

Penerapan Metode *CHAID* (*Chi-Squared Automatic Interaction Detection*) pada Kelanjutan Siswa yang Mengikuti Lembaga Bimbingan Belajar (Studi Kasus Nusagama Grup pada Masa Pandemi Covid-19)

Ulimazzada Islamy¹ Sekti Kartika Dini²
Universitas Islam Indonesia^{1,2}
18611128@students.uii.ac.id¹

Dampak dari pandemi Covid-19 sangat terasa terutama dalam aspek pendidikan. Model pembelajaran yang dulu tatap muka sekarang berubah menjadi pembelajaran secara daring. Hal ini berpengaruh terhadap lembaga pendidikan non formal salah satunya yaitu Nusagama Grup yang mengalami penurunan jumlah siswa yang cukup signifikan dibandingkan dengan tahun sebelumnya. Penelitian ini bertujuan untuk menentukan kelanjutan para siswa dalam mengikuti lembaga bimbingan berdasarkan karakteristik siswa lembaga serta banyaknya pertemuan yang dilakukan. Kelanjutan mengikuti lembaga bimbingan ini bertujuan untuk mengidentifikasi materi dan jenjang apa yang diminati para siswa sehingga akan lanjut mengikuti lembaga studi bimbingan Nusagama Grup. Penelitian ini menggunakan data sekunder dari 78 siswa. Metode yang digunakan adalah analisis *CHAID* (*Chi-Squared Automatic Interaction Detection Analysis*) yang akan digambarkan dengan diagram pohon keputusan. Berdasarkan hasil analisis metode *CHAID* maka diperoleh 4 segmen pengelompokan, tingkat nilai data *training* sebesar 75.8% dan nilai akurasi data *testing* sebesar 86.67%. Hasil presisi untuk klasifikasi siswa baru dari data *testing* sebesar 100% dan presisi klasifikasi siswa lama sebesar 81.8%. Jika dilihat dari nilai ketepatan yang didapatkan, dapat disimpulkan bahwa metode *CHAID* merupakan metode yang cukup baik untuk mengklasifikasi data pada penelitian ini.

Kata kunci: *CHAID, Klasifikasi, Pendidikan non-formal*

I. PENDAHULUAN

Pendidikan memiliki peran penting dalam mempersiapkan sumber daya manusia dalam pembangunan suatu negara [1]. Pendidikan merupakan salah satu pondasi sehingga semakin baik kualitas pendidikan yang diselenggarakan, maka semakin baik pula kualitas negara tersebut. Dalam UUD 1945 Pasal 31 dan UU No. 2 Tahun 2003 tentang Sistem Pendidikan Nasional, penyelenggaraan pendidikan menjadi tanggung jawab bersama antara 2 elemen yaitu pemerintah dan masyarakat. Pada UU tersebut juga dijelaskan bahwa pendidikan terdiri atas pendidikan formal, nonformal dan informal yang dapat saling melengkapi satu sama lain. Pendidikan formal merupakan jalur pendidikan yang terstruktur dan berjenjang yang terdiri atas pendidikan dasar, pendidikan menengah, dan pendidikan tinggi. Pendidikan informal merupakan jalur pendidikan keluarga dan lingkungan berbentuk kegiatan belajar secara mandiri. Pendidikan nonformal merupakan jalur pendidikan di luar pendidikan formal yang dapat dilaksanakan secara terstruktur dan berjenjang[2].

Adanya pandemi Covid-19 telah menghambat proses belajar dan pembelajaran baik pendidikan formal maupun pendidikan non formal [3]. Penutupan sementara lembaga pendidikan di seluruh Indonesia sebagai upaya pencegahan persebaran virus berdampak juga bagi bidang bisnis jasa pendidikan. Lembaga non formal seperti bimbingan belajar harus selalu siap dengan perubahan-perubahan yang dikeluarkan pemerintah di bidang pendidikan, juga harus siap bersaing dengan bimbingan belajar sejenis. Dalam meningkatkan kualitas pendidikan nasional, khususnya Provinsi Daerah Istimewa Yogyakarta, Nusagama Grup sebagai salah satu lembaga yang menyediakan layanan bimbingan belajar dengan waktu dan tempat yang bisa disesuaikan oleh siswa sendiri. Ada beragam materi yang disediakan oleh Nusagama Grup sehingga siswa dapat memilih materi apa yang ingin dipelajari.

Masalah yang dihadapi dalam lembaga bimbingan ini adalah penurunan siswa yang cukup signifikan dari tahun sebelum adanya pandemi Covid-19. Untuk itu, Nusagama Grup ingin mengetahui bagaimana klasifikasi akan kelanjutan siswa mengikuti bimbingan belajar. Klasifikasi adalah sebuah proses dengan tujuan mempelajari fitur dari sebuah atau sekelompok data yang telah diketahui kelasnya sehingga

menghasilkan sebuah model klasifikasi yang dapat digunakan untuk memprediksi kelas atau label data yang baru [4]. Klasifikasi ini membagi kumpulan dari nilai data independen kedalam sub kelas. Pada statistika ada beberapa metode klasifikasi, yaitu *CHAID* (*Chi-Square Automatic Interaction Detection*), *KNN* (*K-Nearest Neighbour*), *CART* (*Classification and Regression Tree*), dan *Naïve Bayes*.

Klasifikasi penelitian ini menggunakan metode *Chi-Squared Automatic Interaction Detection* atau dikenal dengan *CHAID* yaitu sebuah metode untuk mengklasifikasikan data kategori dimana tujuan dari prosedurnya adalah untuk membagi rangkaian data menjadi subgrup-subgrup berdasarkan pada variabel dependennya [5]. *CHAID* pertama kali dikenalkan oleh Dr. G. V. Kass di sebuah artikel berjudul “*An Exploratory Technique for Investigating Large Quantities of Categorical Data*” dalam buku *Applied Statistics* di tahun 1980. *CHAID* biasanya digunakan untuk membentuk segmentasi dengan membagi sebuah sampel menjadi dua atau lebih berdasarkan kriteria tertentu. Kemudian membagi kelompok tersebut menjadi kelompok yang lebih kecil berdasarkan variabel independennya [6]. Prosesnya ini berlanjut sampai tidak ada lagi variabel independen yang signifikan secara statistik.

Penelitian yang menjadi acuan dalam karya tulis ilmiah ini berjudul “Klasifikasi Tingkat Korban Kecelakaan Lalu Lintas dengan Metode *Chi-Square Automatic Interaction Detection (CHAID)* with *Synthetic Minority Oversampling Technique (SMOTE)*” yaitu penelitian terdahulu yang pernah dilakukan sebelumnya bertujuan untuk menganalisis kecelakaan lalu lintas dengan menggunakan metode *CHAID*. Berdasarkan hasil yang diperoleh, terdapat 46 segmen atau aturan. Model klasifikasi *CHAID* memiliki akurasi sebesar 81.5% dan memiliki proporsi yang cukup seimbang dan model sudah cukup baik [7].

Adapun penelitian yang berhasil ditemukan adalah penelitian dengan judul “Klasifikasi Siswa SMAN 1 Banuhampu yang Tidak Melanjutkan Pendidikan ke Perguruan Tinggi Lulusan Tahun 2015-2017 Menggunakan Analisis *CHAID*” yang bertujuan untuk mengetahui variabel-variabel yang memiliki hubungan terhadap SMAN 1 Banuhampu lulusan tahun 2015-2017 yang tidak melanjutkan pendidikan ke perguruan tinggi. Jenis data yang didapatkan yaitu data sekunder. Hasil dari penelitian menggunakan metode *CHAID* berupa diagram pohon yang terdiri dari 3 kelompok dengan variabel yang memiliki hubungan yaitu jurusan dan pekerjaan ibu, dimana kelompok tersebut yaitu: 1) siswa dari jurusan IPA yang ibu bekerja sebagai PNS/POLRI/TNI dan wiraswasta/wirusaha/dll, 2) siswa dari jurusan IPA yang ibu tidak bekerja dan ibu bekerja sebagai petani dan 3) siswa dari jurusan IPS [8].

Penelitian ketiga yang berkaitan dengan analisis dengan *CHAID* dengan judul “Penerapan Metode *CHAID (CHI-SQUARED AUTOMATIC INTERACTION DETECTION)* dan *CART (CLASSIFICATION AND REGRESSION TREES)* pada Klasifikasi Preeklampsia (Studi Kasus: Ibu Hamil di RS PKU Muhammadiyah Yogyakarta)”. Tujuan dalam penelitian ini untuk mengetahui klasifikasi faktor resiko apa saja yang berpengaruh terhadap Preeklampsia (hipertensi pada kehamilan). Hasilnya didapatkan bahwa pohon klasifikasi metode *CHAID* dan *CART* berbeda, serta metode *CART* memiliki pohon klasifikasi yang lebih simpleks dengan nilai akurasi yang tinggi dengan nilai akurasi *CHAID* sebesar 67% dan *CART* sebesar 74%. [9].

Pada penelitian ini, bertujuan untuk melakukan klasifikasi akan kelanjutan siswa mengikuti bimbingan belajar agar dapat mempertahankan eksistensi bisnis jasa di bidang pendidikan non formal pada masa pandemi COVID-19. Berdasarkan latar belakang di atas, penelitian yang berjudul “Penerapan Metode *CHAID (Chi-Squared Automatic Interaction Detection)* pada Kelanjutan Siswa yang Mengikuti Lembaga Bimbingan Belajar pada Masa Pandemi COVID-19” ini dilaksanakan, dengan harapan dapat memberikan manfaat bagi Nusagama Grup.

II. METODE PENELITIAN

Penelitian ini dilakukan pada tanggal 18 Januari 2021 sampai dengan 18 Februari 2021 di lembaga Nusagama Grup menggunakan data sekunder, yaitu data yang tersimpan dalam *database* di kantor Nusagama Grup. Adapun variabel yang digunakan dalam penelitian ini seperti pada tabel 1 berikut:

TABEL 1. VARIABEL PENELITIAN

Variabel	Kategori	Definisi Operasional
Y	Tipe Siswa	Jumlah perulangan siswa yang mengikuti bimbingan di Nusagama Grup. Dikatakan siswa lama jika telah melakukan setidaknya satu kali bimbingan dan siswa baru jika belum melakukan bimbingan sama sekali
X ₁	Jenis Kelamin	Jenis kelamin berupa laki-laki dan perempuan yang melakukan bimbingan di Nusagama Grup

X ₂	Durasi Pertemuan	Jumlah pertemuan dalam satuan jam yang diambil dari siswa yang mengikuti bimbingan belajar di Nusagama Grup.
X ₃	Tingkat Pendidikan	Jenjang siswa yang melakukan bimbingan di Nusagama Grup mulai dari TK, SD, SMP, SMA, Mahasiswa dan UMUM
X ₄	Materi	Bentuk bahan pembelajaran dalam kegiatan belajar mengajar pada Nusagama Grup
X ₅	Tipe Pembelajaran	Model pembelajaran pada Nusagama Grup, baik dilakukan secara <i>online</i> maupun <i>offline</i>

Metode penelitian yang digunakan adalah analisis statistika deskriptif yang merupakan metode yang berkaitan dengan pengumpulan dan penyajian suatu data sehingga memberikan informasi yang berguna [10]. Statistik deskriptif berfungsi untuk mendeskripsikan atau memberi gambaran terhadap objek yang diteliti melalui data sampel atau populasi [11]. Data yang disajikan dalam statistik deskriptif biasanya dalam bentuk ukuran pemusatan data [12]. Analisis deskriptif digunakan sebagai gambaran umum Nusagama Grup saat periode pandemi Covid-19 dari pengelompokan setiap variabel yang digunakan dalam data. Serta, analisis *CHAID* dengan *output* berupa diagram pohon klasifikasi yang menggambarkan pembentukan segmen. Diagram *CHAID* terdiri dari batang pohon (*tree trunk*) dengan membagi menjadi lebih kecil berupa cabang-cabang (*brances*).

Model klasifikasi dapat diketahui tingkat ketelitiannya dengan menggunakan tiga buah parameter, yaitu akurasi, *recall*, dan presisi. Akurasi merupakan ukuran dari seberapa bagus model bisa mengklasifikasikan dokumen dengan benar. Presisi merupakan tingkat ketelitian atau ketepatan dalam pengklasifikasian sementara *recall* memiliki fungsi untuk mengukur proporsi positif aktual yang benar diidentifikasi [13]. Berikut merupakan rumus akurasi, presisi dan *recall*:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

dimana:

TP: *true positive*

TN: *true negative*

FP: *false positive*

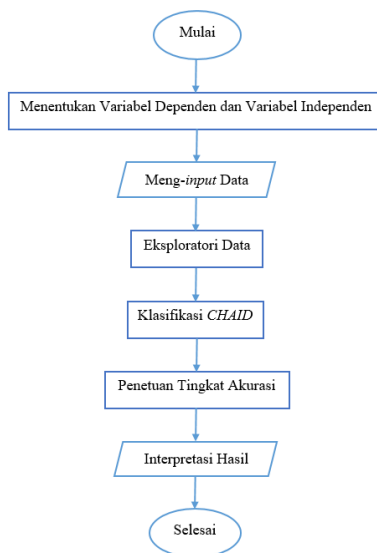
FN: *false negative*

Parameter akurasi, *recall*, dan presisi ketika memiliki nilai yang seimbang maka model klasifikasi dikatakan baik. Selain itu, *Confusion matrix* dapat juga digunakan untuk melakukan evaluasi kinerja dari klasifikasi. *Confusion matrix* merupakan alat ukur berbentuk matrix yang digunakan untuk memperoleh jumlah ketepatan klasifikasi terhadap kelas dengan algoritma yang dipakai

TABEL 2. *CONFUSION MATRIX*

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

Software yang digunakan dalam pengolahan data yaitu *Microsoft Excel* dan *Sipina Research*. Gambar 1 merupakan diagram alir atau langkah-langkah yang dilakukan dalam penelitian.



GAMBAR 1. DIAGRAM ALIR PENELITIAN

III. HASIL DAN PEMBAHASAN

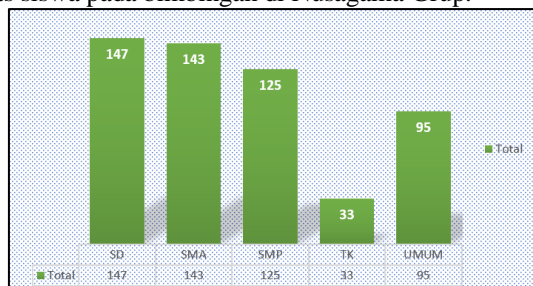
Pada penelitian ini akan dilakukan analisis deskriptif sebagai gambaran umum Nusagama Grup saat periode pandemic Covid-19 dari pengelompokan setiap variabel yang digunakan dalam data. Penelitian ini terdiri atas 78 siswa bimbingan belajar yang tercatat di lembaga Nusagama Grup. Berikut ini grafik yang ditampilkan sebagai hasil analisis deskriptifnya

A. Analisis Deskriptif



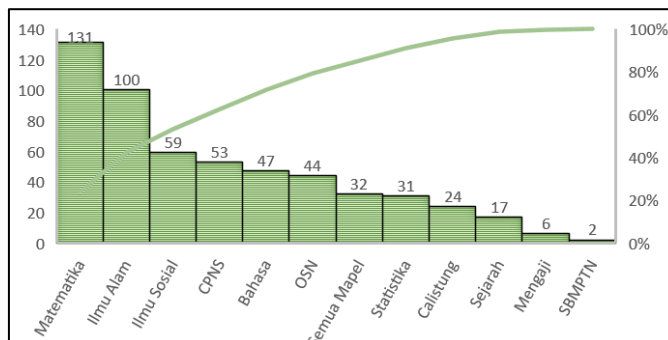
GAMBAR 2. GRAFIK JENIS KELAMIN SISWA

Berdasarkan grafik pada Gambar 2 dapat dilihat bahwa dari keseluruhan data siswa, jenis kelamin laki-laki merupakan mayoritas siswa pada bimbingan di Nusagama Grup.



GAMBAR 3. GRAFIK PERTEMUAN BERDASARKAN JENJANG PENDIDIKAN

Berdasarkan grafik pada gambar 3, dapat dilihat bahwa jumlah durasi pertemuan tertinggi bimbingan merupakan 147 jam yang dilakukan siswa pada tingkat pendidikan SD dan 143 jam pada tingkat pendidikan SMA. Pertemuan terendah merupakan jenjang TK yang menghabiskan waktu bimbingan 33 jam pertemuan. Siswa yang berada pada tingkat Umum mengikuti bimbingan belajar sebanyak 95 jam pertemuan.



GAMBAR 4. GRAFIK PERTEMUAN BERDASARKAN MATERI

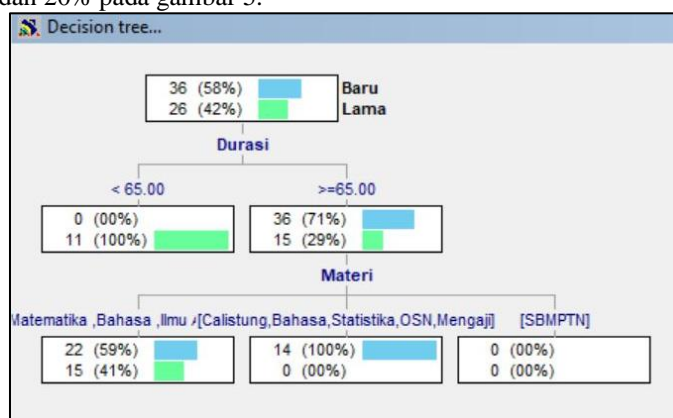
Grafik pada gambar 4 menjelaskan banyaknya menit pertemuan yang dilakukan berdasarkan materi yang diajarkan. Materi dengan pertemuan tertinggi yaitu matematika sebanyak 131 jam pertemuan dan materi dengan peminat terendah yaitu SBMPTN sebanyak 2 jam pertemuan. Hal ini dipengaruhi karena matematika merupakan ilmu pasti dan dianggap sebagai salah satu mata pelajaran yang sulit. Implementasinya juga penting untuk menyelesaikan banyak permasalahan. Sehingga, dari semua jenjang masih cenderung membutuhkan bimbingan matematika. Untuk materi SBMPTN, yang paling membutuhkan bimbingan merupakan siswa SMA. Siswa mulai fokus mempelajari SBMPTN setelah kelulusan.

B. Data Training dan Data Testing

Data yang digunakan merupakan data siswa bimbingan belajar di Nusagama Grup yang berjumlah 78 data. Data tersebut akan dibagi menjadi dua yaitu data *training* dan data *testing*. Dalam penelitian ini, akan dibagi menjadi 80% untuk data *training* dan sisanya yaitu 20% untuk data *testing*

C. Analisis CHAID

Didapatkan klasifikasi dari analisis *CHAID* yang menggunakan pembagian data *training* serta data *testing* sebesar 80% dan 20% pada gambar 5.



GAMBAR 5. POHON KEPUTUSAN METODE CHAID

Berdasarkan klasifikasi metode *CHAID* diatas, terbentuk 4 segmen yang berbeda yaitu sebagai berikut:

1. Jika durasi pertemuan kurang dari 65 menit atau 1.08 jam maka kemungkinan siswa lama akan mengambil materi bimbingan sebanyak 11 siswa.
2. Jika banyaknya jumlah pertemuan lebih dari 65 menit atau 1.08 jam dengan materi yang diambil yaitu Matematika, Bahasa dan Ilmu Alam maka kemungkinan ada 22 siswa baru dan 15 siswa lama yang mengambil materi tersebut.

3. Jika banyaknya jumlah pertemuan lebih dari 65 menit atau 1.08 jam dengan materi yang diambil yaitu Calistung, Statistika, OSN dan mengaji maka kemungkinan ada 14 siswa baru yang mengambil ketiga materi tersebut.
4. Jika banyaknya jumlah pertemuan lebih dari 65 menit atau 1.08 jam dan materi bimbingan yang diambil adalah SBMPTN maka tidak ada kemungkinan siswa akan mengambil materi tersebut.

D. Nilai Akurasi dan Presisi dari Analisis CHAID

Model yang diperoleh dari hasil analisis *CHAID* akan dilakukan perhitungan nilai akurasi dan presisi supaya dapat mengetahui sejauh mana model dalam mengklasifikasikan dengan benar. Berikut ini merupakan tabel tingkat akurasi dan presisi dengan menggunakan metode *CHAID*.

TABEL 3. NILAI AKURASI DAN PRESISI ANALISIS *CHAID*

Data	Hasil Seleksi	Hasil Prediksi		Hasil Akurasi
		Siswa Baru	Siswa Lama	
<i>Training</i>	Siswa Baru	36	0	75.8%
	Siswa Lama	15	11	
Data	Hasil Seleksi	Hasil Prediksi		Hasil Akurasi
		Siswa Baru	Siswa Lama	
<i>Testing</i>	Siswa Baru	4	2	86.67%
	Siswa Lama	0	9	

Berdasarkan tabel 3, hasil perhitungan akurasi dan presisi dapat dilihat bahwa klasifikasi yang terbentuk dari analisis *CHAID* menggunakan perbandingan data *training* dan data *testing* sebesar 80:20 diperoleh nilai akurasi data *training* sebesar 75.8% dan nilai akurasi data *testing* sebesar 86.67%. sehingga, hasil presisi untuk klasifikasi siswa baru dari data *testing* sebesar 100% dan presisi klasifikasi siswa lama sebesar 81.8% bisa dikatakan cukup baik. Oleh karena itu, dapat disimpulkan bahwa metode *CHAID* merupakan metode yang cukup baik untuk melakukan klasifikasi data pada penelitian ini.

IV. SIMPULAN DAN SARAN

Berdasarkan analisis tentang klasifikasi yang terbentuk pada kelanjutan para siswa yang mengikuti Lembaga bimbingan belajar pada masa pandemi Covid-19 yaitu tahun 2020 maka dapat ditarik kesimpulan sebagai berikut:

1. Secara keseluruhan dari empat indikator tipe siswa, mayoritas siswa yang mengikuti bimbingan belajar dengan jumlah jam pertemuan paling tinggi adalah tingkat pendidikan SD dan yang paling rendah adalah tingkat pendidikan TK. Materi yang paling digemari adalah Matematika dan Ilmu Alam.
2. Dari analisis *CHAID* yang dilakukan, kemungkinan paling banyak bahwa siswa lama akan mengikuti kembali bimbingan jika memilih pertemuan dengan waktu kurang dari 65 menit atau 1.08 jam, begitupun dengan siswa baru.
3. Berdasarkan hasil dari klasifikasi dapat disimpulkan bahwa klasifikasi menggunakan metode *CHAID* cukup baik karena diperoleh nilai akurasi data *training* sebesar 78.8% dan nilai akurasi data *testing* sebesar 86.67%. Sehingga, hasil presisi untuk klasifikasi siswa baru dari data *testing* sebesar 100% dan presisi klasifikasi siswa lama sebesar 81.8%.

DAFTAR PUSTAKA

[1] Musyaddad, K. "Problematika Pendidikan di Indonesia". *Education and Biology Journal*, 4, 51 – 57. April 2013.

[2] Darlis, A. "Hakikat Pendidikan Islam: Telaah antara Hubungan Pendidikan Informal, Non Formal dan Formal". *Jurnal Tarbiyah*. 2017.

[3] Musyafa Ali, C. P. Inovasi Model Pembelajaran Jarak Jauh Program Kesetaraan Paket C di Masa Pandemi Covid-19. *Jurnal Ilmiah Pendidik dan Tenaga Kependidikan Pendidikan Non Formal* , Vol 15 No 2. Agustus 2020.

[4] Bing, L. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers. 2012.

[5] Lehmann, T., & Eherler, D. (2001). Diambil kembali dari Responder Profiling with CHAID and Dependency. Analysis: www.informatik.unifreiburg.de/~ml/ecmlpkdd/WSProceedings/w10/lehm_ann.pdf. 2001.

[6] Hasna, Y. K. (2006). Analisis CHAID Sebagai Alat Bantu Statistika untuk Segmentasi Pasar. *Jurnal Manajemen Universitas Kristen Petra Surabaya*, Vol.1 No 2. 2006.

[7] Alfikri, B. *Dspace*. Diambil kembali dari <https://dspace.uui.ac.id/handle/123456789/6512>. April 2018

- [8] Yunanda. Klasifikasi Siswa SMAN 1 Banuhampu yang Tidak Melanjutkan Pendidikan ke Perguruan Tinggi Lulusan Tahun 2015-2017 Menggunakan Analisis CHAID. *Universitas Negeri Padang*. 2020.
- [9] Nazar, R. R. Penerapan Metode CHAID (CHI-SQUARED AUTOMATIC INTERACTION DETECTION) dan CART (CLASSIFICATION AND REGRESSION TREES) pada Klasifikasi Preeklampsia (Studi Kasus: Ibu Hamil di RS PKU Muhammadiyah Yogyakarta) . Diambil kembali dari DSPACE UII: <https://dspace.uii.ac.id/handle/123456789/6567>. April 2018.
- [10] Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2011). *Probability & Statistics for Engineers & Scientists 9th Ed.* USA: Pearson.
- [11] Sugiyono. Metode Penelitian Kuantitatif Kualitatif dan R&D. Bandung: Alfabeta. 2007.
- [12] Kuswanto, D. *Statistik Untuk Pemula dan Orang Awam*. Jakarta: Laskar Aksara. 2012.
- [13] Witten, I., & Frank, E. *Data Mining : Pratical Machine Learning Tools and Techniques Second Edition*. San Francisco: Morgan Kauffman. 2005.