Literasi Informasi yang memiliki jumlah yang sama yaitu 50% atau 6 sampel.

Kinerja Klasifikasi

Kinerja klasifikasi merupakan proses pengujian hasil klasifikasi yang telah dibentuk sebelumnya. Kinerja klasifikasi yang diuji berdasarkan *Testing Data* untuk melihat ketepatan hasil pengklasifikasian pada *Training Data*. Penelitian ini menggunakan dua statistik uji yaitu *Confusion Matrix* yang digunakan untuk melihat akurasi, sensitifitas, dan spesifisitas (Visa, Ramsay, Ralescu & Knaap, 2011). Uji *Press's Q* digunakan untuk melihat kestabilan hasil klasifikasi atau signifikan secara statistik. Perhitungan kinerja klasifikasi menggunakan *Confusion Matrix* dan *Press's Q* dapat dilihat pada

Lampiran 4. Adapun hasil pengujian kinerja klasifikasi *Improved CHAID Testing Data* 40% ditunjukkan pada TABEL 4.

TABEL 4. Kinerja Klasifikasi *Testing Data* 40%

Uji	Hasil
Akurasi	77%
Sensitifitas	78%
Spesifisitas	74%
<i>Press's Q</i>	27,38

Dilihat dari TABEL 4 bahwa hasil kinerja klasifikasi pada *Testing Data* 40% yaitu nilai akurasi 77%, yang menunjukkan bahwa dari 95 sampel terdapat 73 sampel tepat diklasifikasi. Adapun nilai sensitifitas menunjukkan keberhasilan kinerja klasifikasi *Improved CHAID* untuk pengklasifikasian data dengan kategori memiliki Literasi Informasi sebesar 78%, sedangkan nilai spesifisitas menunjukkan keberhasilan kinerja klasifikasi *Improved CHAID* untuk pengklasifikasian data dengan kategori tidak memiliki Literasi Informasi sebesar 74%. Hasil uji *Press's Q* 27,38 pada *Testing Data* menunjukkan bahwa lebih besar dari nilai X^2_{tabel} dengan db 1 atau 3,841 yang berarti hasil klasifikasi stabil atau signifikan secara statistik.

KESIMPULAN

Faktor-faktor yang signifikan mempengaruhi tingkat kemampuan Literasi Informasi generasi muda Manangu berdasarkan hasil analisis *Improved CHAID* yaitu Kebiasaan Membaca (83%), Kebutuhan Informasi (89%), dan Berpikir Kritis (94%).

Hasil klasifikasi menggunakan analisis *Improved CHAID* dengan *Training Data* 60% menghasilkan 4 *segmen* yaitu yang tidak memiliki Kebiasaan Membaca, tidak memiliki Kebutuhan Informasi, Berpikir Kritis, dan tidak Berpikir Kritis. Dalam hal ini menghasilkan 3 variabel independen yaitu Kebiasaan Membaca, Kebutuhan Informasi, dan Berpikir Kritis.

Hasil kinerja klasifikasi dari *Testing Data* 40% dengan akurasi klasifikasi yaitu 77% atau dari 95 sampel terdapat 73 sampel tepat diklasifikasi. Adapun sensitifitas 78% menunjukkan hasil klasifikasi mampu memprediksi sampel yang memiliki Literasi informasi, spesifisitas 74% menunjukkan hasil klasifikasi mampu memprediksi sampel yang tidak memiliki Literasi informasi dan *Press's Q* 27,38 menunjukkan stabilnya klasifikasi atau signifikan secara statistik.

REFERENSI

- Aritonang, N.E., 2016, 'Klasifikasi Status Kerja Pada Angkatan Kerja Kota Semarang Tahun 2014 Menggunakan Metode CHAID dan CART', *GAUSSIAN*, 5(1), 183–192.
- Balasubramaniam, C.P. and Gunasundari, D.R., 2017, 'Supply Chain Enhancement Using Improved Chaid Algorithm for Classifying the Customer Groups', *International Journal of Computer Applications Technology and Research*, 6(8), 2319–8656, doi.org/10.7753/IJCATR0608.1006.
- Bergsma, W., 2013, 'A bias-correction for Cram 'er ' s V and Tschuprow ' s T', *Journal of the Korean Statistical Society*, 42(3), 323–328, doi.org/10.1016/j.jkss.2012.10.002.
- Damayanti, C., 2018, 'Perbandingan Hasil Pembentukan Pohon Klasifikasi Metode CHAID dan Improved CHAID', *BIMASTER*, 7(4), 379–384.
- Díaz-p, F.M., García-gonz, C.G. & Fyall, A., 2020, 'The use of the CHAID Algorithm for Determining

- Tourism Segmentation: A Purposeful Outcome', *Heliyon*, 6(7), 1–11, doi.org/10.1016/j.heliyon.2020.e04256.
- Fanggidae, J.J.R., Ekowati, C.K., Nenohai, J.M.H. & Udil, P.A., 2021, 'Klasifikasi Faktor-faktor yang Mempengaruhi Prestasi Akademik Mahasiswa Pendidikan Matematika FKIP UNDANA dengan Metode CHAID', *Jurnal Matematika dan Pendidikan Matematika*, 2(1), 23–33.
- Fitri, H., 2020, 'Klasifikasi Faktor-Faktor Yang Mempengaruhi Masa Penyelesaian Skripsi Mahasiswa Dengan Metode CHAID', *Jurnal Sains Matematika dan Statistika*, 6(1), 9–22, doi.org/10.24014/jsms.v6i1.9248.
- Irianto, P.O. & Febrianti, L.Y., 2017, 'Pentingnya Penguasaan Literasi bagi Generasi Muda dalam Menghadapi MEA', *ELIC*, 640–647.
- Islamy, U., 2021, 'Penerapan Metode CHAID (Chi-Squared Automatic Interaction Detection) pada Kelanjutan Siswa yang Mengikuti Lembaga Bimbingan Belajar', *Prosiding Pendidikan Matematika dan Matematika*, 4, 1–7.
- Juwita, P., 2021, 'Ketepatan Klasifikasi Metode Regresi Logistik dan Metode CHAID dengan Pembobotan Sampel', *Indonesian Journal of Mathematics and Natural Sciences*, 44(1), 22–33.
- Khodijatunnuriyah, S. & Pratiwi, H., 2019, 'Klasifikasi Jenis Pencabutan Layanan oleh Pelanggan Indihome Menggunakan Metode Chi-Square Automatic Interaction Detection', *Indonesian Journal Of Applied Statistics*, 2(2), 80–93, doi.org/10.13057/ijas.v2i2.34526.
- Muhajir, M., 2016, 'Metode Improved CHAID (Chi-Square Automatic Interaction Detection) pada Analisis Kredit Macet BMT (Baitul Mal Tamwil)', *Eksakta:Jurnal ilmu-ilmu MIPA*, 55–63, doi.org/10.20885/eksakta.vol16.iss1.art7.
- Mulyono, D., 2020, 'Literasi Informasi dalam Kerangka Pengembangan Pendidikan Masyarakat', *Jurnal COMM-EDU*, 3(1), 2622–5492, doi.org/10.22460/comm-edu.v3i1.3516.
- Rizki, M., Isnaini, M., Umam, H. & Hamzah, M.L., 2020, 'Aplikasi Data Mining dengan Metode CHAID dalam Menentukan Status Kredit', *Jurnal Sains Teknologi dan Industri*, 18(1), 29–33, doi.org/10.24014/sitekin.v18i1.11421.
- Suhendra, M.A., 2020, 'Ketepatan Klasifikasi Pemberian Kartu Keluarga Sejahtera Di Kota Semarang Menggunakan Metode Regresi Logistik Biner dan Metode CHAID', *GAUSSIAN*, 9(1), 64–74, doi.org/10.14710/j.gauss.v9i1.27524.
- Visa, S., Ramsay, B., Ralescu, A. & Knaap, E. Van Der, 2011, 'Confusion Matrix-Based Feature Selection', *CEUR Workshop Proceefings*, 710, 120–127.
- Yuliasari, P.D., 2021, 'Klasifikasi Rumah Tanga Miskin di Kecamatan Kaubun Tahun 2020 dengan Menggunakan Metode Improved Chi-Square Automatic Interaction', *Eksponensial*, 12(1), 83–92.