

Penerapan Algoritma C4.5 Untuk Prediksi Penerimaan Beasiswa di SD 4 Pelangian

Selviana Yunita*¹, Vivi Nur Alaeyda²

^{1,2} Program Studi Sistem Informasi Universitas Darwan Ali
E-mail: *selviana.yunita@unda.ac.id, vnurlyda@gmail.com

Abstrak

Pendidikan merupakan salah satu faktor terpenting untuk meningkatkan kualitas dan mengembangkan potensi diri. Program wajib belajar di Indonesia mewajibkan anak Indonesia untuk menempuh pendidikan selama 12 tahun. Namun dalam implementasinya pemenuhan wajib belajar 12 tahun tidak sepenuhnya berjalan tanpa kendala, salah satu kendala yang banyak dialami oleh para siswa adalah biaya pendidikan. Dalam upaya menanggulangi kendala tersebut, maka pemerintah mengadakan program beasiswa yang ditujukan untuk anak-anak kurang mampu. Terdapat proses seleksi untuk pemberian program beasiswa ini oleh pihak sekolah, beasiswa yang dimaksud adalah Program Indonesia Pintar (PIP). Penelitian ini menggunakan algoritma data mining klasifikasi C4.5 dengan memprediksi penerimaan beasiswa di SD 4 Pelangian khususnya kelas 5. Kriteria yang digunakan adalah jenis kelamin, pekerjaan orang tua, pendapatan orang tua, rata-rata nilai raport per semester dari kelas 1 sampai dengan 4. Penelitian ini menghasilkan rule yang dapat membantu pihak sekolah untuk melakukan proses klasifikasi untuk menentukan kualifikasi siswa yang berhak menerima beasiswa. Metode algoritma C.45 ini dipilih karena mampu dalam membantu sebuah klasifikasi dengan melakukan prediksi hingga menghasilkan tingkat nilai akurasi yang cukup tinggi untuk memprediksi penerimaan beasiswa. Hasil penelitian menunjukkan akurasi sebesar 85,36%.

KataKunci—Klasifikasi, C4.5, Data mining, Beasiswa

Abstract

Education is one of the most important factors to improve quality and develop self-potential. The compulsory education program in Indonesia requires Indonesian children to study for 12 years. However, in its implementation, the fulfillment of the 12-year compulsory education is not completely without problems, one of the obstacles that many students experience is the cost of education. In an effort to overcome these obstacles, the government held a scholarship program aimed at underprivileged children. There is a selection process for the award of this scholarship program by schools, the scholarship in question is the Smart Indonesia Program (PIP). This study uses the C4.5 classification data mining algorithm by predicting scholarship receipts at SD 4 Pelangian, especially grade 5. The criteria used are gender, parental occupation, average grades of report cards per semester from grades 1 to 4, which can help the school to carry out the classification process to determine the qualifications of students who are eligible to receive scholarships. The C.45 algorithm method was chosen because it is able to help classification by making predictions to produce a high enough level of accuracy to predict scholarship acceptance. The results showed an accuracy of 85.36%.

Keywords—Classification, C4.5, Data mining, Scholarship

1. PENDAHULUAN

Pendidikan adalah menjadi factor terpenting untuk mengembangkan potensi dan kualitas diri seseorang. Kualitas pendidikan bagi siswa juga menjadi salah satu tujuan dari sekolah. Karena sekolah dapat meningkatkan kualitasnya dari tahun ke tahun sebagai tolak ukur dalam meningkatkan keberhasilan dalam sistem pendidikan. Pendidikan di Indonesia diwajibkan menempuh hingga 12 tahun, namun masih banyak masyarakat yang pada nyatanya tidak bisa melanjutkan sekolah atau berhenti karena masalah biaya pendidikan. Biaya pendidikan menjadi salah satu factor terpenting untuk mewujudkan keberhasilan tujuan dalam pendidikan [1].

Mengetahui betapa pentingnya kualitas pembelajaran dalam keinginan mencerdaskan kehidupan bangsa, maka pemerintah akan terus mencari solusi untuk mewujudkan amanat tersebut dengan mencoba berbagai usaha pembangunan pendidikan dengan menjadi yang lebih berkualitas seperti halnya, pengembangan kurikulum, sistem mengevaluasi, dan perbaikan sarana dalam pembelajaran. Tetapi ada beberapa factor yang mempengaruhi siswa dalam menempuh pembelajaran atau bersekolah yaitu keadaan ekonomi keluarga yang kurang memadai yang berpengaruh terhadap siswa yang berkeinginan untuk sekolah.

Lembaga yang bertanggung jawab dalam penyaluran ilmu pengetahuan ialah lembaga pendidikan. Dalam suatu lembaga pendidikan bertanggung jawab terhadap mengajar, mendidik dan memberikan hak maupun kewajiban kepada semua siswanya, bertujuan untuk meningkatkan dan menguatkan karakter seorang siswa, dengan memperluas ilmu pengetahuan dan menguatkan iman ketaqwaan pada tuhan Maha kuasa. Salah satu lembaga pendidikan ialah Sekolah Dasar, yang mendidik, mengajari dan memberikan hak maupun kewajiban dalam pendidikan dasar hingga menguatkan karakter anak.

Pada UUD 1945 pasal 31 ayat 1-2 berisi Setiap warga Negara Indonesia berhak mendapatkan pendidikan. Tetapi tidak seluruh warga Indonesia dapat berkesempatan menerima pendidikan yang layak dan sesuai [2]. Dengan demikian bagi siswa-siswi yang kurang mampu berhak untuk mendapatkan biaya pendidikan dan bagi siswa-siswi yang berprestasi berhak mendapatkan beasiswa [3]. Tapi, tidak bisa dihindari bahwa masih banyak anak yang belum dapat merasakan dan menikmati dunia pendidikan masalahnya karena kurangnya biaya maka dari itu perlukan dalam berbagi solusi, salah satu contohnya adalah program beasiswa. Beasiswa ialah sebuah pembiayaan yang tidak berasal dari dana sendiri ataupun orang tua siswa, karena beasiswa ini bertujuan untuk memberikan keringanan biaya pendidikan dan untuk memotivasi belajar anak bagi yang menerimanya. Beasiswa juga pemberian bantuan berupa uang yang diberikan secara perorangan dengan tujuan untuk kelanjutan dalam pendidikan yang ditempuh. Namun, sering terjadi pemberian beasiswa tidak tepat sasaran dikarenakan kesulitan dalam penyeleksian siswa.

Disetiap lembaga pendidikan banyak sekali yang menyumbangkan beasiswa yang ditujukan kepada siswa seperti beasiswa yang diberikan oleh Lembaga pemerintah, yayasan atau juga pada perusahaan. Salah satu program beasiswa di Indonesia adalah Program Indonesia Pintar (PIP). Program tersebut bermaksud bantuan yang berupa uang tunai untuk memenuhi kebutuhan pokok selama sekolah dari pemerintah diberikan pada peserta didik yang orang tuanya kurang mampu membiayai pendidikan anaknya dan beasiswa tersebut diseleksi oleh pihak sekolah. Beasiswa ini juga agar bisa dimanfaatkan sebaik mungkin bagi yang mendapatkannya dari bantuan tersebut, sehingga siswa dapat memenuhi kebutuhannya selama masih sekolah atau mengejar pendidikan. Banyaknya siswa yang mengajukan permohonan untuk mendapatkan beasiswa dan syarat penilaian yang banyak pun tidak semua siswa diterima untuk menerima beasiswa ini, sehingga pihak sekolah menyeleksi anak dari nilai akhir semester jika nilainya masuk kriteria maka itu dapat menerimanya[4].

Penelitian ini dilakukan di SD 4 Pelangsian, berlokasi di Kotawaringin Timur, Kalimantan tengah. Data yang diambil khususnya kelas 5 tahun ajaran 2020/2021 yang berjumlah 41 orang terbagi menjadi 2 kelas. Data yang dipakai adalah pekerjaan orang tua, penghasilan orang tua dan nilai akhir kelas yang akan diteliti tersebut diambil dari kelas 1 sampai kelas 4 per semesternya. Penelitian ini bertujuan untuk memprediksi siswa yang penerima beasiswa dengan menggunakan teknik *classification*

data mining, dengan algoritma C4.5. Penelitian ini menghasilkan rule yang dapat membantu pihak sekolah dalam proses klasifikasi untuk menentukan kualifikasi siswa yang dapat menerima beasiswa. Dengan adanya penelitian ini diharapkan dapat membantu pihak sekolah dalam mencari solusi dan bisa melihat tingkat siswa berprestasi, selain itu juga dapat meningkatkan kualitas pendidikan sekolah karena dengan adanya siswa-siswa berprestasi[5][6]. Tujuan adanya beasiswa ini, khususnya untuk anak sekolah dasar yaitu agar dapat membantu para pelajar ataupun mahasiswa untuk mencari ilmu sesuai dengan bidang yang mau dipelajari lebih dalam. Khususnya bagi yang memiliki permasalahan biaya, mewujudkan kesetaraan ilmu pengetahuan dan pendidikan kepada masing-masing siswa yang membutuhkan, menciptakan generasi penerus atau generasi baru yang lebih cerdas dan pintar.

Data mining adalah proses penggalian atau pengumpulan informasi pada sebuah data penting dengan kadar besar dan menggunakan berbagai macam bidang ilmu didalamnya seperti *artificial intelligence*, *machine learning* atau pembelajaran mesin, teknik statistika maupun matematika [7]. Pengertian lain tentang data mining adalah sebuah serangkaian yang dipekerjakan satu atau lebih dengan teknik pembelajaran computer untuk mengkaji dan menyaring pengetahuan secara otomatis atau serangkaian proses untuk menggali nilai tambah pada suatu kumpulan data yang berupa pengetahuan yang selama ini tidak diketahui dalam cara manual. Data mining ialah suatu pemrosesan yang menemukan hubungan yang beartikan kecenderungan sehingga memeriksa sekumpulan data besar yang penyimpanannya masih tersimpan pada teknik pemahaman pada pola seperti teknik matematika dan teknik statistika. Data mining juga adalah suatu kegiatan yang mendapatkan pola yang menarik dari sebuah data yang berukuran besar, penyimpanan data dapat tersimpan dalam *database*, *data warehouse* dan penyimpanan data atau informasi lainnya. Data mining juga didefinisikan sebagai sebuah proses yang menemukan pola-pola dalam data.

Adapun beberapa macam metode dalam data mining yaitu *classification*, *clustering*, *association*, *regression* dan *forecasting*. Istilah lain dari *data mining* dikenal dengan *KnowledgeDiscovery in Database* (KDD) yang merupakan proses penggalian dan menganalisis sebuah data berukuran besar yang dibantu oleh komputer dan mengidentifikasi pengetahuan dalam database. Selain itu, adapula tahapan proses dalam KDD ini yang menjadi salah satu bagian dari data mining. Berikut proses tahapan KDD seperti: Pembersihan data (*data cleaning*), Integrasi data (*data integration*), Pemilihan data (*data selection*), Transformasi data (*data transformation*), Penambangan data (*data mining*), Evaluasi pola (*pattern evaluation*), Presentasi pengetahuan (*knowledge presentation*) [8][9][10][11].

Decision tree (pohon keputusan) adalah suatu gambaran dalam bentuk pohon seperti diagram alir, yang dimana pada atribut pengujiannya mewakili setiap simpul bagian dalam, setiap cabang terdapat hasil pengujian dan sebuah label pada kelas mewakili simpul daun. Pada sebuah alur akan menjelajahi dari akar pohon sampai simpul terminal yang terdapat menyimpan suatu prediksi pada kelas untuk tuple (record)[12]. Suatu aturan klasifikasi dapat merubah pohon keputusan menjadi sangat mudah. Pohon keputusan ini juga menggunakan model prediktif yang bertujuan untuk menggambarkan sebuah pengamatan pada suatu hal untuk menuju kearah kesimpulan[13].

Algoritma C4.5 ini menjadi bagian gabungan dari algoritma decision tree. Pada langkah ini dilakukan dengan perancangan dari model sistem yang menggunakan algoritma C4.5 agar dapat membentuk sebuah pohon keputusan hingga dapat menghasilkan rule prediksi penerimaan beasiswa. Algoritma C4.5 adalah sebuah pohon keputusan yang dapat mengubah suatu fakta dengan menjadikan pohon keputusan atau rule dengan menguraikan aturan [14]. Algoritma C4.5 dianggap algoritma yang dapat membantu dalam melakukan sebuah klasifikasi data karena dalam karakter data yang diklasifikasikan dapat diperoleh dengan baik dan tepat, maka dapat memudahkan peneliti menggali informasi terhadap data tersebut[15].

Algoritma C4.5 yang dianggap sebagai yang sangat dalam membantu klasifikasi data karena karakter data yang akan diklasifikasikan terdapat dengan jelas, baik dalam gambaran struktur pohon keputusan (rule) ataupun ke dalam aturan *if then*, sehingga dapat memudahkan *user* dalam melakukan observasi informasi dengan data yang berkaitan. Pada penyeleksian atribut dalam penelitian ini sangat mempengaruhi dalam pengolahan algoritma C4.5 karena keputusan ini sangat bergantung pada atribut yang dipilih[16][17].

Algoritma juga dapat digunakan dengan metode klasifikasi untuk memprediksi profit pada perusahaan dengan memperhatikan nilai gain tertinggi dan menganalisa set pada data sampel hingga dapat membuat sebuah klasifikasi sehingga harus memiliki sebuah maksimum agar dapat mengatur sebuah kasus latihan dan kasus percobaan yang akurat [18][19]. Penelitian ini bertujuan untuk menyelesaikan permasalahan tersebut dengan menggunakan metode klasifikasi algoritma C4.5 dan dioperasikan menggunakan *software* Weka.

WEKA (*Waikato Environment for Knowledge*) adalah sebetuk software yang digunakan oleh algoritma *machine learning* agar dapat melakukan pemrosesan data mining. Penggalan informasi pada basis data dapat memudahkan pengguna. Ada beberapa fitur unggulan dari WEKA yaitu: *Classification, Regression, Clustering, Association Rules, Visualization*. Penelitian ini menggunakan salah satu fitur weka yaitu *Classification*. *Classification* adalah algoritma yang mendukung untuk proses klasifikasi pada objek untuk dilakukan implementasi secara langsung, dan ada beberapa representasi data yang menghasilkan nilai akurasi hingga tingkat kesalahan dari proses *Classification*.

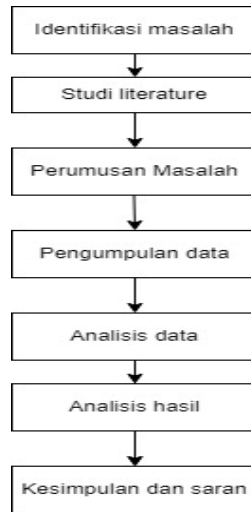
2. METODE PENELITIAN

Terdapat beberapa tahap dalam menyelesaikan sebuah permasalahan ini, tahapan penelitian ini secara garis besar yang dihasilkan ialah identifikasi masalah, studi literature, perumusan masalah, pengumpulan data, menganalisis data, menganalisis hasil, kesimpulan hingga saran. Penelitian ini dikerjakan sesuai dengan tahapan yang telah dibuat. Penetapan tahap penelitian ini agar bisa dikerjakan dengan sesuai yang dituju pada ketetapan-ketetapan standar dalam konsep ilmiah ini. Maka langkah-langkah penelitian ini dapat diuraikan sebagai berikut :

- a. Identifikasi masalah ialah sebuah tahap awal dalam mengatasi suatu masalah yang dimana sebuah tujuan dan keadaan tertentu kita bisa mengenali suatu masalah.
- b. *Studi literature* dilakukan dengan mencari sebuah karya ilmiah yang berhubungan dengan masalah yang diteliti, dan selain itu mengumpulkan data-data sebagai pendukung dalam teori penelitian ini.
- c. Setelah melakukan sebuah identifikasi masalah dan *studi literature* maka didapatkannya suatu perumusan masalah yang akan difokuskan dalam penelitian ini yang bertujuan memperjelas batasan yang tercakup pada suatu masalah.
- d. Pengumpulan data dibuat untuk mendapatkan sebuah informasi yang akan diperlukan dalam bentuk mencapai tujuan penelitian. Dalam tahap ini pengumpulan data dilakukan dengan data primer. Data primer adalah sebuah data yang diperoleh secara langsung dari objek yang diteliti seperti pada penelitian ini proses pengumpulan data dilakukan dengan cara meminta langsung kepada pihak sekolah yaitu SD 4 Pelangsan. Adapun data yang dikumpulkan pada penelitian mencakup dari variabel yang akan diolah yaitu:
 1. Jenis kelamin, variabel jenis kelamin ini berisikan Laki-laki dan perempuan.
 2. Pekerjaan orangtua, variabel pekerjaan orangtua ini berisikan pegawai swasta dan PNS.
 3. Pendapatan orangtua, variabel pendapatan seperti penghasilan tetap tiap bulannya.
 4. Nilai akhir, variabel nilai akhir berisikan akumulasi nilai per semesternya yang sudah ditetapkan.
 5. Status, variabel status adalah data yang akan berfungsi untuk menentukan hasil keputusan seperti, DITERIMA dan DITOLAK.
- e. Analisis data dilakukan dengan pengolahan data yang sudah lulus diseleksi, dan proses data pada penelitian ini sudah memasuki perkembangan dataset dan tahap pembersihan data (*data cleansing*).
- f. Analisis hasil dalam penelitian ini dilakukan dengan mengumpulkan data-data yang akan memasuki proses analisis data dan validasi data yang memberikan hasil pada penelitian tersebut. Pada penelitian ini algoritma C4.5 menggunakan *software* Weka 3.9.5. Implementasi algoritma C4.5 yaitu:
 1. Menetapkan atribut sebagai akar dan menghitung jumlah *information gain*.
 2. Menyusun *tree*(pohon) awal.
 3. Menyusun lanjutan *tree*.

4. Mengubah *tree* menjadi rule (pohon keputusan).
- g. Tahap terakhir yaitu tahap kesimpulan dan saran, yang akan membuktikan terjawab atau tidaknya pertanyaan pada penelitian ini.

Berikut adalah diagram alir yang akan menggambarkan secara singkat tahap penelitian ini dapat dilihat pada gambar 1.



Gambar1.Diagram Alir Penelitian

3. HASIL DAN PEMBAHASAN

Data yang akan dibuat beberapa kelompok untuk memasuki syarat dalam pengolahan data mining dengan menggunakan algoritma C4.5. algoritma C4.5 ini dimulai dari menentukan nilai *entropy* dari data dan dilanjut menentukan kriteria dengan menggunakan *gain*, *split info* dan *gain rasio* atau *index gini*, agar bisa menentukan akar dari pohon keputusan sehingga dapat memprediksi penerimaan beasiswa anak kelas 5 SD 4 Pelangsiyan. Berikut adalah data awal yang masih belum dijadikan klasifikasi atau data sebelum *pre-processing* dapat dilihat pada gambar 2.

NO	NAMA SISWA	JK	Pekerj.OT	Pend.OT	Kelas 1 - smst 1	Kelas 1 - smst 2	Kelas 2 - smst 1	Kelas 2 - smst 2	Kelas 3 - smst 1	Kelas 3 - smst 2	Kelas 4 - smst 1	Kelas 4 - smst 2	Status
1	Siswa 1	Laki-laki	Swasta	1.000.000 - 2.000.000	65	75	60	70	65	78	72	80	Ditolak
2	Siswa 2	Laki-laki	Swasta	1.000.000 - 2.000.000	65	70	70	60	60	82	69	78	Ditolak
3	Siswa 3	Perempuan	Swasta	3.000.000 - 4.000.000	65	77	75	70	70	69	78	75	Diterima
4	Siswa 4	Perempuan	Swasta	2.000.000 - 3.000.000	69	78	65	75	70	77	75	65	Ditolak
5	Siswa 5	Laki-laki	Swasta	1.000.000 - 2.000.000	60	65	80	68	65	60	65	80	Ditolak
6	Siswa 6	Laki-laki	PNS	2.000.000 - 3.000.000	70	80	85	87	70	73	89	74	Diterima
7	Siswa 7	Laki-laki	Swasta	1.000.000 - 2.000.000	78	90	73	89	68	65	69	70	Diterima
8	Siswa 8	Perempuan	Swasta	3.000.000 - 4.000.000	68	68	79	67	84	70	88	75	Ditolak
9	Siswa 9	Perempuan	PNS	1.000.000 - 2.000.000	68	80	76	65	83	78	78	60	Ditolak
10	Siswa 10	Laki-laki	PNS	3.000.000 - 4.000.000	84	78	82	65	80	73	86	90	Diterima
11	Siswa 11	Laki-laki	Swasta	1.000.000 - 2.000.000	83	78	69	70	78	68	69	61	Ditolak
12	Siswa 12	Laki-laki	Swasta	1.000.000 - 2.000.000	80	70	88	78	87	70	60	85	Diterima
13	Siswa 13	Laki-laki	Swasta	2.000.000 - 3.000.000	78	70	78	73	89	68	70	78	Diterima
14	Siswa 14	Laki-laki	Swasta	1.000.000 - 2.000.000	87	78	86	68	67	84	78	82	Diterima
15	Siswa 15	Laki-laki	Swasta	3.000.000 - 4.000.000	89	73	70	74	75	75	68	69	Diterima
16	Siswa 16	Perempuan	PNS	3.000.000 - 4.000.000	67	68	75	76	70	65	68	80	Ditolak
17	Siswa 17	Perempuan	Swasta	500.000 - 1.000.000	65	74	60	76	77	65	85	69	Ditolak
18	Siswa 18	Perempuan	PNS	3.000.000 - 4.000.000	65	76	63	82	90	69	73	60	Ditolak
19	Siswa 19	Perempuan	Swasta	3.000.000 - 4.000.000	60	88	61	69	77	60	79	70	Ditolak
20	Siswa 20	Perempuan	PNS	2.000.000 - 3.000.000	70	84	85	88	82	70	76	78	Diterima
21	Siswa 21	Perempuan	Swasta	1.000.000 - 2.000.000	76	86	75	78	65	90	82	68	Diterima
22	Siswa 22	Laki-laki	Swasta	2.000.000 - 3.000.000	70	62	87	86	80	78	89	90	Diterima

Gambar 2. Data sebelum *processing* Melakukan *Pre-Processing*

a. *Data Clearance*

Variabel ini ditentukan berdasarkan atribut pada data, sebagai berikut :

1. Mengelompokkan jenis kelamin, klasifikasi jenis kelamin dapat dilihat pada tabel 1.

Tabel 1. kelompok jenis kelamin

Jenis kelamin	Klasifikasi
Laki - laki	L
Perempuan	P

2. Mengelompokkan pendapatan orangtua, pengelompokkan pendapatan orang tua dari siswa ini diambil dengan rata-rata penghasilan orangtua perbulannya. Berikut klasifikasi pendapatan orangtua pada tabel 2.

Tabel 2. Penghasilan Orangtua

Pendapatan orangtua	Klasifikasi
3.000.000 – 4.000.000	A
2.000.000– 3.000.000	B
1.000.000 – 2.000.000	C
500.000 – 1.000.000	D

3. Mengelompokkan nilai per semester, pengelompokkan ini diambil dari kelas 5 dengan nilai akhir kelas 1 (semester1 dan semester2), kelas 2 (semester1 dan semester2), kelas 3 (semester1 dan semester2) dan kelas 4 (semester1 dan semester2). Berikut klasifikasi nilai per semester siswa pada tabel 3.

Tabel 3. Pengelompokkan Nilai

Nilai Akhir	Klasifikasi
< (dibawah nilai rata-rata yang ditetapkan)	Rendah
= (batas nilai yang ditetapkan)	Cukup
≥ (nilai diatas rata-rata)	Tuntas

4. Mengelompokkan pekerjaan orangtua, berikut klasifikasi pekerjaan orangtua pada tabel 4.

Tabel 4. Pengelompokkan pekerjaan orangtua

Pekerjaan orgtua	Klasifikasi
Petani, buruk pabrik, penjahit, dll	Swasta
PNS/TNI/POLRI	PNS

Adapun data yang diperoleh dapat dilihat pada gambar 3 sebagai berikut.

NO	NISN	NAMA	JK	Pekerj.OT	Pend.OT	smst 1	smst 2	smst 1	smst 2	smst 1	smst 2	smst 1	smst 2	Status
1	083932	Siswa 1	L	Swasta	C	65	≥ 70	< 68	70	< 72	≥ 69	< 75	≥ 78	Ditolak
2	135195	Siswa 2	L	Swasta	C	65	< 70	≥ 68	< 70	< 72	≥ 69	< 75	78	Ditolak
3	191066	Siswa 3	P	Swasta	A	65	≥ 70	≥ 68	70	< 72	69	≥ 75	< 78	Diterima
4	145839	Siswa 4	P	Swasta	B	≥ 65	≥ 70	< 68	≥ 70	< 72	≥ 69	75	< 78	Ditolak
5	067457	Siswa 5	L	Swasta	C	< 65	< 70	≥ 68	< 70	< 72	< 69	< 75	≥ 78	Ditolak
6	031355	Siswa 6	L	PNS	B	≥ 65	≥ 70	≥ 68	≥ 70	< 72	≥ 69	≥ 75	< 78	Diterima
7	049251	Siswa 7	L	Swasta	C	≥ 65	≥ 70	≥ 68	≥ 70	< 72	< 69	< 75	< 78	Diterima
8	195577	Siswa 8	P	Swasta	A	≥ 65	< 70	≥ 68	< 70	≥ 72	≥ 69	≥ 75	< 78	Ditolak
9	066840	Siswa 9	P	PNS	C	≥ 65	≥ 70	≥ 68	< 70	≥ 72	≥ 69	≥ 75	< 78	Ditolak
10	019594	Siswa 10	L	PNS	A	≥ 65	≥ 70	≥ 68	< 70	≥ 72	≥ 69	≥ 75	≥ 78	Diterima
11	131797	Siswa 11	L	Swasta	C	≥ 65	≥ 70	< 68	70	≥ 72	< 69	< 75	< 78	Ditolak
12	135777	Siswa 12	L	Swasta	C	≥ 65	70	≥ 68	≥ 70	≥ 72	≥ 69	< 75	≥ 78	Diterima
13	081574	Siswa 13	L	Swasta	B	≥ 65	70	≥ 68	≥ 70	≥ 72	< 69	< 75	≥ 78	Diterima
14	177003	Siswa 14	L	Swasta	C	≥ 65	≥ 70	≥ 68	< 70	< 72	≥ 69	≥ 75	≥ 78	Diterima
15	0986199	Siswa 15	L	Swasta	A	≥ 65	≥ 70	≥ 68	≥ 70	≥ 72	≥ 69	< 75	< 78	Diterima
16	021708	Siswa 16	P	PNS	A	≥ 65	< 70	≥ 68	≥ 70	< 72	< 69	< 75	≥ 78	Ditolak
17	142321	Siswa 17	P	Swasta	D	65	≥ 70	< 68	≥ 70	≥ 72	< 69	< 75	< 78	Ditolak
18	092714	Siswa 18	P	PNS	A	65	≥ 70	< 68	≥ 70	≥ 72	< 69	< 75	< 78	Ditolak
19	032201	Siswa 19	P	Swasta	A	< 65	≥ 70	< 68	< 70	≥ 72	< 69	≥ 75	< 78	Ditolak
...
...
41	821647	Siswa 41	L	Swasta	A	≥ 65	≥ 70	≥ 68	≥ 70	< 72	< 69	< 75	< 78	Ditolak

Gambar 3. Data pre-processing

b. Data integration

Dalam tahap ini dilakukan dengan proses penggabungan data. Maka hal ini dilakukan karena pengambilan data nilai dilakukan dengan nilai akhir per semester yang berbeda untuk setiap atributnya, oleh karena itu diperlukannya integrasi atau penggabungan data.

c. Data selection

Dalam tahapan ini melakukan proses seleksi data yang bermaksud dalam pengambilan data tersebut lebih signifikan yang dimana akan diproses analisis dalam metode *data mining*. Pada tahap ini juga membuat kumpulan data target dan suatu penentuan kumpulan data yang dimana akan dilakukan. Pada penelitian ini atribut data yang digunakan ialah jenis kelamin, pendapatan orangtua, pekerjaan orangtua, nilai akhir per semester dari kelas 1 sampai kelas 4. Data NISN dan Nama siswa dapat dihapuskan karena data tersebut termasuk identitas dan bukan merupakan data yang masuk kriteria dalam *data mining* penelitian ini. Data yang sudah dilakukan proses seleksi dapat dilihat pada gambar 4.

JK	Pekerj.OT	Pend.OT	as 1	smst 2	smst 1	smst 2	smst 1	smst 2	smst 1	smst 2	Status
L	Swasta	C	65	≥ 70	< 68	70	< 72	≥ 69	< 75	≥ 78	Ditolak
L	Swasta	C	65	< 70	≥ 68	< 70	< 72	≥ 69	< 75	78	Ditolak
P	Swasta	A	65	≥ 70	≥ 68	70	< 72	69	≥ 75	< 78	Diterima
P	Swasta	B	≥ 65	≥ 70	< 68	≥ 70	< 72	≥ 69	75	< 78	Ditolak
L	Swasta	C	< 65	< 70	≥ 68	< 70	< 72	< 69	< 75	≥ 78	Ditolak
L	PNS	B	≥ 65	≥ 70	≥ 68	≥ 70	< 72	≥ 69	≥ 75	< 78	Diterima
L	Swasta	C	≥ 65	≥ 70	≥ 68	≥ 70	< 72	< 69	< 75	< 78	Diterima
P	Swasta	A	≥ 65	< 70	≥ 68	< 70	≥ 72	≥ 69	≥ 75	< 78	Ditolak
P	PNS	C	≥ 65	≥ 70	≥ 68	< 70	≥ 72	≥ 69	≥ 75	< 78	Ditolak
L	PNS	A	≥ 65	≥ 70	≥ 68	< 70	≥ 72	≥ 69	≥ 75	≥ 78	Diterima
L	Swasta	C	≥ 65	≥ 70	< 68	70	≥ 72	< 69	< 75	< 78	Ditolak
L	Swasta	C	≥ 65	70	≥ 68	≥ 70	≥ 72	≥ 69	< 75	≥ 78	Diterima
L	Swasta	B	≥ 65	70	≥ 68	≥ 70	≥ 72	< 69	< 75	≥ 78	Diterima
L	Swasta	C	≥ 65	≥ 70	≥ 68	< 70	< 72	≥ 69	≥ 75	≥ 78	Diterima
L	Swasta	A	≥ 65	≥ 70	≥ 68	≥ 70	≥ 72	≥ 69	< 75	< 78	Diterima
P	PNS	A	≥ 65	< 70	≥ 68	≥ 70	< 72	< 69	< 75	≥ 78	Ditolak
P	Swasta	D	65	≥ 70	< 68	≥ 70	≥ 72	< 69	< 75	< 78	Ditolak
P	PNS	A	65	≥ 70	≥ 68	≥ 70	≥ 72	< 69	< 75	< 78	Ditolak
P	Swasta	A	< 65	≥ 70	< 68	< 70	≥ 72	< 69	≥ 75	< 78	Ditolak
...
...
L	Swasta	A	≥ 65	≥ 70	≥ 68	≥ 70	< 72	< 69	< 75	< 78	Ditolak

Gambar 4. Data setelah proses seleksi

d. *Data transformation*

Dalam tahap data transformasi ini diproses ke dalam bentuk yang sudah ditetapkan agar bisa dilakukan penggalan data. Data transformation ini dilakukan buat atribut yang banyak tidak memiliki nilai kesamaan atau beracak-acakan. Dalam penelitian ini atribut yang dipandang banyak tidak memiliki nilai kesamaan atau beracak-acakan ialah pada atribut nilai akhir. Nilai akhir pada data penelitian ini terdapat 8 kolom karena yang diambil per semester yaitu kelas 1 semester 1, kelas 1 semester 2, kelas 2 semester 1, kelas 2 semester 2, kelas 3 semester 1, kelas 3 semester 2, kelas 4 semester1 dan kelas 4 semester 2. Pada kedelapan kolom tersebut juga dilakukannya sebuah proses dikritisasi data dan diproses juga menggunakan software weka. *Data discretization* atau data dikritisasi ialah sebuah teknik yang dilakukan untuk memberi rentang atribut agar dapat mengurangi jumlah nilai pada atribut[11]. Rentang nilai yang dimaksud pada proses dikritisasi dikerjakan dengan pengaturan langsung pada software weka. Hasil dikritisasi data dapat dilihat pada gambar 5.

JK	pekerja	pend	Kelas 1	Kelas 1 - s	Kelas 2 - s	Kelas 2 - s	Kelas 3 - s	Kelas 3 - s	Kelas 4 - s	Kelas 4 - s	Status
L	Swasta	C	Cukup	Tuntas	Rendah	Cukup	Rendah	Tuntas	Rendah	Tuntas	Ditolak
L	Swasta	C	Cukup	Rendah	Tuntas	Rendah	Rendah	Tuntas	Rendah	Cukup	Ditolak
P	Swasta	A	Cukup	Tuntas	Tuntas	Cukup	Rendah	Cukup	Tuntas	Rendah	Diterima
P	Swasta	B	Tuntas	Tuntas	Rendah	Tuntas	Rendah	Tuntas	Cukup	Rendah	Ditolak
L	Swasta	C	Rendah	Rendah	Tuntas	Rendah	Rendah	Rendah	Cukup	Tuntas	Ditolak
L	PNS	B	Tuntas	Tuntas	Tuntas	Tuntas	Rendah	Tuntas	Tuntas	Rendah	Diterima
L	Swasta	C	Tuntas	Tuntas	Tuntas	Tuntas	Rendah	Rendah	Rendah	Rendah	Diterima
P	Swasta	A	Tuntas	Rendah	Tuntas	Rendah	Tuntas	Tuntas	Tuntas	Rendah	Ditolak
P	PNS	C	Tuntas	Tuntas	Tuntas	Rendah	Tuntas	Tuntas	Tuntas	Rendah	Ditolak
L	PNS	A	Tuntas	Tuntas	Tuntas	Rendah	Tuntas	Tuntas	Tuntas	Tuntas	Diterima
L	Swasta	C	Tuntas	Tuntas	Rendah	Cukup	Tuntas	Rendah	Rendah	Rendah	Ditolak
L	Swasta	C	Tuntas	Cukup	Tuntas	Tuntas	Tuntas	Tuntas	Rendah	Tuntas	Diterima
L	Swasta	B	Tuntas	Cukup	Tuntas	Tuntas	Tuntas	Rendah	Rendah	Tuntas	Diterima
L	Swasta	C	Tuntas	Tuntas	Tuntas	Rendah	Rendah	Tuntas	Tuntas	Tuntas	Diterima
L	Swasta	A	Tuntas	Tuntas	Tuntas	Tuntas	Tuntas	Tuntas	Rendah	Rendah	Diterima
P	PNS	A	Tuntas	Rendah	Tuntas	Tuntas	Rendah	Rendah	Rendah	Tuntas	Ditolak
P	Swasta	D	Cukup	Tuntas	Rendah	Tuntas	Tuntas	Rendah	Rendah	Rendah	Ditolak
P	PNS	A	Cukup	Tuntas	Rendah	Tuntas	Tuntas	Rendah	Rendah	Rendah	Ditolak
P	Swasta	A	Rendah	Tuntas	Rendah	Rendah	Tuntas	Rendah	Tuntas	Rendah	Ditolak
P	PNS	B	Tuntas	Tuntas	Tuntas	Tuntas	Tuntas	Tuntas	Rendah	Cukup	Diterima
...
...
L	PNS	A	Tuntas	Tuntas	Tuntas	Tuntas	Rendah	Tuntas	Tuntas	Rendah	Diterima
L	Swasta	A	Tuntas	Tuntas	Tuntas	Tuntas	Rendah	Rendah	Rendah	Rendah	Ditolak

Gambar 5. Hasil dikritisasi data

e. *Data mining*

Pada tahapan ini dilakukan proses penggalian data adalah suatu proses dalam pemakaian sebuah metode tertentu agar dapat menghasilkan pola-pola yang terdapat pada data. Proses perancangan data yang dilakukan dengan menggunakan test option *cross validation* dengan *folds* 10. Pemrosesan ini dilakukan dengan algoritma C4.5 dan jumlah data yang digunakan untuk memperoleh pola adalah sejumlah 41 data. Proses ini menggunakan *software* Weka.

Klasifikasi dan Pengujian dengan Algoritma C4.5

Pengolahan data dilakukan dengan melakukan proses perhitungan menggunakan algoritma C4.5. Tabel akurasi untuk proses perhitungan Algoritma C.45 dapat dilihat pada tabel 4.

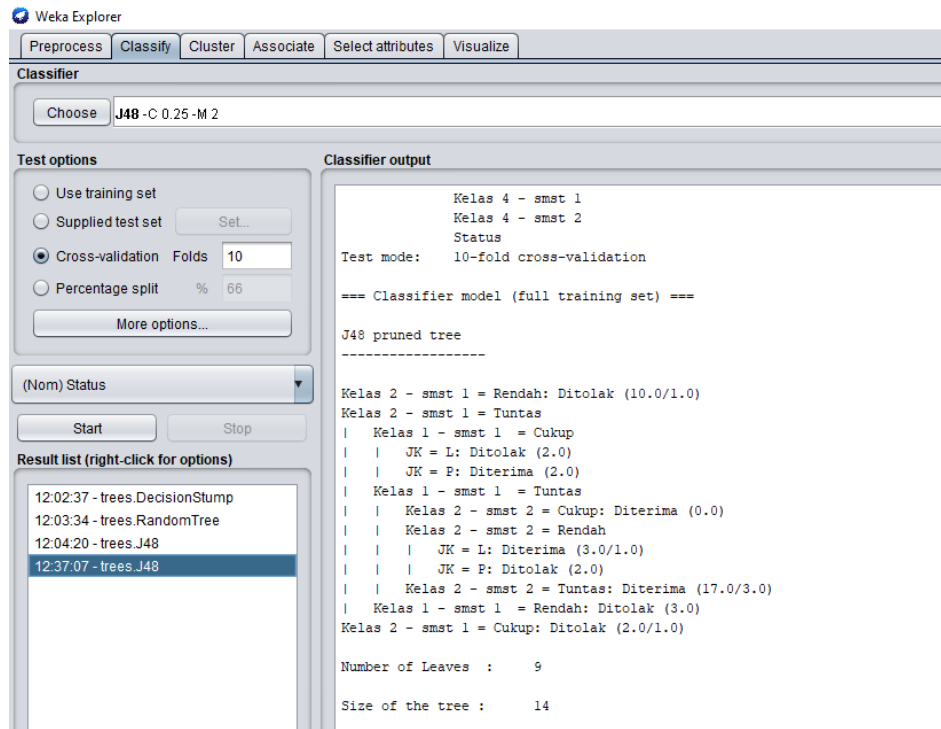
TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.810	0.100	0.895	0.810	0.850	Ditolak
0.900	0.190	0.818	0.900	0.857	Diterima
0.854	0.144	0.857	0.854	0.853	

Tabel 5. Tabel Akurasi C4.5

Berdasarkan tabel 4, diketahui jika True Positive Rate untuk class ditolak adalah 0.810 dengan False Positive Rate sebesar 0.100. Sedangkan untuk class diterima, True Positive Rate adalah 0.900 dan False Positive Rate adalah 0.190.

Classifier output yang berisi *running* data dengan diinputkan yang menjadi atribut-atribut hingga menjadi pohon keputusan. Contoh informasi yang ditampilkan adalah jumlah kasus dan keputusannya, berapa

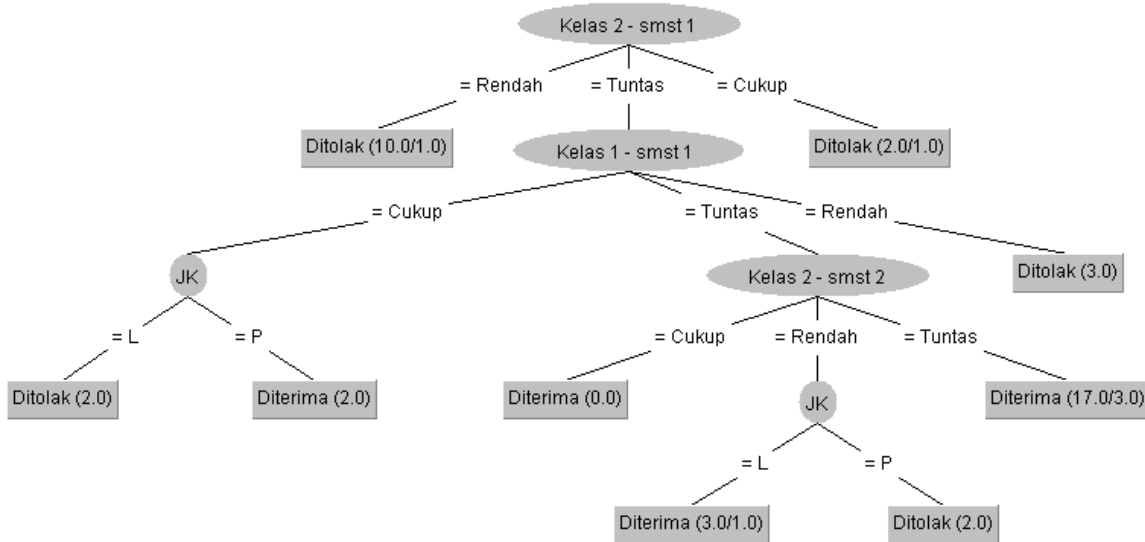
banyaknya cabang dari pohon keputusan tersebut. Berikut *Classifier output* pada software WEKA dapat dilihat pada gambar 6.



Gambar 6. *Classifier output*

f. Pohon Keputusan

Dalam pengolahan pohon keputusan yang wajib dilakukan ialah menghitung jumlah kasus, jumlah kasus keputusan “DITERIMA” jumlah kasus keputusan “DITOLAK” dan jumlah kasus yang dibagi berdasarkan jenis kelamin, pekerjaan orangtua, pendapatan orangtua, nilai akhir per semester dari kelas 1 sampai kelas 4 (semester 1 dan semester 2). Berikut ialah pohon keputusan atau Rule yang dihasilkan oleh pengolahan data menggunakan algoritma C4.5 menggunakan perangkat lunak WEKA dapat dilihat pada gambar 7.



Gambar 7. Pohon Keputusan

Jika dilihat berdasarkan hasil pohon keputusan prediksi penerimaan beasiswa pada gambar 7 ternyata atribut yang mempunyai pengaruh utama dalam mendapatkan prediksi adalah variabel nilai akhir kelas 2 semester 1 yang menempati simpul pada akar pohon keputusan tersebut, berikut adalah keterangan lebih jelas seperti dibawah ini:

- Jika kelas 2 semester 1 nilai rendah maka ditolak
- Jika kelas 2 semester 1, nilai tuntas maka memperhatikan,
 - Jika kelas 1 semester1 Nilai cukup maka perlu memperhatikan,
 - jenis kelamin laki-laki : ditolak
 - jenis kelamin perempuan : diterima
 - Jika kelas 1 semester1 Nilai tuntas maka perlu memperhatikan,
 - Jika Nilai akhir kelas 2 semester 2 maka,
 - Nilai cukup : diterima
 - Nilai rendah perlu memperhatikan,
 - Jenis kelamin laki-laki : diterima
 - Jenis kelamin perempuan : ditolak
 - Nilai tuntas : diterima
 - Jika kelas 1 semester1 Nilai rendah maka ditolak
 - Jika kelas 2 semester 1 nilai cukup maka ditolak

Bisa dilihat dari hasil pohon keputusan dan model yang berbentuk teks penjelasan diatas bahwa, jika calon siswa mendapatkan beasiswa dengan memiliki nilai tuntas maka calon siswa tersebut diprediksi diterima tanpa melihat jenis kelamin, pekerjaan orangtua dan pendapatan orangtua.

4. KESIMPULAN

Pengujian algoritma C4.5 dalam penelitian ini bisa diukur berdasarkan nilai accuracy, precision dan eror rate. Pengujian ini juga menggunakan information gain dengan 11 kriteria yaitu: pekerjaan orangtua, pendapatan orangtua, nilai akhir per semester kelas 1 sampai kelas 4 dan jenis kelamin hingga menghasilkan nilai akurasi sebesar 85,36 %, precision sebesar 85,7 %, eror rate sebesar 6,55 % lantaran

telah memiliki nilai ketepatan prediksi dengan benar tertinggi hingga persentase total dari data yang diprediksi dengan terendah. Maka hasil ini merupakan sebuah model yang pas jika dibandingkan dengan scenario yang lain untuk memprediksi penerimaan beasiswa di SD 4 Pelangian karena semua algoritma klasifikasi dapat membentuk model dengan akurasi tinggi namun *error rate* rendah. Pemilihan variabel (kriteria atau atribut) yang digunakan untuk memprediksi penerimaan beasiswa ini sangat mempengaruhi hasil rule yang dihasilkan. Algoritma C4.5 dengan menggunakan metode pohon keputusan dapat memberikan sebuah informasi berupa rule prediksi dengan menggambarkan proses yang bersangkutan dengan prediksi penerimaan beasiswa yang ditolak. Pengujian menggunakan perangkat lunak WEKA lebih efektif dan fleksibel sehingga dapat sangat membantu saat melakukan pengujian dalam penelitian.

5. SARAN

Pada penelitian selanjutnya dapat menggunakan data satu sekolah atau seluruh kelas agar penerimaan beasiswa dapat dilihat dengan secara detail hingga dapat meningkatkan nilai akurasi menjadi ke lebih baik dalam penerapan algoritma C4.5 untuk prediksi penerimaan beasiswa. Melakukan perbandingan dengan metode atau algoritma lain yang bisa prediksi penerima beasiswa.

DAFTAR PUSTAKA

- [1] Janam JH, Astuti NIP (2018) Penerapan Algoritma C4.5 untuk Penentuan Penerimaan Beasiswa (studi kasus: SDN Karawang Kulon III), Vol. 3 No: 1. ISSN: 2503-054X
- [2] F. Wiranata (2016), Perancangan Sistem Pendukung Keputusan Penerimaan Mahasiswa Beasiswa SMA Muhammadiyah 1 Gisting dengan Metode SAW (simple additive weighting), Vol, 4 No.1. ISSN: 2337-3032
- [3] Safitri, (2020), Penerapan Algoritma Naïve Bayes Untuk Penentuan Calon Penerimaan Beasiswa Pada Sd Negeri 6 Ketapang, JUISI, Vol, 06 No.1. ISSN: 2460-1306
- [4] Hijriana, MR., (2017), Penerapan Metode Decision Tree Algoritma C4.5 Untuk Seleksi Calon Penerima Beasiswa Tingkat Universitas, Vol. 3, no. 1.
- [5] Guterres (2015), Kelayakan Algoritma C4.5 Sebagai Pendukung Keputusan Dalam Pengajuan Penerima Beasiswa, pp. 142–147
- [6] Gunawan, R.P.K, R. R. Wigati,(2013), Pengembangan Sistem Penunjang Keputusan Penentuan Pemberian Beasiswa Tingkat Sekolah, JSM STMIK Mikroskil
- [7] Yunita, N., (2022) Metode Klasifikasi Data Mining Algoritma C4.5 Dan Part Untuk Prediksi Waktu Kelulusan Mahasiswa Di Universitas Darwan Ali, Jurnal Informasi Interaktif Vol. 7 No. 1. ISSN 2527-5240
- [8] H Susanto, S., (2014), Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan Dan Prestasi Masa Lalu, Jurnal Pendidikan Vokasi, Vol 4, Nomor 2. pp. 222-231
- [9] N. Hijriana ,M. Rasyidan,.(2017), Penerapan Metode Decision Tree Algoritma C4.5 Untuk Seleksi Calon Penerima Beasiswa Tingkat Universitas, Al Ulum Sains Dan Teknologi., vol. 3, no. 1
- [10] A. Putra, D. Y. Hardiyanti,. (2011), Penentuan Penerima Beasiswa Dengan Menggunakan Fuzzy Multiple Attribute Decision Making, JSI
- [11] Hera Wasiati, D.Wijayanti, (2014), Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naïve bayes (Studi Kasus: Di PT. Karyatama Mitra Sejati Yogyakarta), IJNS, vol 45
- [12] Rismayanti, (2018), Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus : Fakultas Teknik dan Komputer Universitas Harapan Medan), Jurnal Sistem Informasi, Vol. 02, No. 01
- [13] S.Riswanto, (2011) Penerapan Data Mining Untuk Memprediksi Fluktuasi Harga Saham

- Menggunakan Metode Classification Dengan Teknik Decision Tree,” Universitas Komputer Indonesia
- [14] R.Wajhillah, Ita Y. (2017), Penerapan algoritma C4.5 untuk prediksi penggunaan jenis kontrasepsi berbasis web, KJIK, Vol.04, No.2
- [15] R.Rafiska, S.D., G.W.N,(2018) , Analisis Rekam Medis Untuk Menentukan Pola Kelompok Penyakit Menggunakan Algoritma C4.5”, Jurnal RESTI, Vol. 2, No. 1, 391-396
- [16] Sembiring, M.A., (2016), Penerapan Metode Decission Tree Algoritma C4.5 untuk Memprediksi Hasil Belajar Mahasiswa Berdasarkan Riwayat Akademik, JTISI, Vol.3 No.01. ISSN: 2407-1811
- [17] B. Hermanto, A.S.N, (2017), Klasifikasi Nilai Kelayakan Calon Debitur Baru Menggunakan Decision Tree C4.5, IJCCS, Vol. 11, No. 01, pp. 43-54
- [18] E. Elisa,. (2018), Prediksi profit pada perusahaan dengan klasifikasi algoritma, KJIK, C4.5, Vol.5, No.02
- [19] Lorena, S,. (2014), Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik. Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)