

**KLASIFIKASI FAKTOR PENYEBAB SISWA PUTUS
SEKOLAH MENGGUNAKAN METODE *DECISION TREE ID3***

SKRIPSI

**OLEH
LARAZANI ARUM BARUNA PUTRI
NIM. 17610068**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022**

**KLASIFIKASI FAKTOR PENYEBAB SISWA PUTUS
SEKOLAH MENGGUNAKAN METODE *DECISION TREE ID3***

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
LARAZANI ARUM BARUNA PUTRI
NIM. 17610068**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022**

**KLASIFIKASI FAKTOR PENYEBAB SISWA PUTUS
SEKOLAH MENGGUNAKAN METODE *DECISION TREE ID3***

SKRIPSI

**Oleh
Larazani Arum Baruna Putri
NIM. 17610068**

Telah Diperiksa dan Disetujui untuk Diuji

Malang, 27 Desember 2022

Dosen Pembimbing I



Angga Dwi Mulyanto, M.Si
NIP. 19890813 201903 1 012

Dosen Pembimbing II



Erna Herawati, M.Pd
NIDT. 19760723 20180201 2 222

Mengetahui,
Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc

NIP. 19741129 200012 2 005

**KLASIFIKASI FAKTOR PENYEBAB SISWA PUTUS
SEKOLAH MENGGUNAKAN METODE *DECISION TREE* ID3**

SKRIPSI

**Oleh
Larazani Arum Baruna Putri
NIM. 17610068**

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

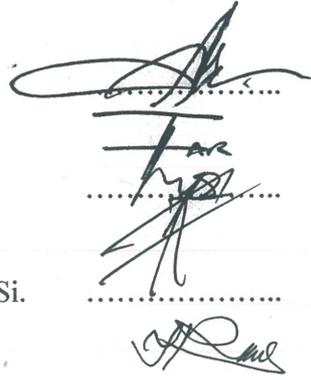
Tanggal, 27 Desember 2022

Ketua Penguji : Abdul Aziz, M.Si.

Anggota Penguji 1 : Fachrur Rozi, M.Si.

Anggota Penguji 2 : Angga Dwi Mulyanto, M.Si.

Anggota Penguji 3 : Erna Herawati, M.Pd



Mengetahui,
Ketua Program Studi Matematika




Dr. Elly Susanti, M.Sc.
NIP. 119741129 200012 2 005

PERNYATAAN KEASLIAN TULIISAN

Nama : Larazani Arum Baruna Putri
NIM : 17610068
Program Studi : Matematika
Fakultas : Sains dan Teknologi
Judul Skripsi : Klasifikasi Faktor penyebab Siswa Putus Sekolah
Menggunakan Metode *Decision Tree ID3*.

Menyatakan dengan sebenar-benarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri. Bukan merupakan pengambilan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya, kecuali dengan mencantumkan sumber pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 27 Desember 2022
Yang membuat pernyataan,



Larazani Arum Baruna Putri
NIM.17610068

MOTO

“Kita tidak bisa melawan ombak, tetapi kita bisa berenang mengikuti ombak”

PERSEMBAHAN

Skripsi ini penulis persembahkan untuk:

Kedua orang tua peneliti Rizani dan Tusnilah, yang telah memberikan dukungan berupa tenaga, pikiran dan finansial dengan sepenuh hati, serta tidak lupa Agha Zaki Athaya Samudra dan Azhar Bahari Zain selaku adek peneliti yang selalu memberi semangat dan menjadi penghibur dikala peneliti merasa jenuh. Tidak lupa Dhiya Khairunnisa selaku saudara sepupu peneliti yang selalu ada dikala susah maupun senang serta membantu dalam penulisan skripsi, beserta sahabat peneliti Arum Cahya yang bersedia meminjamkan laptop untuk keperluan skripsi, Mita yang selalu menemani dan mendukung saat penyusunan skripsi, Arinda dan Zaidina yang selalu menjadi tempat berkeluh kesah ketika peneliti merasa lelah, serta teman-teman yang bersedia membantu dalam penyusunan skripsi ini. Tanpa kalian semua skripsi ini tidak dapat berjalan dengan lancar.

KATA PENGANTAR

Pertama-tama penulis panjatkan puji syukur atas kehadiran Allah SWT karena telah melimpahkan rahmat, taufik dan hidayah-Nya sehingga penulis mampu menyelesaikan penulisan proposal skripsi yang berjudul “Klasifikasi Faktor Penyebab Siswa Putus Sekolah Dengan Metode *Decision Tree ID3*” yang merupakan bagian dari syarat untuk memperoleh gelar sarjana dari Universitas Islam Negeri Maulana Malik Ibrahim Malang Fakultas Sains dan Teknologi Jurusan Matematika.

Untuk kepentingan penulisan makalah ini, peneliti mengucapkan terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim.
2. Dr. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.
3. Dr. Elly Susanti, M.Sc., selaku ketua Program Studi Matematika, Universitas Islam Negeri Maulana Malik Ibrahim.
4. Bapak Angga Dwi Mulyanto, M.Si., selaku dosen pembimbing I yang dengan sabar memberikan bimbingan, arahan serta dukungan dan waktunya untuk membantu penulis menyelesaikan skripsi ini.
5. Ibu Erna Herawati, M.Pd., selaku dosen pembimbing II yang dengan sabar membimbing serta mengarahkan penulis dalam menyusun skripsi ini.
6. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim.
7. Kedua orang tua penulis yang senantiasa membimbing dan mengarahkan penulis untuk selalu berjalan di jalan yang benar.
8. Kedua adik penulis yang selalu mendukung saya.
9. Keluarga besar penulis yang senantiasa memotivasi saya untuk tidak menyerah dan terus berjuang menyelesaikan perkuliahan saya.
10. Sahabat-sahabat penulis yang bersedia meluangkan waktunya untuk mendengarkan keluh kesah saya saat menjalani skripsi ini.

11. Seluruh mahasiswa angkatan 2017 yang bersedia untuk berdiskusi dengan penulis.

Sekian ucapan terimakasih kepada pihak-pihak terkait yang berkontribusi besar pada penulisan skripsi ini. Penulis berharap dengan ditulisnya skripsi ini mampu mendatangkankan kebaikan untuk penulis dan orang banyak. Aamiin.

Malang, 2022

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGUJIAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
ABSTRAK	xv
ABSTRACT	xvi
مستخلص البحث	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan Penelitian	6
1.4 Manfaat Penelitian	6
1.5 Batasan Masalah	7
1.6 Definisi Istilah.....	7
BAB II KAJIAN PUSTAKA	9
2.1 Teori Pendukung	9
2.1.1 <i>Data Mining</i>	9
2.1.2 Klasifikasi	11
2.1.3 <i>Decision Tree</i>	12
2.1.4 <i>Algoritma Iterative Dichotomiser (ID3)</i>	14
2.1.5 <i>Cross Validation</i>	17
2.1.6 <i>Confusion Matrix</i>	18
2.1.7 Putus Sekolah	19
2.2 Pentingnya Pendidikan Dalam Al-Qur'an	24
2.3 Kajian Putus Sekolah dengan Metode Klasifikasi <i>Decision Tree</i>	27
BAB III METODE PENELITIAN	29
3.1 Jenis Penelitian.....	29
3.2 Data dan Sumber Data	29
3.3 Teknik Pengumpulan Data	30
3.4 Variabel Penelitian	30
3.5 Teknik Analisis Data.....	31
3.3.1 Analisis Deskriptif Data	31
3.3.2 <i>Cross Validation</i>	31
3.3.3 <i>Data Selection</i>	31
3.3.4 <i>Preprocessing (Data Cleaning)</i>	32
3.3.5 Transformasi Data	32
3.3.6 <i>Classifier</i>	32

3.3.7	Evaluasi	32
BAB IV	HASIL DAN PEMBAHASAN	33
4.1	Analisis Deskriptif Data	33
4.2	<i>Cross Validation</i>	39
4.3	<i>Data Selection</i>	41
4.4	<i>Preprocessing</i>	44
4.5	Transformasi Data	45
4.6	<i>Classifier</i>	45
4.7	Evaluasi	53
4.8	Kajian Keislaman dengan Hasil Penelitian	54
BAB V	PENUTUP.....	56
5.1	Kesimpulan.....	56
5.2	Saran Untuk Penelitian Lanjutan.....	57
DAFTAR PUSTAKA	58
LAMPIRAN	61
RIWAYAT HIDUP		

DAFTAR TABEL

Tabel 2.1 <i>Rule of thumb</i> Nilai <i>Accuracy</i>	19
Tabel 3.1 Tabel Atribut <i>Decision Tree</i>	30
Tabel 4.1 Nilai <i>Accuracy</i> dari Kombinasi Data <i>Testing</i> dan Data <i>Training</i>	40
Tabel 4.2 Tabel Proses <i>Data Selection</i>	41
Tabel 4.3 Tabel Hasil Perhitungan Menentukan <i>Root Node</i>	46
Tabel 4.4 Tabel Hasil Perhitungan Menentukan <i>Internal Node</i>	49
Tabel 4.5 Tabel Hasil Perhitungan Menentukan <i>Leaf Node</i>	50
Tabel 4.6 Tabel Hasil Perhitungan <i>Confusion Matrix</i>	53

DAFTAR GAMBAR

Gambar 2.1 Proses Abstrak <i>Data Mining</i>	10
Gambar 2.2 <i>Output Decision Tree</i>	13
Gambar 2.3 Proses Seleksi Data <i>K-Fold Cross</i>	17
Gambar 2.4 Tabel Perhitungan <i>Confusion Matrix</i>	18
Gambar 4.1 Grafik Jumlah Siswa Berdasarkan Status Siswa	33
Gambar 4.2 Grafik jumlah Siswa Berdasarkan Jenis Kelamin	34
Gambar 4.3 Prosentase Atribut Peendidikan Ayah.....	35
Gambar 4.4 Prosentase Atribut Pekerjaan Ayah.....	36
Gambar 4.5 Prosentase Atribut Pendidikan Ibu	36
Gambar 4.6 Prosentase Atribut Pekerjaan Ibu	37
Gambar 4.7 Prosentase Atribut Jumlah Anggota Keluarga	38
Gambar 4.8 Prosentase Atribut Pendapatan Keluarga	38
Gambar 4.9 Potongan <i>Output Decision Tree</i> dengan <i>RapidMiner</i>	51

DAFTAR LAMPIRAN

- Lampiran 1** Potongan Data Asli Siswa
- Lampiran 2** Potongan Data Putus Sekolah Setelah di-*Preprocessing*
- Lampiran 3** Proses *Cross Validation* pada *Rapidminer*
- Lampiran 4** Kolom *Number of Fold*
- Lampiran 5** Kombinasi Data *Testing* dan Data *Training*
- Lampiran 6** Tabel Hasil Perhitungan Untuk menentukan *Internal Node* dan *Leaf Node*
- Lampiran 7** *Output Decision Tree*
- Lampiran 8** *Output Confusion Matrix RapidMiner*

ABSTRAK

Putri, Larazani Arum Baruna .2022. **Klasifikasi Faktor Penyebab Siswa Putus Sekolah Menggunakan Metode *Decision Tree ID3***. Skripsi Program Studi Matematika. Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing (I) Angga Dwi Mulyanto, M.Si (II) Erna Herawati, M.Pd.

Kata Kunci: *Decision Tree*, ID3, Klasifikasi, Putus Sekolah.

Pendidikan menjadi faktor paling penting dalam mencetak generasi bangsa yang berkualitas. Oleh sebab itu pemerintah mengeluarkan kebijakan wajib belajar 12 tahun. Namun, kesadaran masyarakat Indonesia mengenai hal ini masih sangat kurang, terlebih bagi mereka yang tinggal di daerah pedalaman. Pada salah satu sekolah menengah pertama di Desa Panangan Kecamatan Gapura Kabupaten Sumenep masih ditemukan siswa-siswi yang memilih untuk putus sekolah setiap tahunnya. Keputusan ini diambil berdasarkan beberapa faktor yang menjadi penyebab siswa memilih jalan tersebut. Faktor yang menjadi penyebab siswa putus sekolah menarik untuk diteliti lebih dalam, agar dimasa yang akan datang hal ini tidak terjadi lagi. Terdapat 329 data yang diteliti dalam penelitian ini dimana 6 diantaranya memutuskan untuk putus sekolah. Data tersebut merupakan data siswa-siswi dari sekolah yang bersangkutan selama 3 tahun terakhir. Pada prosesnya peneliti akan mengklasifikasikan faktor penyebab siswa putus sekolah yang nantinya faktor-faktor tersebut akan disusun menjadi *decision tree* yang dibangun dengan algoritma *ID3*. Nilai akurasi dari model klasifikasi adalah 98,48%. Berdasarkan hasil tersebut Algoritma *ID3* dapat mengoptimalkan *decision tree* dalam mengklasifikasikan faktor penyebab siswa putus sekolah.

ABSTRACT

Putri, Larazani Arum Baruna. 2022. **Classification of Factors Causing Students to Drop Out of School Using the ID3 Decision Tree Method**. Thesis for the Mathematics Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisor (I) Angga Dwi Mulyanto, M.Si (II) Erna Herawati, M.Pd.

Keywords: Decision Tree, ID3, Classification, Drop Out.

Education is the most important factor in producing a quality generation of the nation. Therefore the government issued a 12 year compulsory education policy. However, the awareness of the Indonesian people regarding this matter is still very lacking, especially for those who live in rural areas. In one of the junior high schools in Panangan Village, Gapura District, Sumenep Regency, there are still students who choose to drop out of school every year. This decision was taken based on several factors that caused students to choose that path. The factors that cause students to drop out of school are interesting to study more deeply, so that in the future this will not happen again. There are 329 data that be examined in this study where 6 of them decided to drop out of school. The data is data on students from the school concerned for the last three years. In the process, researchers will classify the factors that cause students to drop out of school, which later these factors will be arranged into a decision tree built with the ID3 algorithm. The accuracy value of the classification model is 98.48%. Based on these results, the ID3 Algorithm can optimize the decision tree in classifying the factors that cause students to drop out of school.

مستخلص البحث

لارازاني أروم بارونا بوتري. ٢٠٢٢. تصنيف العوامل السببية على توقف الطلاب الدراسة باستخدام طريقة *Decision Tree ID3*. البحث الجامعي لقسم الرياضيات. كلية العلوم والتكنولوجيا. جامعة مولانا مالك إبراهيم الإسلامية الحكومية بمالانج. المشرف: (١) أنجا دوي موليانطا الماجستير، (٢) إيرنا هيراواتي الماجستير.

الكلمات المفتاحية: *Decision Tree*، ID3، التصنيف، توقف الدراسة.

التربية هو أهم عامل في إنتاج لوجودة الأمة. أصدرت الحكومة سياسة التعليم الإلزامي لمدة ١٢ عامًا. فإن ذلك لا يزال وعي الشعب الإندونيسي فيما يتعلق بهذه المسألة مفقودا للغاية، خاصة بالنسبة لأولئك الذين يعيشون في المناطق الريفية. في إحدى المدارس المتوسطة في قرية بانانجان، منطقة غابورا، منطقة فرعية سومينيب، لا يزال الطلاب الذين يختارون ترك المدرسة كل عام. تم اتخاذ هذا السبب بناءً على عدة عوامل تسبب في اختيار الطالب للمتقّس. العوامل التي تسبب على توقف الطلاب من المدرسة مثيرة للاهتمام للدراسة بشكل أعمق، بحيث لا يحدث هذا مرة أخرى في المستقبل. تمت ٣٢٩ بيانات في هذه الدراسة، قرر ٦ منها توقف للدراسة. البيانات هي بيانات عن طلاب من المدرسة المعنية لآخر ٣ سنوات. في هذه العملية، سيقوم الباحثة بتصنيف العوامل السببية على توقف الطلاب الدراسة، والتي سيتم تجميع هذه العوامل لاحقًا في *Decision Tree* باستخدام خوارزمية ID3. قيمة دقة نموذج التصنيف هي ٩٨,٤٨٪. بناءً على هذه النتائج، الخوارزمية ID3 تستطيع لترقية *Decision Tree* في تصنيف العوامل السببية على توقف الطلاب الدراسة.

BAB I PENDAHULUAN

1.1 Latar Belakang

Algoritma *ID3* merupakan salah satu algoritma klasifikasi *data mining* yang mampu menghasilkan model prediktif. Algoritma ini dapat digunakan sebagai pembangun *decision tree*. Dimana *decision tree* merupakan algoritma *machine learning* yang memiliki seperangkat aturan untuk menemukan model prediksi, berdasarkan struktur pohon (Hafizan & Nadia, 2020). Dibandingkan dengan algoritma lainnya dalam memprediksi jumlah data yang besar *decision tree* memiliki tingkat akurasi yang lebih baik. Algoritma ini umumnya digunakan untuk mengklasifikasikan *data mining* agar dapat memprediksi tingkah laku dimasa yang akan datang (Jollyta, dkk., 2020). Efektivitas sistem prediksi dapat membantu orang untuk mengetahui resiko dimasa yang akan datang berdasarkan pengelompokkan tingkah laku dimasa sekarang.

Klasifikasi sendiri merupakan salah satu metode pada *data mining* yang bertujuan untuk mengelompokkan data yang relatif berukuran besar untuk dapat menemukan sebuah informasi baru. Sedangkan *data mining* sendiri merupakan sebuah praktik penyelidikan sampel berasal dari basis data besar yang dikumpulkan dari berbagai sumber serta, memiliki kemungkinan berguna dalam membentuk pola penyelesaian. *Data mining* dianggap sebagai jantung dari upaya analitik di berbagai industri dan ilmu yang berbeda seperti komunikasi, manufaktur, pendidikan, media sosial, asuransi, dan ritel (Safii, 2018). Dengan berkembangnya data mining di bidang bioinformatika, *decision tree* menjadi populer.

Di era globalisasi seperti ini pendidikan merupakan salah satu sektor terpenting bagi sebuah negara dalam usaha mengembangkan kualitas Sumber Daya Manusia (SDM). Berdasarkan UU SISDIKNAS (2011:03) menegaskan bahwa fungsi pendidikan adalah sebagai media untuk membentuk kepribadian yang bermartabat dalam mencerdaskan serta menggali potensi dari masing-masing individu, agar dapat menjadikan generasi bangsa yang taat kepada Tuhan Yang Maha Esa, berperilaku baik, bermartabat serta berwawasan tinggi (Dewi, dkk., 2014). Namun, tidak semua masyarakat memahami pentingnya menuntut ilmu, tidak sedikit anak bangsa yang memutuskan putus sekolah dilandasi berbagai macam faktor.

Putus sekolah merupakan masalah pendidikan yang harus ditangani sampai ke-akarnya. Khususnya pada jenjang SMK atau sederajat yang merupakan tahap akhir dari skema pemerintah, wajib belajar 12 tahun, sebab setelah itu barulah seseorang dianggap layak memasuki dunia kerja. Beberapa faktor penyebabnya antara lain faktor ekonomi, kondisi geografis, dan keinginan siswa itu sendiri. Putri (2017) dalam penelitiannya menyebutkan bahwa faktor internal menjadi faktor paling dominan penyebab siswa putus sekolah. Rendahnya kesadaran siswa mengenai semangat belajar dan pentingnya pendidikan menjadi faktor utama. Namun beberapa faktor eksternal juga mempengaruhi siswa putus sekolah (Nigtyas, dkk., 2019). Oleh sebab itu dalam penelitian ini, peneliti berencana untuk mengklasifikasikan faktor eksternal yang menjadi penyebab seorang siswa memutuskan untuk putus sekolah, dengan menggunakan metode *decision tree* yang dibangun dengan algoritma *ID3*.

Salah satu penelitian yang pernah membahas tentang penggunaan metode *decision tree* adalah penelitian yang dilakukan Medina, dkk (2020). Pada penelitiannya percobaan dilakukan terhadap 500 dataset, yang didapat dari responden mahasiswa sarjana dari 5 kampus swasta berbeda di Peru. Dari 500 data tersebut 45% diantaranya adalah perempuan (225 orang) sedangkan sisanya laki-laki (275 orang), kemudian data divalidasi menggunakan metode implementasi model *cross-validation*. Pada penelitian ini *decision tree* dioptimalkan dengan algoritma *adaptive boosting*, sehingga kinerja dari *decision tree* memiliki akurasi sebesar 61,92% secara keseluruhan (Medina, dkk., 2020).

Algoritma *ID3* dinilai mampu mengoptimalkan *decision tree* dalam penelitian Aritonang (2021). Pada penelitiannya terdapat 5 jenis pupuk yang akan diuji, untuk dihitung nilai *entropy* dan *gain*. Dalam penelitian tersebut, algoritma yang digunakan adalah Algoritma *ID3*, dan hasil dari penelitian tersebut berupa *decision tree*. Pada penelitian tersebut dapat disimpulkan bahwa *decision tree* dapat dibentuk dengan menghitung nilai *entropy* dan *gain* menggunakan Algoritma *ID3*.

Klasifikasi merupakan sebuah metode pengelompokkan data untuk dapat memudahkan kita dalam mengamati suatu hal. Dalam penelitian ini pengelompokkan data tersebut dilakukan dengan tujuan untuk dapat memprediksi tingkah laku dimasa yang akan datang dengan menggunakan algoritma *decision tree* yang dibangun oleh algoritma *ID3*. Prinsipnya algoritma *decision tree* akan mengelompokkan dan menyusun atribut pada dataset menjadi atribut label (*Root node*), atribut percabangan (*Internal node*), dan atribut hasil (*Leaf node*). Berdasarkan penyusunan tersebut nantinya, kita dapat melihat atribut apa saja

yang saling berhubungan sehingga dapat memprediksikan tingkah laku dimasa yang akan datang.

Seperti metode klasifikasi *data mining* dengan algoritma *decision tree* dalam sebuah Hadist juga disebutkan peramalan kejadian yang terjadi di masa kini. Prediksi di masa depan juga disebutkan oleh Rasulullah SAW, sebagaimana dalam hadits yang berbunyi sebagaimana berikut:

عَلِّمُوا أَوْلَادَكُمْ فَإِنَّهُمْ سَيَعِيشُ فِي زَمَانِهِمْ غَيْرَ زَمَانِكُمْ فَإِنَّهُمْ خُلِقُوا لِزَمَانِهِمْ وَنَحْنُ خُلِقْنَا لِزَمَانِنَا (رواه عمر ابن الخطاب)

Artinya: “*Ajarilah anak-anakmu sesuai dengan zamannya, karena mereka hidup di zaman mereka bukan pada zamanmu. Sesungguhnya mereka diciptakan untuk zamannya, sedangkan kalian diciptakan untuk zaman kalian. (HR. Umar Bin Khattab)*”

Dari hadits tersebut sangat jelas bahwa waktu adalah dinamis. Waktu selalu berkembang. Bahkan Rasulullah S.A.W. sudah menjelaskan bahwa masa orang tua akan berbeda dengan masa anaknya. Oleh karena itu orang tua harus bisa mendidik anaknya sebagaimana zaman anaknya, bukan zaman orang tua. Dari penjelasan ini sudah jelas bahwa Rasulullah pun sudah melakukan prediksi masa depan mengenai karakter waktu maupun zaman yang bersifat dinamis dan selalu berkembang (Ajahari, 2018). Sama halnya dengan pengklasifikasian *data mining* yang diharapkan mampu memprediksi tingkah laku dimasa yang akan datang berdasarkan kejadian dimasa kini.

Penelitian mengenai sistem prediksi dengan menggunakan algoritma *decision tree* sebelumnya pernah dilakukan oleh Kurniawan & Rosadi (2017). Dalam penelitian tersebut percobaan dilakukan terhadap 40 dataset siswa pada SMAN Barabi, dimana 13 diantaranya memutuskan untuk putus sekolah. Akurasi dari kinerja algoritma *decision tree* pada penelitian ini mampu mencapai angka

72,5%, dan setelah dioptimalkan dengan menggunakan PSO akurasinya meningkat menjadi 85%. Sehingga dapat disimpulkan bahwa dengan menerapkan algoritma *ID3* mampu membentuk *decision tree*. Berdasarkan nilai tersebut dapat disimpulkan bahwa hasil klasifikasinya cukup baik sehingga pola ini dapat menjadi tolak ukur untuk mendeteksi faktor yang mempengaruhi siswa putus sekolah. Sehingga peneliti tertarik menggunakan algoritma *ID3* untuk membangun *decision tree* dalam mengklasifikasikan faktor apa saja yang menjadi alasan seorang siswa memutuskan untuk putus sekolah.

Dengan mengklasifikasikan faktor-faktor eksternal yang dinilai berkemungkinan menjadi alasan seorang siswa putus sekolah, khususnya siswa kelas IX pada salah satu sekolah menengah di Sumenep, diharapkan mampu meminimalisir resiko putus sekolah disekolah tersebut kedepannya. Peneliti berharap dengan dilakukannya penelitian ini dapat mengetahui faktor apa yang paling mempengaruhi seorang siswa dalam mengambil keputusan untuk melanjutkan atau putus sekolah.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas adapun rumusan masalah pada penelitian ini yaitu:

1. Bagaimana hasil kinerja algoritma *ID3* dalam mengoptimalkan *decision tree* untuk mengklasifikasikan faktor yang melatarbelakangi siswa putus sekolah?

2. Atribut apakah yang paling berpengaruh terhadap siswa dalam mengambil keputusan untuk putus ataupun lanjut sekolah setelah diklasifikasikan dengan metode *decision tree* yang dibangun oleh algoritma *ID3*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas maka, penelitian ini bertujuan untuk:

1. Mengetahui hasil kinerja algoritma *ID3* dalam mengoptimalkan *decision tree* untuk mengklasifikasikan faktor yang melatar belakangi siswa putus sekolah.
2. Mengetahui atribut yang paling berpengaruh terhadap siswa dalam mengambil keputusan untuk putus ataupun lanjut sekolah setelah diklasifikasikan dengan metode *decision tree* yang dibangun oleh algoritma *ID3*.

1.4 Manfaat Penelitian

Adapun manfaat yang didapat dalam penelitian ini sebagai berikut:

1. Mampu mempraktikkan secara langsung ilmu statistika yang diperoleh dalam metode pengklasifikasian data *mining* menggunakan algoritma *decesion tree* yang diaplikasikan dengan algoritma *Iterative Dichotomiser 3 (ID3)*.
2. Dapat mengetahui faktor apa saja yang mempengaruhi siswa dalam mengambil keputusan untuk putus sekolah.
3. Dapat meminimalisir siswa putus sekolah.

1.5 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan hanya berfokus pada siswa kelas IX 3 tahun terakhir (2019-2022) dari salah satu sekolah menengah pertama di Desa Panangan, Kecamatan Gapura, Kabupaten Sumenep Jawa Timur.
2. Penilaian hasil kinerja dari algoritma *ID3* dalam mengoptimalkan *decision tree* untuk mengklasifikasikan faktor penyebab putus sekolah ditentukan berdasarkan nilai *accuracy*, *precision*, dan *recall*.

1.6 Definisi Istilah

Adapun definisi dari istilah-istilah yang terdapat dalam skripsi ini adalah sebagai berikut:

<i>Accuracy</i>	:	Tingkat kedekatan pengukuran kuantitas terhadap nilai yang sebenarnya.
<i>Entropy</i>	:	Ukuran teoretis informasi yang mencirikan ketidakmurnian, serta homogenitas kumpulan dari suatu data.
<i>Internal node</i>	:	Titik percabangan dalam <i>decision tree</i> .
<i>Leaf node</i>	:	Titik akhir dalam <i>decision tree</i> .
Putus Sekolah	:	Kondisi anak yang tidak berkesempatan untuk menyelesaikan pendidikan hingga tidak memperoleh keterangan tamat belajar atau ijazah yang disebabkan oleh faktor-faktor tertentu.
<i>Precision</i>	:	Kecocokan antara bagian data yang diambil dengan

informasi yang dibutuhkan.

Recall : Tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

Root node : Titik awal dalam *decision tree*.

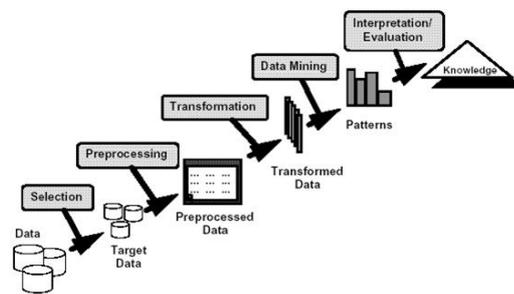
BAB II KAJIAN PUSTAKA

2.1 Teori Pendukung

2.1.1 *Data Mining*

Istilah *data mining* pertama kali muncul pada tahun 1990 (Han, dkk., 2011). Secara umum data adalah sebuah fakta-fakta atau entitas yang tidak berarti apapun, sehingga sering terabaikan. Sedangkan *mining* adalah sebuah proses penambangan data yang dilakukan untuk menghasilkan *output*. Sehingga dapat disimpulkan bahwa, data mining merupakan sebuah proses untuk menambang data, dengan tujuan untuk menghasilkan *output* (keluaran) berupa informasi, pola ataupun pengetahuan baru (Nofriansyah & Nurcahyo, 2015).

John Naisbitt menyatakan bahwa kita sering tenggelam dalam data, namun kita masih kekurangan informasi dan pengetahuan. Dengan berkembangnya ilmu pengetahuan, ditemukanlah teknik pengolahan data yang sekarang kita kenal dengan data mining. Sedangkan menurut Ian H Witten, *data mining* adalah sebuah proses mengekstrak suatu data besar yang terabaikan menjadi sebuah pola, informasi ataupun pengetahuan baru. Dalam prosesnya kita akan sering mendengar istilah *attribut*, maupun *class/label/target*. *Attribut* sendiri merupakan sebuah deskripsi data yang berisi field-field yang saling berkaitan satu sama lain. Sedangkan *class/label/target* bisa dikatakan sebuah *Attribut* keputusan (Nofriansyah & Nurcahyo, 2015).



Gambar 2.1 Proses Abstrak *Data Mining*

Berdasarkan gambar 2.1 dapat kita lihat bahwa dalam prosesnya data akan melalui beberapa tahap, sebagai berikut (Jollyta, dkk., 2020):

1. *Data Selection*

Data selection adalah kegiatan menyeleksi atribut sesuai dengan kriteria yang telah ditentukan. Pada tahap ini atribut yang dipilih adalah atribut yang dapat membangun model *data mining*.

2. *Data Preprocessing*

Preprocessing adalah tahap data mining untuk membersihkan data yang dianggap dapat menghambat jalannya penelitian.

3. *Data Transformation*

Data transformation adalah proses mentransformasi data agar data yang dapat ditelusuri menggunakan metode yang digunakan. Dengan kata lain data akan dipetakan secara kompleks.

4. *Data Mining*

Tahap *data mining* adalah tahap dimana data akan diproses sesuai dengan model *mining* yang telah ditentukan.

5. *Evaluation*

Tahap evaluasi dilakukan untuk menginterpretasi pola data, agar dapat digunakan sebagai pendukung pengambilan keputusan.

Ada sebuah istilah yang memiliki arti serupa dengan *data mining*, yaitu *Knowledge Discovery in Databases* (KDD). Pada dasarnya semua memiliki tujuan yang sama, yaitu menghasilkan informasi yang bermanfaat (Prasetyo E. , 2013). Penambahan Data dan terminologi pengetahuan sering ditemukan di database. (KDD) digunakan secara bergantian untuk menggambarkan proses mengekstraksi informasi tersembunyi basis data yang besar. Sebenarnya kedua istilah ini memiliki konsep yang berbeda, tetapi berhubungan satu sama lain. Data mining dalam prosesnya memerlukan satu atau lebih pemodelan data (Fatihatul, dkk., 2011). Dalam penelitian kali ini metode yang akan digunakan adalah metode kalsifikasi.

2.1.2 Klasifikasi

Klasifikasi merupakan salah satu metode *data mining* yang paling populer. Klasifikasi sendiri merupakan proses *data mining* yang bertujuan menemukan model untuk memprediksi tingkah laku konsumen dimasa yang akan datang, berdasarkan pengklasifikasikan rekaman dari *database*, yang kemudian akan dijelaskan atau dibedakan kelas-kelasnya pada setiap data. Klasifikasi sendiri adalah sebuah proses mengumpulkan data yang memiliki kesamaan menurut peneliti yang telah ditentukan sebelum dianalisis (Jollyta, dkk., 2020). Contohnya, untuk sebuah perusahaan menilai tingkat kepuasan pelanggannya, dilakukanlah sebuah analisis berdasarkan faktor-faktor yang dinilai berkaitan serta dapat dimodelkan untuk memprediksi kepuasan pelanggannya. Agar dimasa yang akan datang perusahaan dapat memperbaiki sistem-sistem yang dinilai masih kurang memuaskan bagi para customer. Proses klasifikasian terdiri atas 3 tahapan, yaitu (Han & Kamber, 2001):

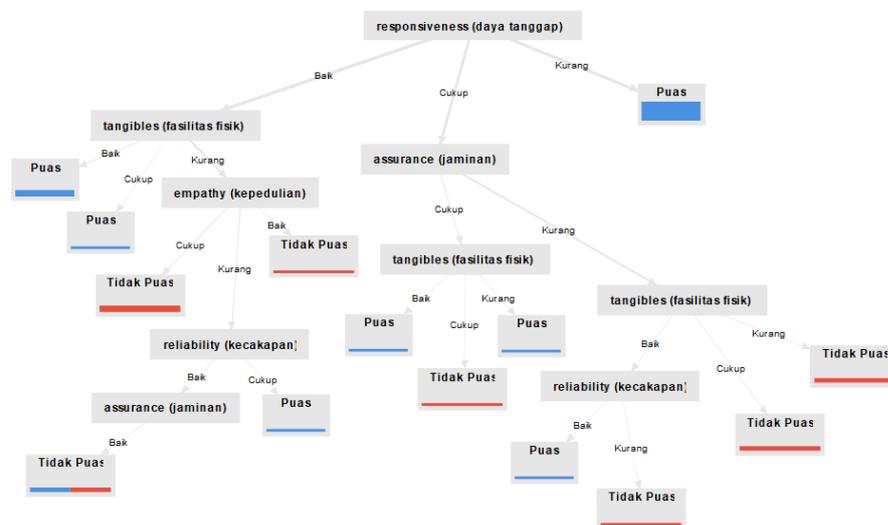
1. Hal pertama yang perlu dilakukan adalah mengumpulkan rekaman dari data yang akan diuji. Kemudian berdasarkan data tersebut akan dibuatlah deskripsi data berdasarkan model klasifikasinya. Tahapan ini disebut dengan *learning*.
2. Pada tahap selanjutnya yang perlu dilakukan adalah model dari data yang telah dibuat akan di lakukan beberapa uji coba dengan menggunakan algoritma tertentu sesuai dengan yang dibutuhkan. Pada tahap ini rekaman dari hasil uji coba harus berbeda dengan data uji (*training set*) sebelumnya. Agar model klasifikasi dapat bekerja pada data uji dengan semestinya.
3. Pada tahap akhir, hasil dari pengklasifikasian akan dimodelkan dengan beberapa model klasifikasi, diantaranya yang populer adalah *neural network*, *decision tree*, dan lain sebagainya.

2.1.3 *Decision Tree*

Basis data yang besar merupakan hal yang biasa kita temui dalam dunia digital saat ini. Ada banyak teknik yang dapat diterapkan dalam pengolahan *data mining*. Salah satu metode yang paling sering digunakan adalah metode klasifikasi data. Di antara berbagai teknik klasifikasi, ditemukanlah *decision tree* yang menjadi populer karena metode ini cukup sederhana, ekspresif, kuat, dan efisien (Panhalkar & Doye, 2021).

Decision Tree memiliki struktur pohon yang terdiri dari *root node* (titik awal), *internal node* (titik percabangan) dan *leaf node* (titik akhir). Setiap *root node* pada *decision tree* akan merepresentasikan keputusan tentang suatu atribut (*attribute*), sedangkan *internal node* akan merepresentasikan *output* atau hasil

keputusan, dan *leaf node* hanya akan memiliki satu input tanpa memiliki *output*. Setelah *decision tree* dibangun, seperangkat aturan dapat dibuat untuk mengevaluasi keadaan agar tegangan sistem tetap stabil. Aturan-aturan ini dapat digunakan sebagai pedoman operasional bagi operator sistem untuk memantau dan mengevaluasi stabilitas tegangan secara *real time* (Meng, dkk., 2020). Adapun contoh *output* dari *decision tree* adalah sebagaimana terlampir pada gambar 2.2.



Gambar 2.2 Output Decision Tree

Tujuan utama dari *decision tree* sendiri adalah untuk menemukan seperangkat aturan keputusan guna memprediksi serangkaian hasil (Yang, dkk., 2017). Metode ini memiliki efisiensi yang cukup tinggi dan kompleksitas waktu yang relatif rendah, serta tidak menggunakan akurasi klasifikasi sebagai fungsi evaluasi.

Decision Tree banyak diterapkan untuk memecahkan masalah dibidang representasi pengetahuan, klasifikasi, optimasi kombinatorial, geometri komputasi, dan lain sebagainya. Manfaat utama dari penggunaan *decision tree* yang menjadikannya cukup populer adalah kemampuannya dalam penjabaran

proses pengambilan keputusan yang kompleks menjadi lebih simpel, sehingga akan lebih memudahkan dalam pengambilan keputusan untuk menginterpretasi solusi dari sebuah permasalahan. Selain itu *decision tree* juga berfungsi untuk mengeksplorasi data, guna menemukan keterkaitan calon variabel *input* dengan variabel *target* (Jollyta, dkk., 2020).

Seperti algoritma data mining pada umumnya, *decision tree* juga memiliki beberapa kekurangan diantaranya. Sangat memungkinkan terjadi tumpang tindih (*overlap*) ketika kelas dan kriteria yang digunakan lebih banyak, sehingga *decision tree* akan lebih mudah dipengaruhi oleh sampel maupun sub-pohon yang diulang berkali-kali dalam pengambilan keputusan. Untuk masalah *over-fitting* ini, kita bisa mengatasinya dengan cara teknologi pemangkasan dan validasi *k-fold cross* (Zhou, dkk., 2021).

2.1.4 Algoritma Iterative Dichotomiser (ID3)

Dalam prosesnya *data mining* membutuhkan satu atau beberapa algoritma untuk bisa menghasilkan pola ataupun informasi baru. Salah satu metode yang umum digunakan banyak orang adalah metode klasifikasi *decision tree*. Namun untuk menentukan *root node* (titik awal), *internal node* (titik percabangan) dan *leaf node* (titik akhir) yang baik kita membutuhkan algoritma. Algoritma *Iterative Dichotomiser Tree* (ID3) adalah salah satu dari sekian banyak algoritma yang cukup baik untuk diaplikasikan pada *decision tree*. Algoritma ID3 pertama kali dikembangkan oleh J. Ross Quinlan pada tahun 1986 (Jollyta, dkk., 2020).

Fungsi utama dari algoritma ID3 adalah untuk mendapatkan informasi yang terkandung dalam kriteria pemilihan atribut untuk perhitungan atribut uji hierarkis dari node *non-leaf* (internal) di *decision tree*. Tujuan dari algoritma ini

adalah untuk mendapatkan informasi kelas maksimum tentang sub-set data saat membuat keputusan pada *internal node*. Secara khusus, langkah pertama yang perlu dilakukan adalah menghitung perolehan informasi untuk setiap atribut. *Root node* dari *decision tree* kemudian dipilih dengan menentukan perolehan informasi maksimum. Berdasarkan percabangan, secara iteratif, algoritma akan bertindak sebagai alat penghitungan informasi, mulai dari pembuatan *internal node* baru, hingga sub-set data di setiap cabang dengan label yang sama pada setiap kelas. Oleh karena itu, *decision tree* yang terbangun digunakan untuk mengklasifikasikan contoh-contoh baru (Yang, dkk., 2017).

Algoritma ID3 dapat diimplementasikan menggunakan fungsi rekursif (fungsi yang dapat memanggil dirinya sendiri). Algoritma ID3 mencoba membangun *decision tree* secara *top-down* (mengurutkan dari atas ke bawah), dengan memeriksa atribut mana yang pantas untuk ditempatkan sebagai *root*. Algoritma ini melakukan pencarian secara menyeluruh (*greedy*) pada semua data yang berkemungkinan akan mempengaruhi *decision tree* (Tyasti, dkk., 2015). Algoritma ID3 memilih atribut uji dengan menghitung dan membandingkan informasi yang telah diperoleh. Algoritma tersebut memiliki konsep *entropy* informasi (Meng, dkk., 2020).

Entropy sendiri adalah ukuran teoretis informasi yang mencirikan ketidakmurnian, serta homogenitas kumpulan dari suatu data. Sebelum menentukan *root node*, *internal node*, dan *leaf node*, kita perlu mengetahui nilai *entropy* secara keseluruhan ($Entropy(S)$). Untuk mendapatkan nilai *entropy* keseluruhan kita dapat menggunakan persamaan, sebagai berikut (Meng, dkk., 2020):

$$Entropy(S) = (-p_+ \log_2 p_+) - (-p_- \log_2 p_-) \quad (2.1)$$

Keterangan:

S = Himpunan kasus

p_+ = Jumlah solusi positif (mendukung)

p_- = Jumlah solusi negatif (tidak mendukung)

Setelah mendapatkan nilai *entropy* secara keseluruhan, kemudian perlu bagi kita mengetahui nilai masing-masing partisi ($Entropy(S_i)$). Untuk mendapatkan nilai partisi ke- i , kita dapat menggunakan persamaan sebagai berikut (Meng, dkk., 2020):

$$Entropy(S_i) = (-p_+ \log_2 p_+) - (-p_- \log_2 p_-) \quad (2.2)$$

Keterangan:

S_i = Jumlah kasus pada partisi ke- i

Berdasarkan nilai *entropy*, kemudian kita dapat menghitung nilai *information gain* dari setiap atribut untuk menentukan *internal node* pada *decision tree* (Meng, dkk., 2020).

$$Gain(S, A) = Entropy(S) - \sum_i^n \frac{|S_i|}{|S|} Entropy(S_i) \quad (2.3)$$

Keterangan:

A = Atribut

n = Jumlah partisi S

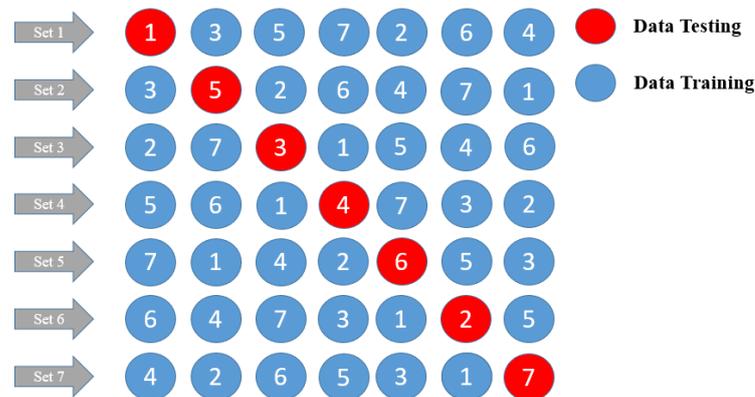
i = Jumlah atribut ke- i

Jadi, secara ringkas pada tahap pertama hal yang harus kita lakukan adalah menghitung nilai *entropy* keseluruhan ($Entropy(S)$) dan nilai *gain* pada setiap atribut yang ada ($Gain(S, A)$). Keduanya berfungsi sebagai pembanding untuk menentukan *root node* dan membuat simpul (*internal node*) pada *decision tree*. Proses perhitungan *information gain* akan terus dilakukan sampai semua atribut yang ada, masuk kedalam kelas-kelas dalam simpul *decision tree*. Atribut yang telah mencapai titik keputusan tidak perlu di hitung *gain*-nya lagi. Contohnya, dalam sebuah data terdapat faktor-faktor yang mempengaruhi kepuasan pelanggan

pada suatu toko, dimisalkan pada data tersebut semua yang memillih sangat baik pada atribut pelayanan toko, juga memilih puas dengan toko tersebut. Maka, simpul ini tidak perlu untuk dihitung ulang.

2.1.5 Cross Validation

Cross validation merupakan sebuah metode pada *data mining* yang berfungsi sebagai alat untuk mengavaluasi model algoritma dengan beberapa tahap. Dalam prosesnya data yang akan digunakan untuk membuat sebuah model akan disebut data latih atau data *training* sedangkan data yang digunakan untuk memvalidasi disebut data uji atau data *testing*. Pada umumnya peneliti akan membagi data menjadi 10 bagian pada tahap *cross validation*. Dalam prosesnya data set akan dilakukan 10 kali *record* sampai seluruh data mendapat kesempatan memnjadi data *testing* ataupun data *training* (Yuniarti & Hartanti, 2019). Hal ini dapat dianalogikan seperti pada gambar 2.3.



Gambar 2.3 Proses Seleksi Data *Cross Validation*

Menurut Indrayanti, dan Karomi (2017) peneliti biasa menggunakan 10 *folds* dikarenakan menghasilkan performa algoritma yang lebih stabil (Indrayanti, & Karomi, 2017).

2.1.6 Confusion Matrix

Evaluasi merupakan salah satu hal yang perlu dilakukan dalam proses pengklasifikasian *data mining*. Proses evaluasi ini bertujuan untuk mengetahui tingkat akurasi dari model algoritma yang digunakan. Evaluasi dalam penelitian ini akan dilakukan dengan *confusion matrix* untuk mendapatkan nilai *precision* dan *recall*. *Confusion matrix* adalah tabel yang terdiri dari beberapa baris data uji yang diprediksi benar ataupun salah, berdasarkan model klasifikasinya. Tabel tersebut diperlukan untuk mengetahui *performance* dari sebuah model klasifikasi (Nasution & Hayaty, 2019).

	Positive	Negative	
Positive	TP	FN	TP + FN
Negative	FP	TN	FP + TN

Gambar 2.4 Tabel Perhitungan Confusion Matrix

Berdasarkan gambar 2.4 TP (*True Positive*) adalah data positif yang terdeteksi benar, sedangkan FP (*False Positive*) adalah data negatif namun terdeteksi positif. Sedangkan TN dan FN merupakan kebalikan dari TP dan FP. TN (*True Negative*) adalah jumlah data negatif yang terdeteksi benar, sedangkan FN (*False negative*) adalah data negatif namun terdeteksi positif.

Akurasi adalah rasio dari keseluruhan prediksi benar, untuk mengetahui besar akurasi dapat dihitung dengan rumus sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.4)$$

Sedangkan *precision* merupakan perbandingan antara *true positive* dengan total data yang berlabel positif. *Precision* dapat kita hitung dengan rumus, sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (2.5)$$

Dan *recall* merupakan ukuran kelengkapan dari suatu model. Kita dapat mengetahui besar nilai *recall* berdasarkan perbandingan *true positive* terhadap total keseluruhan, sebagai berikut :

$$Recall = \frac{TP}{TP+FN} \quad (2.6)$$

Nilai *accuracy* sebagai alat pengukur untuk penilaian kinerja suatu algoritma dalam proses klasifikasi tentu memiliki *rule of thumb* pada nilai *accuracy*-nya. Adapun tingkatan nilai *accuracy* dari sangat baik sampai sangat buruk menurut (Gorunescu, 2011) sebagai terdapat pada tabel 2.1:

Tabel 2.1 *Rule of thumb* Nilai *Accuracy*

No	Nilai <i>Accuracy</i>	Keterangan
1	>0,90 (>90%)	Klasifikasi sangat baik
2	0,80-0,90 (80%-90%)	Klasifikasi baik
3	0,70-0,80 (70%-80%)	Klasifikasi cukup baik
4	0,60-0,70 (60%-70%)	Klasifikasi buruk
5	<0,60 (<60%)	Klasifikasi sangat buruk

Berdasarkan tabel 2.1 dapat kita ketahui bahwa apabila nilai *accuracy* dari suatu model kurang dari 60% maka algoritma tersebut sangat buruk dalam mengklasifikasikan suatu data. Dan apabila hasil *accuracy* dari data tersebut lebih besar dari 90% maka algoritma tersebut dinilai sangat baik dalam mengklasifikasikan data penelitian.

2.1.7 Putus Sekolah

Indonesia memiliki masalah pendidikan yang cukup serius, menyamakan pendidikan dari Sabang-Merauke serta memberikankan pendidikan yang bermutu dan berkualitas bukanlah hal yang mudah. Tantangan tersebut semakin

hari semakin berat, mengingat standar kualitas sumber daya manusia juga semakin tinggi. Pendidikan merupakan salah satu cara bagi pemerintah untuk meningkatkan kualitas sumber daya manusia di Indonesia. Pendidikan aspek utama dalam membuat perubahan ke arah yang lebih baik guna mewujudkan cita-cita bangsa Indonesia. Pendidikan yang berkualitas akan meningkatkan kualitas hidup masyarakat di masa yang akan datang. Jika suatu negara ingin meningkatkan kualitas hidup, pertama-tama harus dimulai dengan perbaikan sistem pendidikan yang mengikuti perkembangan zaman.

Kemiskinan dan terhambatnya pendidikan merupakan dua hal yang tidak dapat dipisahkan. Keduanya memiliki hubungan sebab akibat secara langsung maupun tidak langsung, sehingga dibutuhkan upaya untuk mengatasinya secara bertahap. Kemiskinan menjadi salah satu faktor utama kegagalan bangsa Indonesia untuk dapat memiliki generasi bangsa yang berkualitas. Mahalnya biaya pendidikan menjadi alasan dasar hal ini dasangkut pautkan dengan kemampuan dalam mengenyam pendidikan. Hal ini dilakukan sebagian masyarakat dengan tingkat ekonomi rendah untuk mengurangi beban hidup sehari-hari anggota keluarga. Keputusan ini tentu akan sangat merugikan di kemudian hari karena masyarakat tidak mendapatkan suplai yang cukup untuk bersaing di masa mendatang.

Putus sekolah mengacu pada situasi dimana anak masuk sekolah dasar, sekolah menengah pertama, dan sekolah menengah atas untuk belajar dan menerima kelas, tetapi putus sekolah atau berhenti sekolah tanpa menyelesaikan studi sampai tamat. Putus sekolah juga dapat dijelaskan sebagai seorang anak yang pernah duduk di bangku sekolah kemudian memutuskan untuk berhenti atau

tidak melanjutkan sekolah ke jenjang yang lebih tinggi. Menurut para ahli, ada 3 kriteria putus sekolah yaitu, yang pertama siswa yang berhenti sekolah tetapi masih pada tingkat tertentu, yang ke-2 jika anak tidak sempat menyelesaikan mata pelajaran sekolah tertentu, atau berhenti pada tingkat terakhir, dan yang ke-3 yaitu dimana anak tidak melanjutkan ke tingkat yang lebih tinggi (Yuniar, 2021). Berdasarkan beberapa penelitian sebelumnya atribut yang mempengaruhi siswa dalam mengambil keputusan untuk melanjutkan atau putus sekolah adalah sebagai berikut:

1. Jenis kelamin

Dalam beberapa penelitian terdahulu menyebutkan bahwa jenis kelamin berperan dalam pengambilan keputusan untuk melanjutkan atau putus sekolah. Salah satunya penelitian yang dilakukan oleh Richard (2016), pada penelitian tersebut anak laki-laki memiliki kecenderungan putus sekolah lebih tinggi ketimbang perempuan. Dalam penelitian lainnya yang dilakukan oleh Dalton, dkk (2009) dalam penelitiannya menyebutkan anak laki-laki memiliki kecenderungan untuk putus sekolah lebih besar yaitu 7% sedangkan anak perempuan hanya 6%.

Begitu juga dengan penelitian yang dilakukan oleh Kusbudiyanto dan Adis (2019) pada salah satu SMK di kota Bekasi yang menghasilkan bahwa angka putus sekolah pada SMK tersebut didominasi oleh siswa laki-laki. Hal ini sejalan dengan anak laki-laki sebagai sumber daya tenaga kerja memiliki kewajiban lebih besar dalam menambah penghasilan keluarga dibandingkan anak perempuan. Apabila sebuah keluarga memiliki keterbatasan dalam kondisi keuangan tentu mempengaruhi keputusan anak untuk putus sekolah. Anak laki-

laki yang menjadi harapan untuk bisa membantu tulang punggung keluarga dalam mencari nafkah menjadikan alasan terbesar untuk putus sekolah (Sudarwati, 2009).

2. Pendidikan Orang Tua

Pola asuh adalah cara orangtua melakukan serangkaian tindakan untuk mendidik anaknya (Gunarsa, 2002). Menurut Hurlock (2009) ada beberapa faktor yang dapat mempengaruhi pola asuh orangtua terhadap anak, salah satunya adalah tingkat pendidikan orangtua. Orangtua yang berpendidikan tinggi biasanya memiliki teknik sendiri dalam mengasuh anaknya, dibandingkan orangtua yang kurang dalam hal pendidikan.

Sugianto (2017) dalam penelitiannya mendapatkan 10 responden atau 29,41% kasus putus sekolah disebabkan oleh tingkat pendidikan orangtuanya yang rendah. Oleh sebab itu faktor pendidikan orangtua juga berpengaruh pada anak dalam mengambil keputusan. Namun, tidak sedikit orang tua yang berpendidikan rendah juga mampu menyekolahkan tinggi anaknya bahkan sampai bangku perkuliahan karena tidak ingin anaknya bernasib sama dengan dirinya (Sugianto, 2017).

3. Pekerjaan Orangtua

Pendapatan orang tua yang kurang dari upah minimum menjadi salah satu faktor yang mendorong seorang siswa untuk memilih putus sekolah. Kebutuhan ekonomi yang semakin tinggi dengan pendapatan yang rendah menjadi pendorong untuk anak ikut mencari nafkah sehingga harus mengorbankan sekolahnya. Oleh sebab itu pekerjaan orangtua juga menjadi salah satu faktor yang mempengaruhi anak putus sekolah (Wassahua, 2016).

Jenis pekerjaan menjadi salah satu faktor yang mempengaruhi penghasilan (Cahyono, 1998). Untuk memenuhi kebutuhan manusia membutuhkan penghasilan, dan untuk mendapatkan penghasilan seseorang harus bekerja. Berbeda jenis pekerjaannya juga berbeda penghasilannya (Putri & Setiawina, 2013).

4. Anggota Keluarga

Sugianto (2017) dalam penelitiannya mendapati responden putus sekolah dengan jumlah anggota keluarga lebih dari 4 sebesar 60%. Artinya semakin banyak anggota keluarga maka akan semakin banyak tanggungan yang harus di tanggung dalam keluarga tersebut. Hal ini membuktikan bahwa banyaknya anggota keluarga bisa menjadi salah satu faktor yang dapat mempengaruhi siswa putus sekolah.

5. Pendapatan Keluarga

Sugianto (2017) dalam penelitiannya mendapati variabel pendapatan keluarga memiliki nilai *odds ratio (OR)* 0,74 yang berarti variabel ini berpengaruh dalam keputusan siswa untuk memutuskan putus sekolah. Pola tersebut menggambarkan banyak siswa putus sekolah terjadi pada siswa yang tingkat pendapatan keluarganya rendah. Bukan masalah besar bagi keluarga yang memiliki penghasilan rata-rata perbulan sedang bahkan tinggi untuk menyekolahkan anaknya, hal ini akan berbanding terbalik dengan mereka yang lahir dalam keluarga yang pendapatannya tergolong rendah (Richard, 2016).

2.2 Pentingnya Pendidikan dalam Al-Qur'an

Dunia yang semakin hari semakin bersifat dinamis, tentunya menjadi tantangan tersendiri bagi setiap orang, salah satunya adalah pihak pendidikan. Mengingat pendidikan sebagai tolak ukur kualitas seseorang. Sebagai pendidik harus terus berfikir untuk meningkatkan semangat dan kesadaran tentang pentingnya pendidikan bagi setiap manusia, hal ini bermaksud agar siswa-siswi kedepannya dapat menjadi generasi bangsa yang berwawasan luas dan berintelektual tinggi sehingga mampu bersaing didalam maupun diluar negeri. Hal inilah yang menjadi alasan mengapa generasi bangsa kedepannya harus memahami pentingnya menuntut ilmu.

Begitupun Islam mengajarkan kita untuk menuntut ilmu, bahkan ayat yang pertama kali diturunkan oleh Allah SWT kepada Nabi Muhammad SAW di Goa Hira berbunyi “*Iqro*” yang berarti bacalah. Bahkan kata ini di ulang sebanyak 2 kali sebagaimana Allah SWT berfirman dalam surah Al-alaq ayat 1-5:

أَقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ (١) خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ (٢) اقْرَأْ وَرَبُّكَ الْأَكْرَمُ (٣) الَّذِي عَلَّمَ بِالْقَلَمِ (٤) عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ (٥)

Artinya: “*Bacalah dengan (menyebut) nama Tuhanmu yang menciptakan Dia menciptakan manusia dari segumpal darah. Bacalah dan Tuhanmu maha pemurah. Yang mengajarkan (manusia) dengan perantara kalam. Dia mengajarkan kepada manusia apa yang tidak diketahuinya.*”

Dari potongan ayat di atas jelas Allah memerintahkan Nabi Muhammad SAW beserta umatnya untuk membaca agar memperoleh pengetahuan mengenai ilmu agama maupun ilmu alam dan sosial. Mencari ilmu wajib hukumnya bagi setiap umat muslim mulai dari lahir sampai masuk kedalam kubur. Hal ini sejalan dengan kebijakan pemerintah yang mewajibkan rakyatnya untuk wajib belajar 12 tahun. Menambah pengetahuan harus dilakukan baik dalam lembaga

formal seperti sekolah maupun informal diluar sekolah terutama bagi mereka yang masih berada di usia sekolah (Hidayat, 2014).

Lantas, mengikuti kajian yang disampaikan oleh Rasulullah dan firman Allah SWT, yang tertulis dalam Al-Qur'an menyebutkan, bahwa keistimewaan manusia dibandingkan dengan makhluk lainnya untuk menjalankan kepemimpinan di dunia adalah karena kita diberi akal untuk berfikir dan raga untuk belajar sebagaimana tertulis pada Q.S Al-Baqarah ayat 31-32:

صٰدِقِيْنَ (۳۱) وَعَلَّمَ ءَادَمَ الْاَسْمَاءَ كُلَّهَا ثُمَّ عَرَضَهُمْ عَلَى الْمَلٰٓئِكَةِ فَقَالَ اَنْبِئُوْنِي بِاَسْمَاءِ هٰٓؤُلَآءِ اِنْ كُنْتُمْ قٰلُوْا سُبْحٰنَكَ لَا عِلْمَ لَنَا اِلَّا مَا عَلَّمْتَنَا اِنَّكَ اَنْتَ الْعَلِيْمُ الْحَكِيْمُ (۳۲)
(سورة البقرة: ۳۱-۳۲)

Artinya: *“Dan Dia mengajarkan kepada Adam nama-nama (benda-benda) seluruhnya, kemudian mengemukakannya kepada para malaikat lalu berfirman” Sebutkanlah kepada-Ku nama benda-benda itu jika kamu memang benar orang-orang yang benar!” mereka menjawab:” Maha suci Engkau, tidak ada yang kami ketahui selain daripada apa yang telah Engkau ajarkan kepada kami; Sesungguhnya Engkau lah yang Maha mengetahui lagi Maha Bijaksana. (Q.S. Al-Baqarah : 31-32)”*.

Kata “nama-nama” yang disebut dalam ayat tersebut adalah tentang sifat, ciri dan hukum sesuatu, yang berarti manusia memiliki peluang untuk mengetahui rahasia yang ada di alam semesta (Shihab, 2006). Dalam hal ini secara tidak langsung Allah SWT berfirman bahwa bahwasannya suatu hari nanti manusia akan mengembangkan ilmu pengetahuan atas izin-Nya dengan akal yang diberikan oleh Allah SWT. Quraish Shihab pernah mengutip ungkapan dari A.N. Whitehead dalam bukunya yang berjudul *Science and The Modern World*, “Bila kita menyadari betapa pentingnya agama dan ilmu pengetahuan bagi manusia, maka bukanlah hal yang berlebihan jika sejarah kita dimasa yang akan datang bergantung pada keputusan generasi sekarang tentang hubungan keduanya” (Shihab, 1999).

Berdasarkan kedua dalil di atas yang sudah dijelaskan mengenai pentingnya pendidikan di masa kini maupun dimasa depan, entah itu pendidikan agama maupun pendidikan formal. Penting untuk kedua ilmu tersebut dapat berdampingan dalam membangun generasi yang cerdas berwawasan serta berintelektual tinggi.

Pendidikan dapat ditempuh dengan cara menuntut ilmu atau belajar. Dalam islam, pendidikan sangatlah diperhatikan. Islam memandang bahwa pendidikan harus ditempuh oleh semua, terbukti dengan hadist yang berbunyi (Bukhari, 2012):

حَدَّثَنَا هِشَامُ بْنُ عَمَّارٍ حَدَّثَنَا حَفْصُ بْنُ سُلَيْمَانَ حَدَّثَنَا كَثِيرُ بْنُ شَنْظِيرٍ عَنْ مُحَمَّدِ بْنِ شَيْرِينَ عَنْ أَنَسِ بْنِ مَالِكٍ قَالَ : قَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ طَلَبُ الْعِلْمِ فَرِيضَةٌ عَلَى كُلِّ مُسْلِمٍ وَوَاضِعُ الْعِلْمِ عِنْدَ غَيْرِ أَهْلِهِ كَمُقَلَّدِ النَّحْتَازِيرِ الْجَوْهَرَ وَاللُّؤْلُؤِ وَالذَّهَبِ (رواه ابن ماجه)

Artinya: *Hisyam bin 'Ammar menceritakan kepada kami, Hafs bin Sulaiman menceritakan kepada kami, Katsir bin Syindzir menceritakan kepada kami dari Muhammad bin Syirin, dari Anas bin Malik berkata, Rasulullah SAW. bersabda : "Mencari ilmu itu wajib bagi setiap muslim, dan orang yang meletakkan ilmu pada selain ahlinya bagaikan menggantungkan permata mutiara dan emas pada babi hutan". (HR. Ibnu Majjah).*

Mencari ilmu dalam konteks ini merupakan suatu keharusan bagi setiap muslim. Hal ini tentunya dalam rangka agar setiap muslim dapat berubah dari yang awalnya tidak tau menjadi tau. Selain itu, dalam Islam, seorang pencari Ilmu juga mempunyai keutamaan sebagai yang tertera dalam hadits berikut:

عَنْ أَبِي هُرَيْرَةَ قَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ مَنْ سَلَكَ طَرِيقًا يَلْتَمِسُ فِيهِ عِلْمًا سَهَّلَ اللَّهُ لَهُ طَرِيقًا إِلَى الْجَنَّةِ (رواه ابن ماجه و مسلم)

Artinya: *Dari Abu Hurairah, ia berkata bahwa Rasulullah SAW. pernah bersabda, "Barangsiapa yang menempuh perjalanan untuk mencari ilmu, maka Allah Swt. akan memberikan kepadanya kemudahan jalan menuju surga. " (HR. Ibnu Majjah dan Muslim)*

Maksud dari hadits diatas adalah seorang pencari Ilmu, akan dimudahkan jalannya menuju ke surga oleh Allah. Hal ini menunjukkan bahwa Islam sangat memuliakan orang yang sedang mencari ilmu. Dalam kata lain, seseorang yang menuntut ilmu atau menempuh pendidikan akan mendapat keistimewaan tersendiri dari Allah (Siregar, 2014).

2.3 Kajian Putus Sekolah dengan Metode Klasifikasi *Decision Tree*

Penelitian yang dilakukan kali ini adalah proses klasifikasi faktor penyebab siswa putus sekolah dengan menggunakan metode *decision tree* yang dioptimalkan dengan algoritma *ID3*. Algoritma *decision tree* merupakan algoritma yang paling umum digunakan dalam mengklasifikasikan *data mining* dengan tujuan untuk memprediksi tingkah laku dimasa yang akan datang berdasarkan tingkah laku dimasa sekarang (Jollyta, dkk., 2020). Dalam prosesnya data akan dibagi menjadi dua yaitu data *training* dan data *testing* dengan menggunakan teknik *K-Fold cross validation*.

Pada tahap awal peneliti akan menentukan atribut apa saja yang akan digunakan berdasarkan data yang ada, kemudian atribut tersebut akan melalui tahap *preprocessing* data, tujuannya adalah untuk menyeleksi data atau atribut yang dianggap dapat menghambat jalannya penelitian. Setelahnya data akan dibagi menjadi data *training* dan data *testing* dengan menggunakan teknik *K-Fold cross validation*. Hal ini dilakukan dengan tujuan untuk mendapatkan nilai akurasi yang maksimal. Dalam prosesnya peneliti akan menggunakan software *RapidMiner*. Sehingga nantinya akan didapatkan *output* berupa faktor yang

mempengaruhi siswa dalam mengambil keputusan untuk melanjutkan atau putus sekolah.

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Data dalam penelitian ini bersifat ilmiah yang disusun secara sistematis, sehingga penelitian ini dapat dikategorikan dalam jenis penelitian kuantitatif. Penelitian akan dilakukan dengan menggunakan metode klasifikasi *decision tree* dengan algoritma *ID3*. Pada penelitian ini keputusan akan diambil berdasarkan atribut yang disusun menjadi *node-node* dalam *decision tree*.

3.2 Data dan Sumber Data

Sumber data dalam penelitian ini merupakan data sekunder. Data tersebut merupakan data tahunan yang diambil pada tahun 2019-2022 dari salah satu Sekolah Menengah Pertama di Desa Panagan, Kecamatan Gapura, Kabupaten Sumenep Jawa Timur. Data yang digunakan dalam penelitian ini merupakan keterangan siswa-siswi kelas IX selama 3 tahun terakhir mengenai keputusan siswa-siswi yang bersangkutan untuk melanjutkan ataupun putus sekolah. Data yang ada berjumlah 335 dataset, dimana 7 diantaranya memutuskan untuk putus sekolah dan 348 lainnya memutuskan untuk tetap melanjutkan sekolah. Adapun atribut yang digunakan adalah sebagaimana yang terdapat pada tabel 3.1 berikut:

Tabel 3.1 Atribut *Decision Tree*

Jenis Kelamin	Jenis kelamin siswa.
Pendidikan Ayah	Pendidikan terakhir ayah dari siswa yang bersangkutan.
Pekerjaan Ayah	Mata pencaharian ayah dari siswa yang bersangkutan.
Pendidikan Ibu	Pendidikan terakhir ibu dari siswa yang bersangkutan.
Pekerjaan Ibu	Mata pencaharian ibu dari siswa yang bersangkutan.
Anggota Keluarga	Banyak anggota keluarga dalam satu atap tempat siswa yang bersangkutan tinggal.
Pendapatan Keluarga	Tingkat pendapatan ekonomi keluarga dari siswa yang bersangkutan.

3.3 Teknik Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan data tahunan dari salah satu Sekolah Menengah Pertama di Desa Panagan, Kecamatan Gapura, Kabupaten Sumenep Jawa Timur yang dilaporkan setiap tahunnya. Sehingga data ini termasuk data sekunder, karena data tersebut memang sudah ada setiap tahunnya. Data tersebut didapatkan setelah peneliti mengajukan permohonan izin untuk dapat menggunakan data keterangan siswa-siswi untuk melanjutkan atau putus sekolah ditahun tersebut.

3.4 Variabel Penelitian

Variabel yang digunakan dalam proses penyusunan data skripsi ini, diantaranya terdapat 2 variabel yaitu, variabel bebas (*Independent*) (x) yang meliputi 7 atribut, diantaranya jenis kelamin, pendidikan terakhir ayah, pekerjaan ayah, pendidikan terakhir ibu, pekerjaan ibu, jumlah anggota keluarga dalam satu rumah, dan tingkat pendapatan ekonomi keluarga dari siswa-siswi yang bersangkutan setiap bulannya. Kemudian terdapat juga variabel terikat (*Dependent*) (y) yang dipengaruhi oleh variabel bebas (x) yaitu sebagai label

atau kelas target (*Output*) pada penelitian ini, yang berupa status melanjutkan atau putus sekolah siswa kelas IX pada tahun ajaran 2019-2022.

3.5 Teknik Analisis Data

Dalam penyusunan skripsi ini terdapat beberapa tahapan yang perlu dilakukan. Mulai dari penyusunan atribut yang nantinya akan melalui beberapa tahap mulai dari deskriptif data, *data cleaning*, selanjutnya pada tahap *clasifier* dataset akan diklasifikasikan dengan menggunakan model algoritma *decision tree ID3*. Kemudian pada tahap akhir data akan dievaluasi dengan menggunakan *confusion matrix* untuk mendapatkan nilai *accuracy*, *precision*, dan *recall*.

3.5.1 Analisis Deskriptif Data

Pada tahap awal dalam proses pengolahan data, data akan di deskripsikan. Pada tahap ini peneliti akan mendeskripsikan mengenai informasi apa saja yang terkandung dalam data tersebut.

3.5.2 Cross Validation

Pada tahap awal dalam proses pengolahan data, data akan dibagi menjadi 2, yaitu data *testing* dan data *training* dengan menggunakan *Cross Validation*, dimana terdapat ketentuan bahwa banyak data *testing* lebih dari data *training*. Nantinya data *training* akan melewati uji model klasifikasi *data mining*. Software yang digunakan dalam proses ini adalah *RapidMiner*.

3.5.3 Data Selection

Sebelum diolah, dataset perlu melalui tahap Pada tahap ini data akan diseleksi berdasarkan kriteria tertentu. Peneliti akan memilih atribut sesuai

dengan kriteria yang telah ditentukan. Proses ini akan bergantung dengan tujuan dari proses *data mining*.

3.5.4 Pre-processing (Data Cleaning)

Pada tahap ini data yang memiliki *missing value* dan *noisy* akan disingkirkan karena dapat mengganggu konsistensi dalam dataset. Proses ini bertujuan untuk meminimalisir terjadinya *error* saat proses pengolahan data.

3.5.5 Data Transformasi

Kegiatan ini dilakuakn dengan tujuan untuk mentransformasi data yang kurang sesuai dengan model data yang dibutuhkan. Tujuannya adalah agar data dapat digunakan dan ditelusuri. Dengan kata lain data akan dipetakan secara kompleks.

3.5.6 Classifier

Tahap selanjutnya akan dilakukan pemilihan teknik, metode maupun algoritma yang dianggap sesuai dengan tujuan dan prosesnya. Pada tahap ini peneliti memutuskan untuk menggunakan salah satu teknik klasifikasi *data mining*, dimana dataset yang sudah dikelompokkan akan dilakukan uji menggunakan algoritma *ID3*.

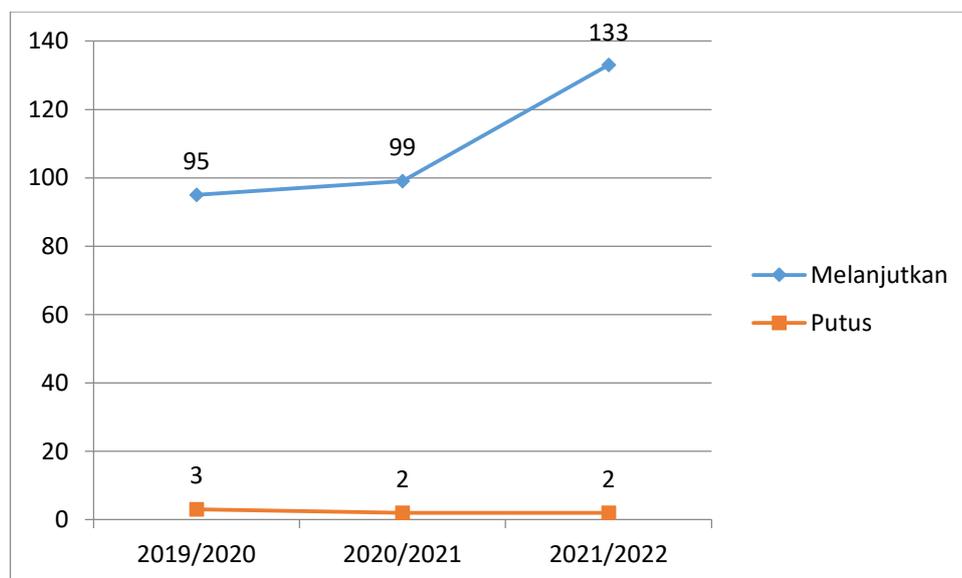
3.5.7 Evaluasi

Pada tahap akhir dalam penelitian ini akan dilakukan evaluasi, tujuannya adalah untuk mengetahui tingkat akurasi dari model algoritma yang digunakan. Nantinya evaluasi dalam penelitian ini akan dilakukan dengan *confusion matrix* untuk mendapatkan nilai *accuracy*, *precision* dan *recall*. *Software* yang digunakan untuk menyelesaikan tahapan ini adalah *RapidMiner*.

BAB IV HASIL DAN PEMBAHASAN

4.1 Analisis Deskriptif Data

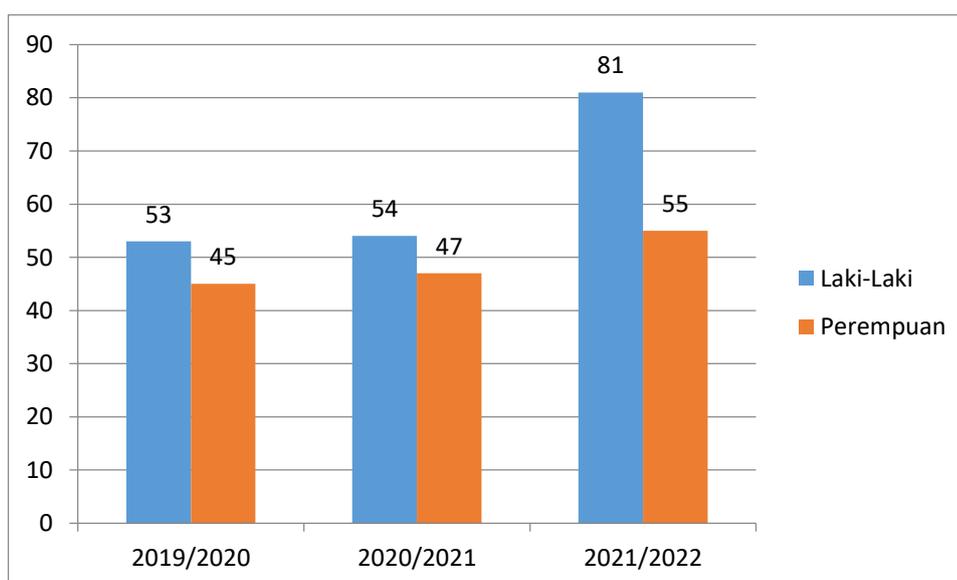
Pada bab ini akan menjelaskan tentang analisis data penelitian, pembahasan serta hasil dari pengklasifikasian *data mining decision tree* terhadap faktor penyebab siswa putus sekolah dengan menggunakan algoritma *ID3*. Data yang didapat untuk penelitian ini merupakan data tahunan yang disediakan salah satu sekolah di Desa Panangan, Kecamatan Gapura, Kabupaten Sumenep Jawa Timur. Data tersebut merupakan data mengenai keputusan siswa untuk melanjutkan atau putus sekolah, khususnya siswa kelas IX tahun ajaran 2019/2020, 2020/2021, dan 2021/2022. Dimana total keseluruhan terdapat 335 data yang akan diteliti (Lampiran 1). Pada data tersebut terdapat faktor-faktor eksternal yang berkemungkinan mendorong siswa untuk melanjutkan atau putus sekolah.



Gambar 4.1 Diagram Jumlah Siswa Berdasarkan Status Siswa

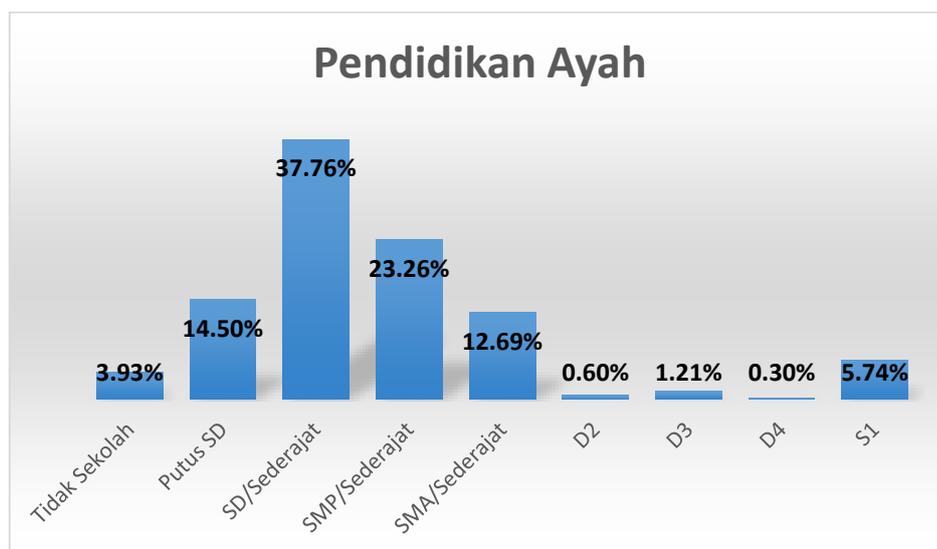
Berdasarkan gambar 4.1 dapat dilihat bahwa seiring berkembangnya zaman kesadaran masyarakat mengenai pentingnya pendidikan semakin tinggi. Hal itu dapat dibuktikan dengan grafik siswa putus sekolah yang menurun dan grafik siswa melanjutkan sekolah yang meningkat. Namun dari diagram tersebut masih ditemukan sebagian kecil siswa yang putus sekolah. Berdasarkan dari data asli sebelum diolah pada lampiran 1 diketahui selama 3 tahun terakhir sebanyak 7 dari 335 siswa memutuskan untuk putus sekolah disebabkan beberapa faktor yang masih belum diketahui sehingga mampu mendorong siswa tersebut untuk mengambil keputusan tidak melanjutkan sekolah.

Informasi yang terkandung dalam lampiran 1 diantaranya adalah Tahun Ajaran, data diri siswa yang terdiri dari NIPD, Nama Siswa, dan NISN. Kemudian juga terdapat data dari ayah siswa yang terdiri dari Nama Ayah, Pendidikan Terakhir Ayah dan Pekerjaan Ayah. Selanjutnya terdapat data ibu dari siswa yang bersangkutan meliputi Nama Ibu, Pendidikan Terakhir Ibu, dan Pekerjaan Ibu. Selain itu juga terdapat informasi mengenai banyak Anggota Keluarga dalam satu rumah serta tingkat Pendapatan Keluarga dalam 1 bulan.



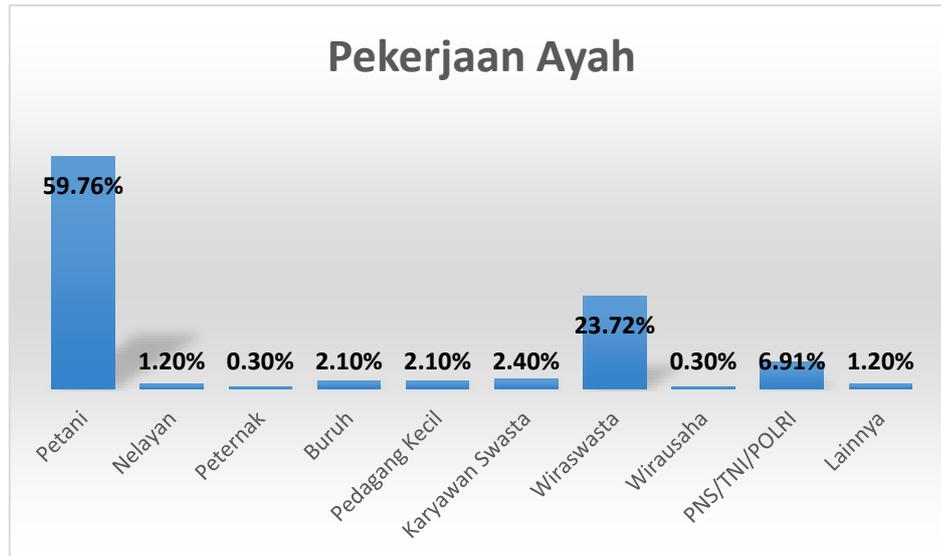
Gambar 4.2 Grafik jumlah Siswa Berdasarkan Jenis Kelamin

Melihat grafik pada gambar 4.2 dapat diketahui bahwa Siswa laki-laki lebih dominan dibandingkan dengan siswa perempuan setiap tahunnya. Namun, banyak siswa laki-laki maupun perempuan yang bersekolah di sekolah tersebut meningkat setiap tahunnya. Hal ini juga sesuai dengan data pada lampiran 1 dimana banyak siswa laki-laki pada tahun ajaran 2019/2020 sebanyak 53 dan banyak siswa perempuan ditahun yang sama adalah 45. Kemudian pada tahun ajaran 2020/2021 terdapat 54 siswa laki-laki dan 47 siswa perempuan. Dan pada tahun 2021/2022 banyak siswa laki-laki adalah 81 sedangkan banyak siswi perempuan adalah 55.



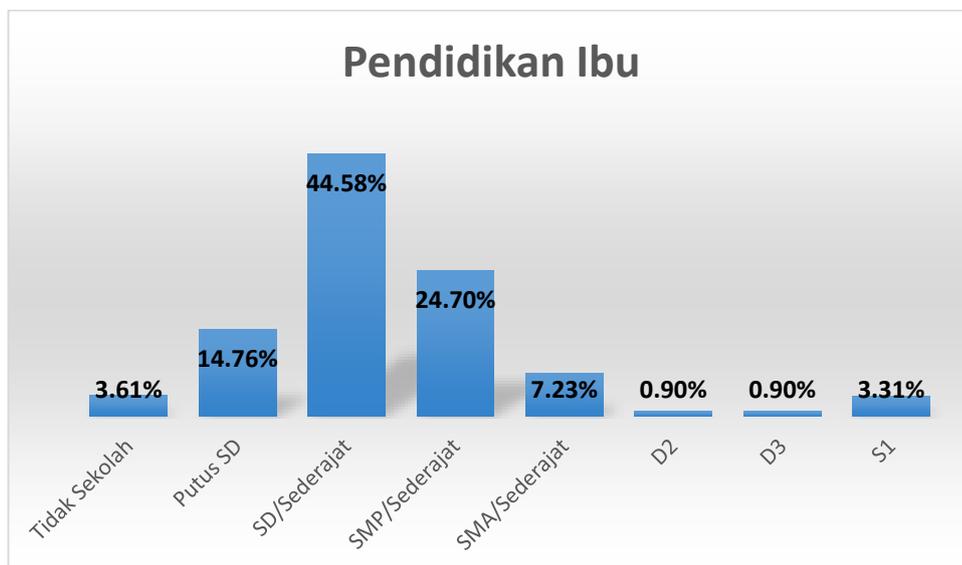
Gambar 4.3 Prosentase Atribut Pendidikan Ayah

Berdasarkan grafik pada gambar 4.2 dapat diketahui bahwa terdapat 9 kategori dari atribut pendidikan ayah yaitu, Tidak Sekolah, Putus SD, SD/ sederajat, SMP/ sederajat, SMA/ sederajat, D2, D3, D4 dan S1. Dari gambar tersebut juga dapat terlihat bahwa siswa dengan pendidikan terakhir ayah SD/ sederajat memiliki prosentase tertinggi yaitu 37,76%, sebanyak 125 siswa menurut data pada lampiran 1.



Gambar 4.4 Prosentase Atribut Pekerjaan Ayah

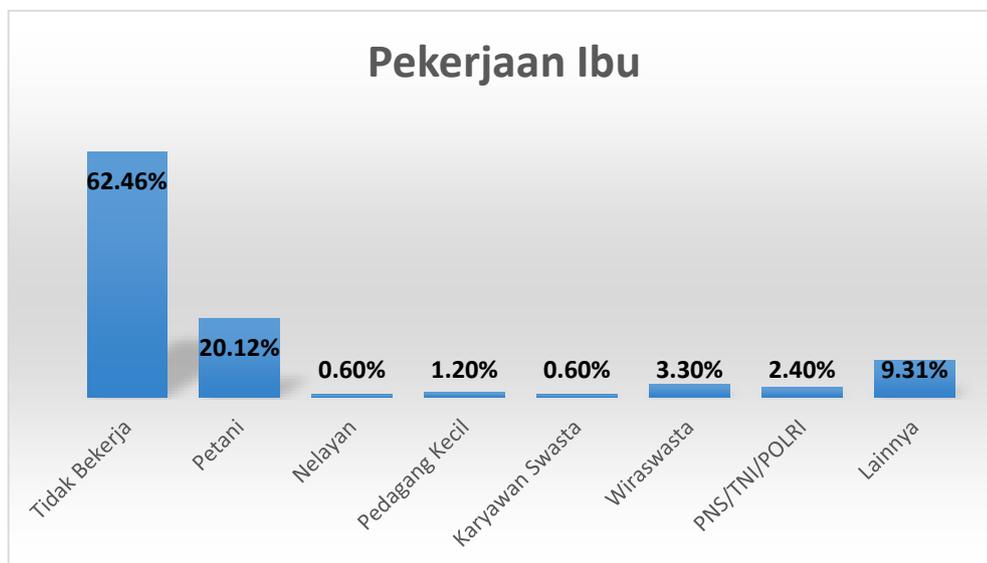
Pada gambar 4.4 terlihat bahwa banyak kategori pada atribut pekerjaan ayah adalah 10. Kategori tersebut meliputi petani, nelayan, peternak, buruh, pedagang kecil, pegawai swasta, wiraswasta, wirausaha, PNS/TNI/POLRI dan lainnya. Siswa dengan pekerjaan ayah sebagai petani memiliki prosentase tertinggi yaitu 59,76% atau 199 siswa berdasarkan data pada lampiran 1.



Gambar 4.5 Prosentase Atribut Pendidikan Ibu

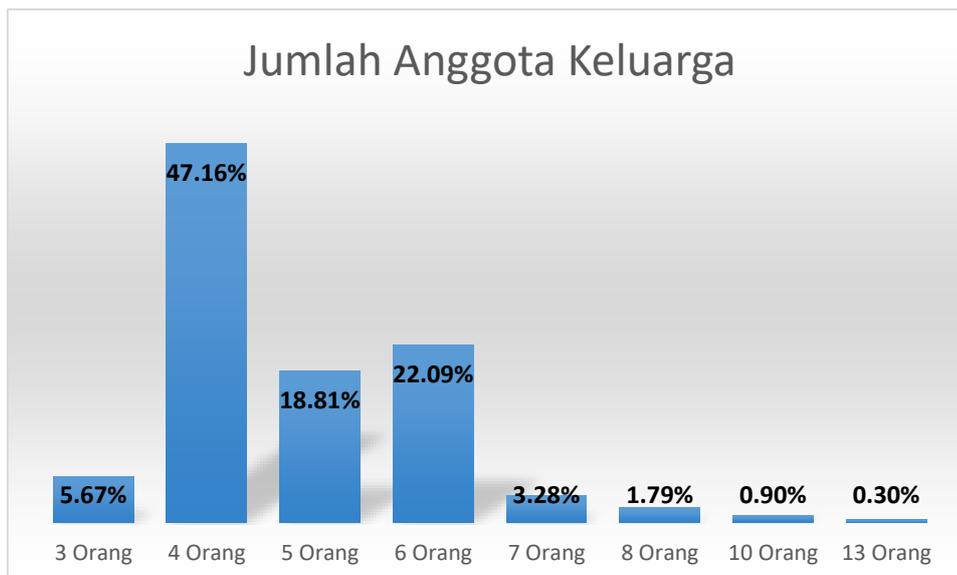
Berdasarkan gambar 4.5 diketahui bahwa terdapat 8 kategori pada atribut Pendidikan Ibu, yang terdiri dari Tidak Sekolah, Putus SD, SD/Sederajat,

SMP/Sederajat, SMA/Sederajat, D2, D3 dan S1. Pada atribut ini diketahui bahwa siswa dengan pendidikan ibu SD/Sederajat mendominasi dengan prosentase sebesar 44,58% atau 148 berdasarkan data pada lampiran 1.



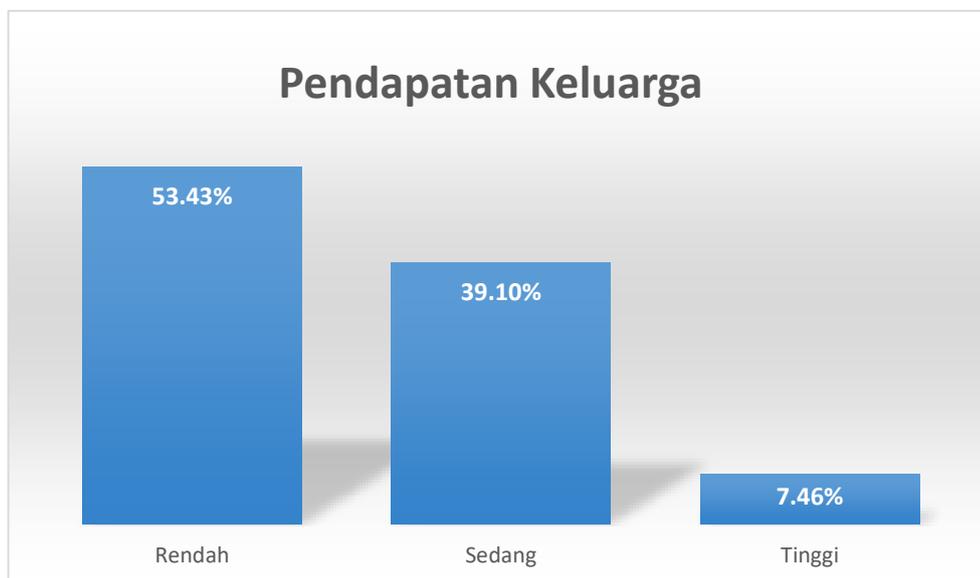
Gambar 4.6 Prosentase Atribut Pekerjaan Ibu

Berdasarkan gambar 4.6 diketahui bahwa dalam atribut Pekerjaan Ibu terdapat 8 kategori yang terdiri dari Tidak Bekerja, Petani, Nelayan, Pedagang Kecil, karyawan swasta, wiraswasta, PNS/TNI/POLRI dan lainnya. Pekerjaan ibu dari siswa yang paling banyak adalah tidak bekerja dengan prosentase sebesar 62,46% atau 208 siswa yang ibunya tidak bekerja menurut data pada lampiran 1.



Gambar 4.7 Prosentase Atribut Jumlah Anggota Keluarga

Berdasarkan gambar 4.7 diketahui bahwa terdapat 8 kategori jumlah anggota keluarga dalam satu rumah. Kebanyakan siswa memiliki 4 anggota keluarga dalam satu rumah. Besar prosentase pada siswa dengan anggota keluarga sebanyak 4 orang adalah 47,16% atau 158 berdasarkan data pada lampiran 1.



Gambar 4.8 Prosentase Atribut Pendapatan Keluarga

Berdasarkan pada gambar 4.8 diketahui bahwa pendapatan keluarga siswa dalam satu bulan dibagi menjadi 3 kategori yaitu, Rendah, Sedang dan Tinggi dimana kebanyakan siswa pada sekolah tersebut berada dalam kategori

pendapatan rendah. Besar prosentase pada siswa dengan pendapatan keluarga rendah adalah 53,43% atau 179 siswa. Pada atribut ini siswa yang memilih untuk putus sekolah seluruhnya adalah mereka yang pendapatan keluarganya rendah.

Data tersebut nantinya akan melalui serangkaian uji, untuk membentuk *decision tree* agar dapat menghasilkan sebuah informasi baru. Pada tahap awal data akan melalui tahap *cross validation*. Dimana pada tahap ini data akan diuji sebanyak k -kali untuk membagi data menjadi data *testing* dan data *training*. Hal ini dilakukan dengan tujuan untuk mengevaluasi kinerja dari algoritma *decision tree* dalam mengklasifikasikan data tersebut. Kemudian data akan di-*preprocessing* untuk menghilangkan data atau atribut yang dianggap dapat mengganggu jalannya penelitian, serta menyesuaikan data sesuai dengan bentuk data yang dibutuhkan dalam penelitian.

Pada tahap selanjutnya untuk mengklasifikasikan data maka, atribut-atribut yang ada akan disusun menjadi struktur pohon pada *decision tree* yang dioptimalkan dengan algoritma *ID3*. Pada tahap akhir kinerja dari algoritma *ID3* dalam mengklasifikasikan siswa putus sekolah dengan *decision tree* akan dinilai dengan menggunakan uji *confusion matrik* untuk dapat mengetahui nilai *accuracy*, *precision* serta *recall* dari algoritma yang digunakan terhadap data yang akan diteliti.

4.2 Cross Validation

Pada tahap ini untuk menentukan banyak presentase dari data *training* maupun data *testing* akan dilakukan 5 uji coba kombinasi presentase, yaitu: (90%,10%), (85%,15%), (80%,20%), (75%,25%), dan (70%,30%). Uji kombinasi

presentase pada *cross validation* ini akan dilakukan dengan menggunakan *software rapidminer*. Pada tahap awal data yang telah di *import* pada *software rapidminer* akan dihubungkan ke dalam kotak *cross validation* sebagaimana terlampir pada lampiran 3. Kemudian pada langkah selanjutnya akan diisikan banyak presentase dari data *testing* yang dibutuhkan ke dalam kotak *number of fold* pada *software rapidminer* sebagaimana terlampir pada lampiran 4.

Data *training* nantinya akan digunakan sebagai penyusun dari pohon klasifikasi *decision tree*, sedangkan data *testing* akan digunakan sebagai alat pengukur kelayakan model dalam mengklasifikasikan data. Kombinasi yang terpilih adalah kombinasi yang memiliki nilai akurasi tertinggi. Hasil dari program *rapidminer* tersebut nantinya akan berbentuk tabel *confusion matrix* dengan menampilkan nilai *accuracy*, *precision*, dan *recall* dimana *output* dari kelima kombinasi data *testing* dan data *training* tersebut terlampir pada lampiran 4. Banyak prosentase pada data *testing* dan data *training* beserta nilai *accuracy*-nya terangkum pada tabel 4.1.

Tabel 4.1 Nilai *Accuracy* dari Kombinasi Data *Testing* dan Data *Training*

No	Prosentase Data <i>Testing</i>	Prosentase Data <i>Training</i>	<i>Accuracy</i>
1.	10%	90%	98,48%
2.	15%	85%	96,96%
3.	20%	80%	98,18%
4.	25%	75%	98,48%
5.	30%	70%	98,18%

Berdasarkan Tabel 4.1 dapat dilihat bahwa, dari 5 kombinasi presentase data *training* dan data *testing*, diketahui bahwa kombinasi 10% data *testing* dan 90% data *training*, serta 25% data *testing* dan 75% data *training* menghasilkan nilai akurasi terbaik, yaitu 98,48%. Menurut Indrayanti, dkk (2017) menyebutkan bahwa pada penelitian-penelitian sebelumnya biasa menggunakan 10 *folds*

dikarenakan menghasilkan performa algoritma yang relatif stabil. Sehingga kombinasi yang digunakan oleh peneliti adalah 90% data *training* dan 10% data *testing*.

4.3 Data Selection

Pada tahap ini data yang dianggap tidak membantu jalannya penelitian akan dibersihkan. Atribut yang dianggap dapat mendukung penelitian akan dipilih sedangkan atribut yang tidak membantu akan dibersihkan. Adapun tabel atribut yang digunakan dan tidak digunakan dalam dalam penelitian ini terdapat pada tabel 4.2.

Tabel 4.2 Proses *Data Selection*

Atribut	Keterangan
Tahun	No
Nama	No
NIPD	No
Jenis Kelamin	Yes
NISN	No
Nama Ayah	No
Pendidikan Ayah	Yes
Pekerjaan Ayah	Yes
Nama Ibu	No
Pendidikan Ibu	Yes
Pekerjaan Ibu	Yes
Anggota Keluarga	Yes
Pendapatan Keluarga	Yes

Berdasarkan tabel 4.2 dapat kita ketahui bahwa atribut Tahun, Nama, NIPD, NISN, Nama Ayah dan Nama Ibu tidak digunakan dalam penelitian ini. Hal itu dilakukan dikarenakan data tersebut bersifat pribadi serta dalam penelitian terdahulu tentang putus sekolah tidak ada peneliti yang mencantumkan keempat atribut tersebut.

Atribut pertama yang dapat mendorong siswa untuk melanjutkan atau putus sekolah adalah jenis kelamin. Banyak siswa yang berjenis kelamin laki-laki adalah 188 dimana 2 diantaranya memilih untuk putus sekolah sedangkan dari 147 siswa berjenis kelamin perempuan 5 diantaranya memilih untuk putus sekolah. Berdasarkan landasan teori 2.1.7 jenis kelamin menjadi salah satu faktor yang mendorong siswa dalam mengambil keputusan. Namun dalam mengambil keputusan ada faktor lain yang juga dapat mempengaruhi siswa. Faktor selanjutnya yang dapat mempengaruhi siswa dalam mengambil keputusan adalah pendidikan ayah.

Berdasarkan 9 kategori pada atribut pendidikan ayah terdapat 2 dari 13 siswa yang ayahnya tidak sekolah, 1 dari 48 siswa yang ayahnya tidak lulus SD serta 3 dari 125 siswa yang ayahnya hanya tamatan SD/ sederajat memilih untuk putus sekolah. Sedangkan siswa yang ayahnya minimal lulusan SD/ sederajat memilih untuk melanjutkan sekolah. Pola ini menggambarkan bahwa pendidikan ayah dapat menjadi salah satu atribut yang berpengaruh bagi siswa dalam mengambil keputusan. Semakin tinggi tingkat pendidikan ayah dapat mendorong minat siswa untuk melanjutkan sekolah.

Pada penelitian sebelumnya atribut pekerjaan ayah menjadi salah satu faktor yang mempengaruhi siswa dalam mengambil keputusan. Diketahui dari data pada lampiran 1 bahwa terdapat 6 dari 199 siswa yang ayahnya berprofesi sebagai petani dan 1 dari 4 siswa yang ayahnya berprofesi sebagai nelayan memutuskan untuk putus sekolah. Profesi dapat mempengaruhi pendapatan seseorang, dan pendapatan dalam keluarga dapat menjadi penentu kemampuan

orang tua dalam menyekolahkan anaknya. Faktor pekerjaan ayah juga dapat menjadi faktor penyebab putus sekolah apabila didukung dengan faktor lainnya.

Selanjutnya terdapat atribut pendidikan berdasarkan lampiran 1 ditemukan terdapat 4 dari 49 siswa dengan pendidikan ibu yang tidak menyelesaikan SD dan 2 dari 12 siswa yang ibunya tidak bersekolah memilih untuk putus sekolah. Sedangkan bagi siswa yang pendidikan terakhir ibunya minimal lulus SD memilih untuk melanjutkan sekolah. Melihat pola tersebut dapat diketahui bahwa semakin tinggi jenjang pendidikan ibu dapat menjadi memotivasi anak untuk melanjutkan sekolah. Mengingat fakta bahwa pada umumnya seorang ibu lebih banyak berinteraksi dengan anak ketimbang anggota keluarga lainnya. Faktor selanjutnya yang menjadi alasan siswa putus sekolah adalah pekerjaan ayah.

Atribut selanjutnya adalah pekerjaan ibu, mencari nafkah memang bukan kewajiban seorang ibu. Namun terkadang karena tuntutan ekonomi membuat seorang ibu harus ikut mencari nafkah. Pada data ini terdapat 8 kategori dalam atribut pekerjaan ibu. Dalam data pada lampiran 1 ditemukan 3 diantara 208 siswa yang ibunya tidak bekerja dan 3 dari 67 siswa yang ibunya bekerja sebagai petani memilih untuk tidak melanjutkan sekolah. Pekerjaan ibu juga dapat mempengaruhi pendapatan keluarga, sedangkan pendapatan keluarga dapat mempengaruhi kemampuan orang tua dalam menyekolahkan anaknya.

Atribut selanjutnya yang mempengaruhi keputusan siswa adalah banyak anggota keluarga dalam satu rumah. Semakin banyak anggota keluarga, maka akan semakin besar kebutuhan yang harus dikeluarkan oleh keluarga tersebut. Oleh sebab itu faktor ini dapat mendorong siswa untuk putus sekolah. Pada data yang terlampir pada lampiran 1 diketahui bahwa banyak anggota keluarga siswa

yang memilih putus sekolah adalah, 2 siswa yang dalam keluarganya beranggota 5 orang, 2 siswa dengan jumlah anggota keluarga 6 orang, 1 siswa dengan anggota keluarga 7 orang dan 1 siswa dengan jumlah anggota keluarga 10 orang. Namun, atribut jumlah anggota keluarga bisa mendorong siswa untuk putus sekolah apabila ada faktor lain yang juga berpengaruh.

Atribut terakhir yang dinilai dapat mempengaruhi siswa dalam mengambil keputusan adalah pendapatan keluarga. Pendapatan keluarga dibagi menjadi 3 kategori yaitu rendah, sedang dan tinggi. Namun berdasarkan data yang ada diketahui bahwa keseluruhan siswa yang memilih untuk putus sekolah adalah mereka yang berkecenderungan memiliki pendapatan rendah, namun tidak sedikit keluarga yang memiliki pendapatan rendah tetapi tetap menyekolahkan anaknya.

4.4 Pre-processing (Data Cleaning)

Proses ini dilakukan dengan tujuan menghilangkan data yang dianggap tidak membantu jalannya penelitian, ataupun data yang dianggap bernilai kosong atau *missing value*. Banyak data yang didapatkan sebelumnya sejumlah 335 data. Namun karena ditemukan 6 dataset yang bernilai kosong atau *missing value* pada data tersebut sehingga tersisa 329 data yang dapat digunakan dalam penelitian (Lampiran 2). Peneliti menggunakan seluruh sampel dikarenakan mempertimbangkan fokus penelitian ini untuk mengetahui atribut apa saja yang mempengaruhi siswa putus sekolah.

4.5 Transformasi Data

Tahap ini merupakan tahap akhir pada *data mining* sebelum data dapat diolah dengan algoritma yang ditentukan. Kegiatan transformasi data dilakukan dengan tujuan untuk mentransformasi agar data yang digunakan dapat dipetakan secara kompleks. Namun dikarenakan pada penelitian ini data yang didapat sudah sesuai dengan data yang dibutuhkan dalam penelitian sehingga data tersebut tidak perlu ditransformasikan lagi.

4.6 Classifier

Pada tahap ini peneliti akan memilih teknik, metode dan algoritma yang akan digunakan untuk membentuk model klasifikasi. Teknik yang digunakan dalam penelitian ini adalah klasifikasi *data mining decision tree* dengan algoritma *ID3*. Untuk membentuk *decision tree* maka harus menghitung nilai *entropy* total maupun nilai *entropy* dari setiap atribut dengan menggunakan persamaan (2.1) sebagai berikut:

$$\begin{aligned} Entropy\ Total &= (-p_+ \log_2 p_+) - (-p_- \log_2 p_-) \\ &= \left(-\frac{323}{329} \log_2 \left(\frac{323}{329}\right)\right) - \left(-\frac{6}{329} \log_2 \left(\frac{6}{329}\right)\right) = 0,131424 \end{aligned}$$

Atribut Jenis Kelamin

a. Laki-laki

$$\begin{aligned} Entropy(S_i) &= (-p_+ \log_2 p_+) - (-p_- \log_2 p_-) \\ &= \left(-\frac{186}{187} \log_2 \left(\frac{186}{187}\right)\right) - \left(-\frac{1}{187} \log_2 \left(\frac{1}{187}\right)\right) \\ &= 0.048052 \end{aligned}$$

b. Perempuan

$$\begin{aligned} Entropy(S_i) &= (-p_+ \log_2 p_+) - (-p_- \log_2 p_-) \\ &= \left(-\frac{137}{142} \log_2 \left(\frac{137}{142}\right)\right) - \left(-\frac{5}{142} \log_2 \left(\frac{5}{142}\right)\right) = 0,219888 \end{aligned}$$

Untuk mendapatkan nilai *entropy* pada setiap atribut maka, dilakukan perhitungan yang sama pada setiap atribut dengan menggunakan persamaan 2.1. Kemudian berdasarkan nilai *entropy* peneliti dapat menghitung nilai *information gain*. Berikut adalah perhitungan *information gain* secara manual dengan menggunakan persamaan 2.3 sebagai berikut:

$$\begin{aligned} Gain(S, A) &= Entropy(S) - \sum_{i=2}^n \frac{|S_i|}{|S|} Entropy(S_i) \\ &= 0,131424 - \left(\frac{187}{329} \times 0,048052 \right) - \left(\frac{137}{329} \times 0,219888 \right) = 0,009206 \end{aligned}$$

Untuk menentukan *root node* pada *decision tree* maka perlu menghitung nilai *information gain* seluruh atribut dengan menggunakan persamaan yang sama. Kemudian atribut dengan nilai *information gain* tertinggi akan terpilih menjadi *root node*. Adapun tabel hasil perhitungan *entropy* dan *information gain* dengan adalah sebagai berikut:

Tabel 4.3 Tabel Hasil Perhitungan Menentukan *Root Node*

No	Atribut	Kelas	Jumlah (S)	Melanjutkan (Si)	Putus Sekolah (Si)	Entropy	Information Gain
			329	323	6	0,131424	
1	Jenis Kelamin						
		Laki-laki	187	186	1	0,048052	0,009206
		Perempuan	142	137	5	0,219888	
2	Pendidikan Ayah						
		Tidak Sekolah	12	10	2	0,650022	0,024339
		Putus SD	48	47	1	0,146094	
		SD/ Sederajat	125	122	3	0,163346	
		SMP/Sederajat	76	76	0	0	
		SMA/Sederajat	42	42	0	0	
		D2	2	2	0	0	
		D3	4	4	0	0	
		D4	1	1	0	0	
		S1	19	19	0	0	
3	Pekerjaan Ayah						
		Nelayan	3	3	0	0	0,013657
		Petani	197	191	6	0,196678	
		Peternak	1	1	0	0	

	Buruh	7	7	0	0	
	Karyawan Swasta	8	8	0	0	
	Pedagang Kecil	7	7	0	0	
	Wiraswasta	78	78	0	0	
	Wirausaha	1	1	0	0	
	PNS/TNI/POLRI	23	23	0	0	
	Lainnya	4	4	0	0	
4	Pendidikan Ibu					
	Tidak Sekolah	12	10	2	0,650022	0,046963
	Putus SD	49	45	4	0,407906	
	SD/ Sederajat	147	147	0	0	
	SMP/Sederajat	80	80	0	0	
	SMA/Sederajat	24	24	0	0	
	D2	3	3	0	0	
	D3	3	3	0	0	
	S1	11	11	0	0	
5	Pekerjaan Ibu					
	Tidak Bekerja	208	205	3	0,108861	0,009291
	Nelayan	4	1	0	0	
	Petani	65	62	3	0,269828	
	Pedagang Kecil	4	4	0	0	
	Karyawan Swasta	2	2	0	0	
	Wiraswasta	11	11	0	0	
	PNS/TNI/POLRI	8	8	0	0	
	Lainnya	30	30	0	0	
6	Jumlah Anggota Keluarga					
	3	17	17	0	0	0,029395
	4	156	156	0	0	
	5	63	61	2	0,203074	
	6	72	70	2	0,183122	
	7	11	10	1	0,439497	
	8	6	6	0	0	
	10	3	2	1	0,918296	
	13	1	1	0	0	
7	Pendapatan Keluarga					
	Rendah	173	167	6	0,217354	0,017132
	Sedang	131	131	0	0	
	Tinggi	25	25	0	0	

Berdasarkan tabel 4.3 peneliti dapat menentukan Pendidikan Ibu sebagai *root node*. Karena pendidikan ibu memiliki nilai *information gain* tertinggi setelah

dilakukan perhitungan dengan menggunakan algoritma *ID3* dan akan menjadi *root node* dari *decision tree*. Hal ini sejalan dengan fakta bahwa pendidikan dasar pertama yang diterima oleh seseorang adalah dari kedua orang tua, terutama ibu. Pada umumnya interaksi yang paling banyak dilakukan seorang anak didalam keluarga adalah dengan ibu. Itulah kenapa perempuan harus berpendidikan, agar dapat mendidik anaknya menjadi pribadi yang cerdas dan berintelektual, karena tingkat pendidikan ibu yang tinggi dapat berdampak pada pola asuh anak (Kusbudiyanto & Munandar, 2019).

Berdasarkan tabel 4.3 dapat kita lihat bahwa pada kategori Tidak Sekolah dan Putus SD ditemukan siswa yang memilih untuk putus sekolah, namun pada siswa dengan pendidikan ibu minimal SD/Sederajat memilih untuk melanjutkan sekolah. Mengamati pola pada tabel 4.3 menggambarkan bahwa anak yang ibunya minimal lulusan SD/Sederajat akan melanjutkan sekolah kejenjang yang lebih tinggi. Sedangkan terdapat beberapa anak yang putus sekolah terdapat pada label pendidikan ibu tidak lulus SD atau bahkan tidak sekolah. Pola ini menunjukkan bahwa semakin rendah pendidikan ibu akan semakin terbuka peluang siswa untuk putus sekolah. Oleh sebab itu semakin tinggi pendidikan ibu akan berdampak semakin banyak pendidikan dan keteladanan yang ibu berikan kepada anak untuk memotivasi anak untuk melanjutkan sekolah. Selain itu kemampuan yang dimiliki anak dalam menerima pelajaran juga bergantung pada pola asuh orangtua dirumah dalam memberikan motivasi kepada anak (Richard, 2016).

Karena pada atribut pendidikan ibu Tidak Sekolah dan Putus SD terdapat siswa yang memilih untuk melanjutkan dan putus sekolah, sehingga dapat kita telusuri faktor pendukung lainnya dengan menentukan *internal node*. Berikut

adalah tabel hasil perhitungan untuk mencari *internal node* dari atribut Pendidikan ibu-Tidak Sekolah:

Tabel 4.4 Tabel Hasil Perhitungan Menentukan *Internal Node*

No	Atribut	Kelas	Jumlah (S)	Melanjutkan (Si)	Putus Sekolah (Si)	Entropy	Information Gain
	Pendidikan Ibu-Tidak Sekolah		12	10	2	0,650022	
1	Jenis Kelamin						
		Laki-laki	7	6	1	0,591673	0,004077
		Perempuan	5	4	1	0,721928	
2	Pendidikan Ayah						
		Tidak Sekolah	5	5	0	0	0,24546
		SD/ Sederajat	5	3	2	0,970951	
		SMP/Sederajat	2	2	0	0	
3	Pekerjaan Ayah						
		Nelayan	1	1	0	0	0,10917
		Petani	8	6	2	0,811278	
		Peternak	0	0	0	0	
		Wiraswasta	3	3	0	0	
5	Pekerjaan Ibu						
		Tidak Bekerja	8	6	2	0,811278	0,10917
		Petani	4	4	0	0	
6	Jumlah Anggota Keluarga						
		4	4	4	0	0	0,420448
		5	3	3	0	0	
		6	3	2	1	0,918296	
		10	1	0	1	0	
		13	1	1	0	0	
7	Pendapatan Keluarga						
		Rendah	6	4	2	0,918296	0,190875
		Sedang	6	6	0	0	

Berdasarkan tabel 4.4 atribut “Jumlah Anggota Keluarga” menjadi *internal node* selanjutnya dikarenakan memiliki nilai *information gain* tertinggi yaitu 0,420448. Bagi siswa dengan pendidikan ibu tidak sekolah dan jumlah anggota keluarga sebanyak 4 dan 5 orang lebih memilih untuk menyekolahkan anaknya. Hal ini dapat kita asumsikan dalam satu keluarga tersebut terdapat 2 orang tua dan yang lainnya adalah anak sehingga beban biaya yang harus

dikeluarkan orang tua masih bisa ditangani. Namun bagi anak yang ibunya tidak sekolah dengan jumlah anggota keluarga lebih dari 5 memilih putus sekolah dikarenakan tanggungan biaya yang dipikul oleh orangtuanya juga akan lebih banyak.

Pola ini sejalan dengan penelitian sebelumnya yang dilakukan oleh Kusbudiyanto dan Adis (2019) juga mendapati banyak anggota keluarga menjadi salah satu faktor penentu dalam keputusan anak untuk putus sekolah. Dalam penelitian tersebut terdapat 60% siswa yang putus sekolah adalah mereka yang anggota keluarganya lebih dari 4 orang. Hal ini menggambarkan bahwa bertambahnya jumlah anggota keluarga akan berdampak pada anak untuk memilih putus sekolah.

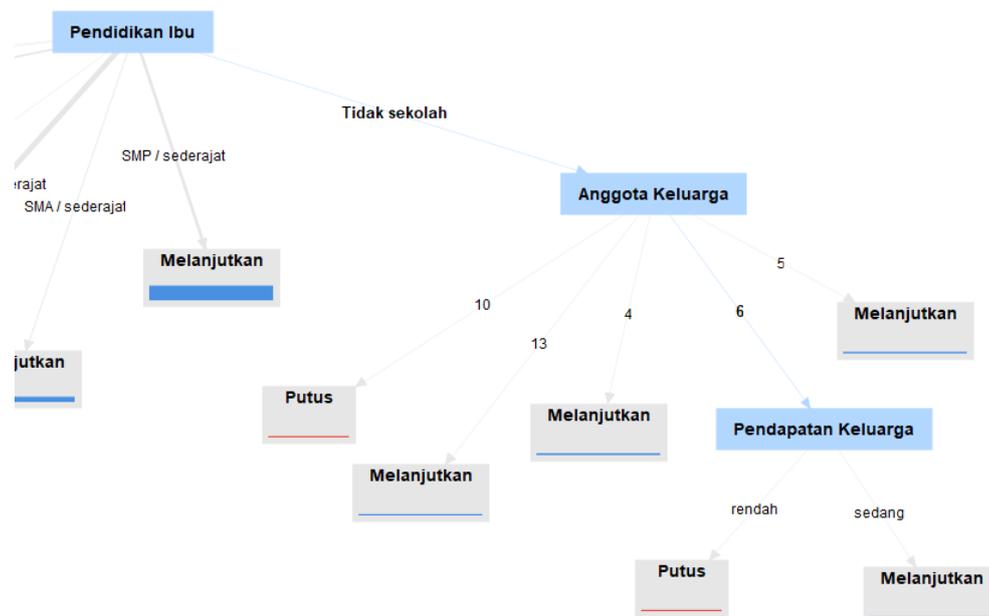
Dikarenakan pada jumlah anggota keluarga sebanyak 6 orang didapati ada siswa yang memilih untuk melanjutkan dan putus sekolah sehingga, pada tahap selanjutnya peneliti akan menghitung kembali nilai *information gain* untuk menentukan *leaf node*. Berikut adalah tabel perhitungan untuk menentukan *leaf node*:

Tabel 4.5 Tabel Hasil Perhitungan menentukan *Leaf Node*

No	Atribut	Kelas	Jumlah (S)	Melanjutkan (Si)	Putus Sekolah (Si)	Entropy	Information Gain
		Pendidikan Ibu-Tidak Sekolah, Anggota Keluarga-6	3	2	1	0,918296	
1	Jenis Kelamin						
		Laki-laki	2	1	1	1	0,251629
		Perempuan	1	1	0	0	
2	Pendidikan Ayah						
		Tidak Sekolah	1	1	0	0	0,918296
		SD/ Sederajat	1	0	1	0	
		SMP/Sederajat	1	1	0	0	
3	Pekerjaan Ayah						
		Petani	3	2	1	0,918296	0

		Peternak	0	0	0	0	
5	Pekerjaan Ibu						
		Tidak Bekerja	2	1	1	0	0,918296
		Petani	1	1	0	0	
7	Pendapatan Keluarga						
		Rendah	1	0	1	0	0,918296
		Sedang	2	2	0	0	

Berdasarkan hasil perhitungan pada tabel 4.4 didapati bahwa atribut pendapatan keluarga menjadi *leaf node* pada *decision tree* dikarenakan memiliki nilai *information gain* yang paling tinggi. Sehingga bagi siswa yang ibunya tidak sekolah dengan jumlah anggota keluarga sebanyak 6 orang, apabila pendapatan keluarganya rendah akan mendorong siswa untuk putus sekolah. Sedangkan apabila tingkat pendapatan keluarganya relatif sedang akan mendorong anaknya untuk melanjutkan sekolah. Adapun bentuk *output decision tree* berdasarkan *software Rapidminer* seperti pada gambar 4.9



Gambar 4.9 Potongan *Output Decision Tree* dengan *RapidMiner*

Berdasarkan gambar 4.9 dapat diketahui bahwa Pendidikan Ibu menjadi *root node* pada *decision tree*. Atribut tersebut terpilih dikarenakan memiliki nilai

information gain tertinggi. Namun hasil ini juga sejalan dengan fakta bahwa pada umumnya keluarga yang paling dekat dan paling banyak ber-interaksi dengan anak adalah ibu, sehingga dengan pendidikan ibu yang memadai dapat mendorong anak untuk sekolah lebih tinggi lagi. Berdasarkan pada gambar 4.9 atribut Pendidikan Ibu dengan kategori tidak sekolah didapatkan bahwa *internal node* selanjutnya adalah jumlah anggota keluarga. Dimana keseluruhan siswa yang ibunya tidak sekolah dengan jumlah anggota keluarga sebanyak 13, 4 dan 5 memilih untuk melanjutkan sekolah, sedangkan siswa dengan jumlah keluarga 10 memilih untuk putus sekolah. Namun siswa dengan jumlah keluarga 6 terdapat siswa yang melanjutkan dan putus sekolah. Keputusan siswa selanjutnya dipengaruhi oleh Atribut Pendapatan Keluarga. Apabila pendapatan keluarganya rendah maka siswa tersebut memilih untuk putus sekolah, sedangkan apabila Pendapatan keluarganya Sedang memilih untuk melanjutkan sekolah. Sedangkan untuk menentukan *internal node* dan *leaf node* pada atribut pendidikan ibu-Putus SD dengan cara dan langkah yang sama seperti pada tabel sebelumnya dengan menggunakan *Ms.Excel* yang terlampir pada lampiran 6.

Penyusunan *node* pada *decision tree* juga dilakukan dengan menggunakan *software RapidMiner* berikut adalah uji model dengan menggunakan *software RapidMiner* seperti pada lampiran 3. Pada proses awal dataset akan di *import*, kemudian data tersebut akan dihubungkan dengan *cross validation* untuk membagi dataset menjadi *data testing* dan *data training*. Setelah di klik dua kali pada menu *cross validation* maka akan muncul menu dan hal pertama yang perlu dilakukan adalah memasukkan menu *ID3* pada bagian *training*. Kemudian masukkan *apply model* dan *performance* pada bagian testing untuk melihat

akurasi dari model *decision tree* yang diaplikasikan dengan algoritma *ID3*. Seperti pada lampiran 3. Setelah itu maka *software RapidMiner* akan memunculkan nilai *accuracy, precision, recall*, serta akan memunculkan *node-node* dari *decision tree* sebagaimana terlampir pada lampiran 7.

4.7 Evaluasi

Nilai akurasi digunakan sebagai alat ukur untuk mengetahui kinerja dari algoritma *ID3* dalam membangun *node-node* pada *decision tree* saat mengklasifikasikan faktor yang mempengaruhi putus sekolah. Dalam proses pengujian ini teknik *k-fold cross validation* sebanyak 10 *fold* dipilih dengan tujuan untuk menghasilkan nilai akurasi yang maksimal. Adapun tabel *confusion matrix* dan nilai *accuracy, precision, dan recall* yang didapatkan berdasarkan perhitungan *software RapidMiner* adalah terlampir pada lampiran 8.

Tabel 4.6 Tabel Hasil Perhitungan *Confusion Matrix*

	True Melanjutkan	True Putus
Pred. Melanjutkan	321	3
Pred. Putus	2	3

Tabel 4.6 merupakan hasil dari perhitungan *confusion matrix* dari data *training* dengan menggunakan *software RapidMiner*. Terdapat 321 *record* data dengan keterangan *true positive* (TP), 3 *record* data dengan keterangan *false negative* (FN), 2 *record* data dengan keterangan *false positive* (FP), dan 3 *record* data dengan keterangan *true negative* (TN). Berdasarkan tabel *confusion matrix* di atas kita dapat menghitung nilai *accuracy, precision, dan recall*. Sedangkan nilai dari *accuracy, precision, dan recall* dengan persamaan (2.4), (2.5), dan (2.6) adalah sebagai berikut:

$$1. \text{ Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{321+3}{321+3+2+3} = \frac{324}{329} \approx 0,9848 = 98,48\%$$

$$2. \text{ Precision} = \frac{TP}{TP+FN} = \frac{321}{321+3} = \frac{321}{324} \approx 0,9907 = 99,07\%$$

$$3. \text{ Recall} = \frac{TP}{TP+FP} = \frac{321}{321+2} = \frac{321}{323} \approx 0,9938 = 99,38\%$$

Berdasarkan hasil perhitungan di atas maka, Algoritma *ID3* dianggap mampu mengoptimalkan *decision tree* berdasarkan uji model *confusion matrix* dengan nilai *accuracy* 98,48%, *precision* 99,07%, dan *recall* sebesar 99,38%.

4.8 Kajian Keislaman dengan Hasil Penelitian

Berdasarkan nilai dari *confusion matrix* yang cukup tinggi membuktikan bahwa Algoritma *ID3* dianggap mampu mengoptimalkan *decision tree* dalam mengklasifikasikan faktor penyebab putus sekolah pada salah satu Sekolah Menengah Pertama di Kabupaten Sumenep. Dalam penelitian ini atribut pendidikan ibu menjadi *root node* dalam *decision tree*. Al-Qur'an juga memerintahkan untuk orang tua membimbing anaknya kejalan yang benar seperti dalam Al-Quran surat At-Tahrim ayat 6:

يَا أَيُّهَا الَّذِينَ ءَامَنُوا قُوا أَنفُسَكُمْ وَأَهْلِيكُمْ نَارًا وَقُودُهَا النَّاسُ وَالْحِجَارَةُ عَلَيْهَا مَلَائِكَةٌ غِلَاظٌ شِدَادٌ لَّا يَعْصُونَ
 ٱللَّهَ مَا أَمَرَهُمْ وَيَفْعَلُونَ مَا يُؤْمَرُونَ ٦

Artinya: “Wahai orang-orang yang beriman! Peliharalah dirimu dan keluargamu dari api neraka yang bahan bakarnya adalah manusia dan batu; Penjaganya malaikat-malaikat yang kasar, dan keras, yang tidak durhaka kepada Allah terhadap apa yang dia perintahkan kepada mereka dan selalu mengerjakan apa yang diperintahkan . (Q.S. At-Tahrim: 6)”

Dalam ayat tersebut Allah SWT memerintahkan dengan jelas bagi seluruh umat muslim untuk menjaga keluarganya, dan para orang tua untuk membimbing anaknya agar berjalan di jalan kebenaran. Sebagai orang tua wajib untuk mendidik anaknya, serta mengingatkan anaknya apabila berbuat kesalahan (Ar-Rifa'i, 2000). Dari penjelasan ayat tersebut diterangkan bahwa wajib

hukumnya bagi orang tua untuk mendidik anaknya kearah kebaikan serta keridhoan Allah SWT. Hal ini sejalan dengan tingkat pendidikan ibu yang menjadi atribut paling berpengaruh dalam keputusan siswa dalam mengambil keputusan untuk lanjut atau putus sekolah.

Oleh sebab itu, sebagai orang tua juga sudah selayaknya untuk lebih banyak belajar karena pada dasarnya anak adalah harta yang dititipkan oleh Allah SWT. Hal ini dijelaskan dalam firman-Nya sebagai berikut:

أَمْالٌ وَالْبُنُونََ زِينَةُ الْحَيَاةِ الدُّنْيَا وَالْبَاقِيَةُ الصَّالِحَةُ خَيْرٌ عِنْدَ رَبِّكَ ثَوَابٌ وَخَيْرٌ أَمَلٌ (سورة الكهف: ٤٦)

Artinya: *“Harta dan anak-anak adalah perhiasan kehidupan dunia tetapi amalan-amalan yang kekal lagi saleh adalah lebih baik pahalanya disisi Tuhanmu serta lebih baik menjadi harapan . (Q.S. Al-Kahfi:46)”*

Berdasarkan ayat tersebut dijelaskan bahwa anak hanyalah sebuah titipan di dunia yang harus dijaga, dirawat serta dididik. Sebab Allah SWT menjelaskan bahwa banyak anak dan harta sering menjadi kebanggaan manusia sehingga membuatnya menjadi takabur. Namun Allah SWT menegaskan bahwa anak hanyalah perhiasan dunia, namun anak juga dapat menjadi penyelamat orang tuanya jika mereka didik dengan baik sehingga menjadi anak yang berwawasan luas serta berintelektual tinggi (Al-Sheikh, 2003). Maka, sebaiknya sebagai orang tua dapat mendukung serta mengarahkan anaknya untuk bisa menjadi pribadi yang lebih baik.

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan rumusan masalah yang terlampir pada bab 1 serta penelitian yang terlampir pada bab 4 maka dapat disimpulkan:

1. Hasil kinerja algoritma *ID3* dalam mengoptimalkan *decision tree* untuk mengklasifikasikan faktor yang melatar belakangi siswa putus sekolah memiliki nilai *accuracy* sebesar 98,48%, *precision* sebesar 99,07%, dan *recall* sebesar 99,38%. Sehingga kinerja dari model algoritma *ID3* dinilai sangat baik untuk mengoptimalkan *decision tree* dalam mengklasifikasikan faktor yang melatar belakangi siswa putus sekolah.
2. Atribut yang paling berpengaruh terhadap siswa dalam mengambil keputusan untuk putus ataupun lanjut sekolah setelah diklasifikasikan dengan metode *decision tree* yang dibangun oleh algoritma *ID3* adalah atribut pendidikan ibu. Dikarenakan atribut ini memiliki nilai *information gain* tertinggi, sehingga pendidikan ibu menjadi *root node* sekaligus sebagai atribut paling berpengaruh bagi siswa dalam mengambil keputusan. Pola ini sejalan dengan kewajiban orangtua terutama ibu yang pada umumnya lebih banyak ber-interaksi dengan anak untuk memdidik dan mengarahkan anaknya ke jalan kebenaran. Karena sejatinya anak adalah titipan yang harus dijaga dan diarahkan. Hal ini juga sesuai dengan Surah At-Tahrim ayat 6 yang memerintahkan untuk menjaga keluarganya dari api neraka, begitu pula dengan kewajiban orang tua untuk memberikan pendidikan kepada anaknya, agar dapat menjadi pribadi

yang berwawasan luas serta ber-intelektual tinggi sehingga jauh dari api neraka.

5.2 Saran untuk Penelitian Lanjutan

Berikut adalah hal-hal yang perlu ditambahkan pada penelitian selanjutnya, sehingga pada penelitian selanjutnya dapat dilakukan lebih komprehensif:

1. Dalam penelitian ini peneliti berfokus pada faktor eksternal yang mempengaruhi siswa-siswi putus sekolah, diharapkan pada penelitian selanjutnya peneliti dapat membandingkan antara faktor eksternal maupun faktor internal dalam memprediksi faktor apa saja yang mempengaruhi siswa-siswi memilih untuk meninggalkan bangku sekolah.
2. Pada penelitian ini peneliti hanya berfokus pada satu instansi diharapkan pada penelitian selanjutnya peneliti dapat memperluas objek penelitian.
3. Pada penelitian ini *decision tree* dioptimalkan hanya dengan algoritma *ID3* diharapkan pada penelitian selanjutnya peneliti dapat membandingkan dua atau lebih algoritma dalam mengoptimalkan *decision tree* agar dapat membandingkan algoritma mana yang dianggap paling baik dalam proses pengklasifikasi.

DAFTAR PUSTAKA

- Ajahari. (2018). *Ulumul Qur'an (Ilmu-Ilmu Al-Qur'an)*. Yogyakarta: Aswaja Pressindo.
- Al-Sheikh, D. A. (2003). *Tafsir Ibnu Katsir Jilid 5*. Bogor: Pustaka Imam Asy-Syafi'i.
- Ar-Rifa'i, M. N. (2000). *Ringkasan Tafsir Ibnu Katsir*. Jakarta: Gema Insani.
- Aritonang, M. (2021). Penerapan Algoritma ID3 dalam Prediksi Kebutuhan Pupuk. *Journal of Information System Research (JOSH)*, 2, 247-253.
- Bukhari, U. (2012). *Hadits Tarbawi (Pendidikan dalam Perspektif Islam)*. Jakarta: Anizah.
- Cahyono, S. A. (1998). Karakteristik Sosial Ekonomi yang Mempengaruhi Pendapatan Rumah Tangga Penyadap Getah Pinus di Desa Somagede, Kebumen, Jawa Timur. *Jurnal Universitas Gajah Mada*, 111-122.
- Dewi, N. A., Zukhri, A., & Dunia, I. K. (2014). Analisis Faktor-Faktor Anak Putus Sekolah Usia Pendidikan Dasar di Kecamatan Gerogak Tahun 2012/2013. *ACADEMIA*, 4, 13-22.
- Fatihatul, F., Setiawan, A., & Rosadi, R. (2011). Asosiasi Data Mining dengan Menggunakan Algoritma FP-Growth untuk Market Basket Analisis. *Academia*, 1-8.
- Gunarsa, S. (2002). *Psikologi Perkembangan Anak dan Remaja*. Jakarta: Gunung Mulia.
- Gorunescu, F. (2011). *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer.
- Hafizan, H., & Nadia Putri, A. (2020). Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi Balita di Kabupaten Simalungan. *Kesatria*, 2, 68-72.
- Han, J., & Kamber, M. (2001). *Data Mining Concept And Techniques*. USA: Academic Press.
- Hidayat, D. I. (2014). *Tafsir Salman (Tafsir Ilmiah Atas Juz'Amma)*. Bandung: Mizan Pustaka.
- Hurlock, E.B. (1999). *Child Development Jilid 2, Terjemah Tjandrasa*. Jakarta: Erlangga
- Indrayanti, D, S., & Karomi, M. A. (2017). Optimasi Parameter K pada Algoritma K-Nearest Neighbour untuk Klasifikasi Penyakit Diabetes Mellitus. *Prosiding SNATIF*, 823-829.
- Jollyta, D., Ramadhan, W., & Zarlis, M. (2020). *Konsep Data Mining dan Penerapan*. Yogyakarta: CV BUDI UTAMA.

- Kurniawan, M. Y., & Rosadi, M. E. (2017 Vol.2(1)). Optimasi Decision Tree Menggunakan Particle Swarm Optimization Pada Data Siswa Putus Sekolah. *Jtiulm*, 15-22.
- Medina, E. C., Chunga, C. B., Aguirre, J. A., & Grandón, E. E. (2020, Vol.15). Predictive Model to Reduce the Dropout Rate of University Students in Perú: Bayesian Networks vs. Decision Trees. *Iberian Conference on Information Systems and Technologies (CISTI)*, 117-123.
- Meng, X., Zhang, P., Xu, Y., & Xie, H. (2020). Construction of Decision Tree based on C4.5 Algorithm for Online Voltage Stability Assessment. *Electrical Power and Energy Systems*, 1-8.
- Nasution, M. R., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika*, 4, 212-218.
- Nigtyas, D. A., Wahyudi, M., & N.Nurjannah. (2019, vol.7). Klasifikasi Siswa SMK Berpotensi Putus Sekolah Menggunakan Algoritma Decision Tree, Support Vector Machine dan Naive Bayes. *Jurnal Khatulistiwa Informatika*, 85-90.
- Nofriansyah, D., & Nurcahyo, G. W. (2015). *Algoritma Data Mining dan Pengujian*. Yogyakarta: CV. Budi Utama.
- Nurjannah, & Ahmad, M. R. (2017 vol 4(1)). Kehidupan Remaja Putus Sekolah di Desa Lara Kecamatan Karossa Kabupaten Mamuju Tengah. *Jurnal Sosialisasi Pendidikan Sosiologi*, 95-99.
- Panhalkar, A. R., & Doye, D. D. (2021). Optimization of Decision Trees using Modified African Buffalo Algorithm. *Journal of King Saud University – Computer and Information Sciences*, 1-10.
- Putri, Tiara Rizki Novinda. (2017). Faktor Penyebab Siswa Putus Sekolah di Sekolah Menengah Atas Kabupaten Bantul Tahun Ajaran 2015/2016. *Jurnal Hanata Widya*. 6(8), 70-82.
- Putri, A. D., & Setiawina, N. D. (2013). Pengaruh Umur, Pendidikan, Pekerjaan Terhadap Pendapatan Rumah Tangga miskin di Desa Bebandem. *E-Jurnal EP Unud* 2(4) , 173-180.
- Richard, A. J. (2016). Tesis: Determinan Peluang Putus Sekolah pada Usia 16-18 Tahun di Indonesia (Analisis Data Susenas 2013) . *Program Pascasarjana, Kajian Kependudukan dan Ketenagakerjaan, Universitas Indonesia*, 298-318.
- Safii, M. (2018). Implementasi Data Mining Dengan Metode Pohon Keputusan Algoritma ID3 untuk Menentukan Status Mahasiswa. *Mantik Penusa*, 82-87.
- Shihab, M. Q. (1999). *Membumikan Al-Qur'an*. Bandung: Mizan.
- Shihab, M. Q. (2006). *Wawasan Al-Qur'an*. Bandung : Mizan.

- Siregar, L. (2014). Pendidikan Dalam Konsep Islam. *Jurnal Da'wah and Communication Sciences faculty: IAIN Padangsidimpun*, 24-36.
- Sudarwati. (2009). Tesis: Perbedaan Resiko Putus Sekolah Anak Usia 7-15 Tahun pada Tahun 1998 dan 2006 di Indonesia. Program Pascasarjana, Kajian Kependudukan dan Ketenagakerjaan, Universitas Indonesia.
- Sugianto, E. (2017). Faktor Penyebab Anak Putus Sekolah Tingkat SMA di Desa Bukit Lipai Kecamatan Batang Cenaku Kabupaten Inderagiri Hulu. *Jurusan Sosiologi Fakultas Ilmu Sosial dan Ilmu Politik Universitas Riau*, 35-44.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung: Alfabeta.
- Tyasti, A. E., Ispriyanti, D., & Hoyyi, A. (2015). Algoritma Iterative Dichotomiser 3 (ID3) untuk Mengidentifikasi Data Rekam Medis. *Gaussian, IV*, 237-246.
- Wassahua, S. (2016). Analisis Faktor-Faktor Penyebab Putus Sekolah di Kampung Wara Negeri Hative Kecil Kota Ambon. *Al-Iltizam*, 93-113.
- Yang, S., Guo, J.-Z., & Jin, J. W. (2017). An Improved ID3 Algorithm for Medical Data Classification. *Computers and Electrical Engineering*, 1-14.
- Yuniar, D. P. (2021). Faktor Penyebab Putus Sekolah Pada Anak di Wilayah Pedesaan (Studi Kasus di Desa Sonowangi Kabupaten Malang) vol.6(1). *E-plus*, 99-107.
- Yuniarti, R., & Hartanti, W. (2019). Regresi Nonparametrik Menggunakan Metode Robust dan Cross Validation. *Jurnal UJMC*, 9-16.
- Zhou, H., Zhang, J., Zhou, Y., Guo, X., & Ma, Y. (2021). A Feature Selection Algorithm of Decision Tree based on Feature Weight . *Expert Systems With Applications*, 1-14.

LAMPIRAN

Lampiran 1 Potongan Data Asli Siswa

Tahun Ajaran	NIPD	Nama	NISN	JK	Data Ayah			Data Ibu			Jumlah Anggota Keluarga	Pendapatan keluarga	Keterangan
					Nama	Jenjang Pendidikan	Pekerjaan	Nama	Jenjang Pendidikan	Pekerjaan			
2019/2020	L	...	SMA / sederajat	Wiraswasta	...	D2	Karyawan Swasta	4	sedang	Melanjutkan
2019/2020	P	...	SMA / sederajat	Karyawan Swasta	...	D2	Lainnya	5	sedang	Melanjutkan
2019/2020	P	...	S1	PNS/TNI/Polri	...	D2	Lainnya	5	tinggi	Melanjutkan
2019/2020	L	...	S1	Wiraswasta	...	D3	Wiraswasta	4	rendah	Melanjutkan
2019/2020	L	...	S1	Karyawan Swasta	...	D3	Lainnya	4	sedang	Melanjutkan
2019/2020	L	...	D3	PNS/TNI/Polri	...	D3	PNS/TNI/Polri	4	tinggi	Melanjutkan
2019/2020	P	...	SD / sederajat	Petani	...	Putus SD	Petani	5	rendah	Melanjutkan
2019/2020	P	...	Tidak sekolah	Petani	...	Putus SD	Petani	5	rendah	Putus
2019/2020	P	...	Tidak Sekolah	Petani	...	Putus SD	Petani	7	rendah	Putus
2019/2020	P	...	Putus SD	Petani	...	Putus SD	Tidak bekerja	6	rendah	Putus
2019/2020	L	...	Putus SD	Petani	...	Putus SD	Lainnya	4	rendah	Melanjutkan
2019/2020	L	...	Putus SD	Petani	...	Putus SD	Petani	4	rendah	Melanjutkan
2019/2020	L	...	Putus SD	Petani	...	Putus SD	Petani	7	rendah	Melanjutkan
2019/2020	L	...	Putus SD	Petani	...	Putus SD	Petani	5	rendah	Melanjutkan

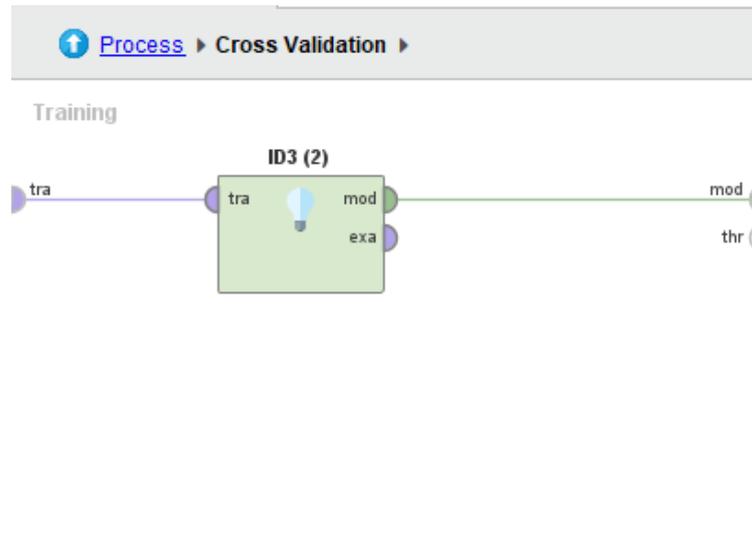
2019/2020	L	...	Putus SD	Petani	...	Putus SD	Petani	5	rendah	Melanjutkan
2019/2020	L	...	SD / sederajat	Petani	...	Putus SD	Petani	4	rendah	Melanjutkan
2019/2020	P	...	Putus SD	Petani	...	Putus SD	Tidak bekerja	4	rendah	Melanjutkan
2019/2020	L	...	Putus SD	Petani	...	Putus SD	Petani	10	rendah	Melanjutkan
2019/2020	P	...	Putus SD	Petani	...	Putus SD	Tidak bekerja	4	rendah	Melanjutkan
2019/2020	L	...	SMP / sederajat	Wiraswasta	...	Putus SD	Tidak bekerja	6	rendah	Melanjutkan
2019/2020	L	...	SD / sederajat	Petani	...	Putus SD	Tidak bekerja	4	rendah	Melanjutkan
2019/2020	P	...	SD / sederajat	Petani	...	Putus SD	Tidak bekerja	4	rendah	Melanjutkan
2019/2020	L	...	SD / sederajat	Petani	...	Putus SD	Tidak bekerja	7	rendah	Melanjutkan
2019/2020	L	...	Putus SD	Petani	...	Putus SD	Tidak bekerja	3	rendah	Melanjutkan
2019/2020	L	...	SMP / sederajat	Karyawan Swasta	...	Putus SD	Tidak bekerja	4	rendah	Melanjutkan
2019/2020	P	...	Putus SD	Petani	...	Putus SD	Tidak bekerja	6	rendah	Melanjutkan
2019/2020	L	...	Putus SD	Petani	...	Putus SD	Tidak bekerja	4	rendah	Melanjutkan
2019/2020	P	...	Putus SD	Petani	...	Putus SD	Tidak bekerja	6	rendah	Melanjutkan
2019/2020	P	...	SD / sederajat	Petani	...	Putus SD	Tidak bekerja	4	rendah	Melanjutkan
2019/2020	P	...	Tidak sekolah	Petani	...	Putus SD	Tidak bekerja	4	rendah	Melanjutkan
2019/2020	P	...	Putus SD	Petani	...	Putus SD	Tidak bekerja	6	rendah	Melanjutkan
2019/2020	L	...	Putus SD	Petani	...	Putus SD	Tidak bekerja	5	rendah	Melanjutkan
...

Lampiran 2 Potongan Data Putus Sekolah Setelah di-Preprocessing

Jenis Kelamin	Pendidikan Ayah	Pekerjaan Ayah	Pendidikan Ibu	Pekerjaan Ibu	Jumlah Anggota Keluarga	Pendapatan Keluarga	Keterangan
L	SMA / sederajat	Wiraswasta	D2	Karyawan Swasta	4	Sedang	Melanjutkan
P	SMA / sederajat	Karyawan Swasta	D2	Lainnya	5	Sedang	Melanjutkan
P	S1	PNS/TNI/Polri	D2	Lainnya	5	Tinggi	Melanjutkan
L	S1	Wiraswasta	D3	Wiraswasta	4	Rendah	Melanjutkan
L	S1	Karyawan Swasta	D3	Lainnya	4	Sedang	Melanjutkan
L	D3	PNS/TNI/Polri	D3	PNS/TNI/Polri	4	Tinggi	Melanjutkan
P	SD / sederajat	Petani	Putus SD	Petani	5	Rendah	Putus
P	Tidak sekolah	Petani	Putus SD	Petani	5	Rendah	Putus
P	Tidak Sekolah	Petani	Putus SD	Petani	7	Rendah	Putus
P	Putus SD	Petani	Putus SD	Tidak bekerja	6	Rendah	Putus
L	Putus SD	Petani	Putus SD	Lainnya	4	Rendah	Melanjutkan
L	Putus SD	Petani	Putus SD	Petani	4	Rendah	Melanjutkan
L	Putus SD	Petani	Putus SD	Petani	7	Rendah	Melanjutkan
L	Putus SD	Petani	Putus SD	Petani	5	Rendah	Melanjutkan
L	Putus SD	Petani	Putus SD	Petani	5	Rendah	Melanjutkan
L	SD / sederajat	Petani	Putus SD	Petani	4	Rendah	Melanjutkan
P	Putus SD	Petani	Putus SD	Tidak bekerja	4	Rendah	Melanjutkan
L	Putus SD	Petani	Putus SD	Petani	10	Rendah	Melanjutkan
P	Putus SD	Petani	Putus SD	Tidak bekerja	4	Rendah	Melanjutkan
L	SMP / sederajat	Wiraswasta	Putus SD	Tidak bekerja	6	Rendah	Melanjutkan
L	SD / sederajat	Petani	Putus SD	Tidak bekerja	4	Rendah	Melanjutkan
P	SD / sederajat	Petani	Putus SD	Tidak bekerja	4	Rendah	Melanjutkan

L	SD / sederajat	Petani	Putus SD	Tidak bekerja	7	Rendah	Melanjutkan
L	Putus SD	Petani	Putus SD	Tidak bekerja	3	Rendah	Melanjutkan
L	SMP / sederajat	Karyawan Swasta	Putus SD	Tidak bekerja	4	Rendah	Melanjutkan
P	Putus SD	Petani	Putus SD	Tidak bekerja	6	Rendah	Melanjutkan
L	Putus SD	Petani	Putus SD	Tidak bekerja	4	Rendah	Melanjutkan
P	Putus SD	Petani	Putus SD	Tidak bekerja	6	Rendah	Melanjutkan
P	SD / sederajat	Petani	Putus SD	Tidak bekerja	4	Rendah	Melanjutkan
P	Tidak sekolah	Petani	Putus SD	Tidak bekerja	4	Rendah	Melanjutkan
P	Putus SD	Petani	Putus SD	Tidak bekerja	6	Rendah	Melanjutkan
L	Putus SD	Petani	Putus SD	Tidak bekerja	5	Rendah	Melanjutkan
L	SMP / sederajat	Wiraswasta	Putus SD	Tidak bekerja	5	Rendah	Melanjutkan
L	Putus SD	Petani	Putus SD	Tidak bekerja	4	Rendah	Melanjutkan
L	SMP / sederajat	Wiraswasta	Putus SD	Tidak bekerja	4	Rendah	Melanjutkan
P	Putus SD	Petani	Putus SD	Tidak bekerja	6	Rendah	Melanjutkan
L	Putus SD	Petani	Putus SD	Tidak bekerja	6	Rendah	Melanjutkan
P	Putus SD	Buruh	Putus SD	Tidak bekerja	5	Rendah	Melanjutkan
L	Putus SD	Petani	Putus SD	Tidak bekerja	3	Rendah	Melanjutkan
P	SMP / sederajat	Petani	Putus SD	Tidak bekerja	6	Rendah	Melanjutkan
P	Putus SD	Petani	Putus SD	Tidak bekerja	4	Rendah	Melanjutkan
P	SMP / sederajat	Petani	Putus SD	Tidak bekerja	6	Rendah	Melanjutkan
L	Tidak sekolah	Petani	Putus SD	Tidak bekerja	6	Rendah	Melanjutkan
.....

Lampiran 3 Proses *Cross Validation* pada *Rapidminer*



Lampiran 4 Kolom *Number of Fold*

🔗 Cross Validation	
<input type="checkbox"/>	leave one out
number of folds	<input type="text" value="10"/>
sampling type	<input type="text" value="automatic"/>

Lampiran 5 Gambar Hasil Perhitungan Kombinasi Data *Testing* dan Data *Training* dengan Menggunakan *Rapidminer*

accuracy: 98.48% +/- 1.60% (micro average: 98.48%)

	true Melanjutkan	true Putus	class precision
pred. Melanjutkan	321	3	99.07%
pred. Putus	2	3	60.00%
class recall	99.38%	50.00%	

accuracy: 96.96% +/- 5.06% (micro average: 96.96%)

	true Melanjutkan	true Putus	class precision
pred. Melanjutkan	316	3	99.06%
pred. Putus	7	3	30.00%
class recall	97.83%	50.00%	

accuracy: 98.18% +/- 2.85% (micro average: 98.18%)

	true Melanjutkan	true Putus	class precision
pred. Melanjutkan	320	3	99.07%
pred. Putus	3	3	50.00%
class recall	99.07%	50.00%	

accuracy: 98.48% +/- 3.10% (micro average: 98.48%)

	true Melanjutkan	true Putus	class precision
pred. Melanjutkan	321	3	99.07%
pred. Putus	2	3	60.00%
class recall	99.38%	50.00%	

accuracy: 98.18% +/- 3.70% (micro average: 98.18%)

	true Melanjutkan	true Putus	class precision
pred. Melanjutkan	320	3	99.07%
pred. Putus	3	3	50.00%
class recall	99.07%	50.00%	

Lampiran 6 Tabel Hasil Perhitungan Untuk menentukan *Internal Node dan Leaf Node*

No	Atribut	Kelas	Jumlah (S)	Melanjutkan (Si)	Putus Sekolah(Si)	Entropy	Information Gain
		Pendidikan Ibu-Putus SD	49	45	4	0,407906	
1		Jenis Kelamin					
		Laki-laki	29	29	0	0	0,113241
		Perempuan	20	16	4	0,721928	
2		Pendidikan Ayah					
		Tidak Sekolah	4	2	2	1	0,104631
		Putus SD	27	26	1	0,228538	
		SD/ Sederajat	10	9	1	0,468996	
		SMP/Sederajat	8	8	0	0	
3		Pekerjaan Ayah					
		Petani	40	36	4	0,468996	0,025052
		Peternak	1	1	0	0	
		Buruh	1	1	0	0	
		Karyawan Swasta	1	1	0	0	
		Wiraswasta	5	5	0	0	
		PNS/TNI/POLRI	1	1	0	0	
5		Pekerjaan Ibu					
		Tidak Bekerja	36	35	1	0,183122	0,074687
		Petani	12	9	3	0,811278	
		Lainnya	1	1	0	0	
6		Jumlah Anggota Keluarga					
		3	3	3	0	0	0,103009
		4	20	20	0	0	
		5	10	8	2	0,721928	
		6	12	11	1	0,413817	
		7	3	2	1	0,918296	
		10	1	1	0	0	
7		Pendapatan Keluarga					
		Rendah	38	34	4	0,485461	0,031426
		Sedang	10	10	0	0	
		Tinggi	1	1	0	0	

No	Atribut	Kelas	Jumlah (S)	Melanjutkan (Si)	Putus Sekolah(Si)	Entropy	Information Gain
		Pendidikan Ibu-Putus SD, Jenis Kelamin-Perempuan	20	16	4	0,721928	
2		Pendidikan Ayah					
		Tidak Sekolah	3	1	2	0,918296	0,180205
		Putus SD	11	10	1	0,439497	

		SD/ Sederajat	4	3	1	0,811278	
		SMP/Sederajat	2	2	0	0	
3	Pekerjaan Ayah						
		Petani	18	14	4	0,764205	0,034144
		Buruh	1	1	0	0	
		Wiraswasta	1	1	0	0	
5	Pekerjaan Ibu						
		Tidak Bekerja	17	16	1	0,322757	0,447585
		Petani	3	0	3	0	
6	Jumlah Anggota Keluarga						
		4	8	8	0	0	0,314843
		5	4	2	2	1	
		6	7	6	1	0,591673	
		7	1	0	1	0	
7	Pendapatan Keluarga						
		Rendah	17	13	4	0,787127	0,05287
		Sedang	3	3	0	0	

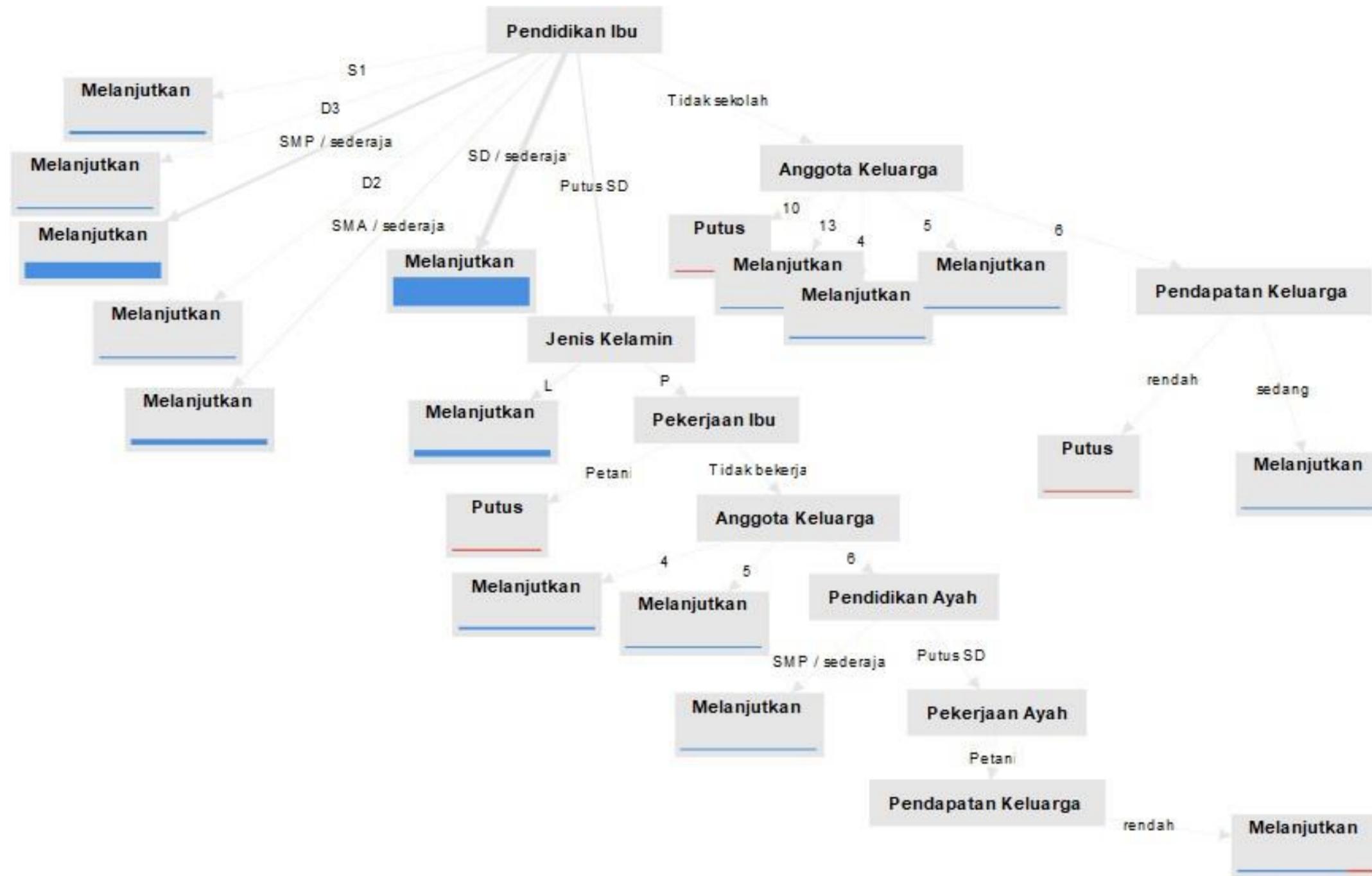
No	Atribut	Kelas	Jumlah (S)	Melanjutkan (Si)	Putus Sekolah(Si)	Entropy	Information Gain
		Pendidikan Ibu-Putus SD, Jenis Kelamin-Perempuan, Pekerjaan Ibu-Tidak Bekerja	17	16	1	0,322757	
2	Pendidikan Ayah						
		Tidak Sekolah	1	1	0	0	0,038377
		Putus SD	11	10	1	0,439497	
		SD/ Sederajat	3	3	0	0	
		SMP/Sederajat	2	2	0	0	
3	Pekerjaan Ayah						
		Petani	15	14	1	0,353359	0,010969
		Buruh	1	1	0	0	
		Wiraswasta	1	1	0	0	
6	Jumlah Anggota Keluarga						
		4	8	8	0	0	0,079127
		5	2	2	0	0	
		6	7	6	1	0,591673	
	Pendapatan Keluarga						
7		Rendah	14	13	1	0,371232	0,017036
		Sedang	3	3	0	0	

No	Atribut	Kelas	Jumlah (S)	Melanjutkan (Si)	Putus Sekolah(Si)	Entropy	Information Gain
		Pendidikan Ibu-Putus SD, Jenis Kelamin-Perempuan,	7	6	1	0,591673	

	Pekerjaan Ibu-Tidak Bekerja, Jumlah Anggota Keluarga-6						
2	Pendidikan Ayah						
		Putus SD	5	4	1	0,721928	0,07601
		SMP/Sederajat	2	2	0	0	
3	Pekerjaan Ayah						
		Petani	7	6	1	0,591673	0
7	Pendapatan Keluarga						
		Rendah	7	6	1	0,591673	0

No	Atribut	Kelas	Jumlah (S)	Melanjutkan (Si)	Putus Sekolah(Si)	Entropy	Information Gain
		Pendidikan Ibu-Putus SD, Jenis Kelamin-Perempuan, Pekerjaan Ibu-Tidak Bekerja, Jumlah Anggota Keluarga-6, Pendidikan Ayah-Putus SD	5	4	1	0,721928	
3	Pekerjaan Ayah						
		Petani	5	4	1	0,721928	0
7	Pendapatan Keluarga						
		Rendah	5	4	1	0,721928	0

Lampiran 7 Output Decision Tree



Lampiran 8 *Output Confusion Matrix RapidMiner*

accuracy: 98.48% +/- 1.60% (micro average: 98.48%)

	true Melanjutkan	true Putus	class precision
pred. Melanjutkan	321	3	99.07%
pred. Putus	2	3	60.00%
class recall	99.38%	50.00%	

RIWAYAT HIDUP



Skripsi ini ditulis oleh Larazani Arum Baruna Putri, yang biasa dipanggil Laras. Dilahirkan di Banyumas pada tanggal 26 November 1998. Bertempat tinggal di Perumahan Taman Puspa Sari blok S-05, Kecamatan Candi, Kabupaten Sidoarjo Jawa Timur. Lahir dari pasangan Rizani dan Tusnilah, merupakan anak pertama dari 3 bersaudara serta memiliki 2 saudara kandung yang bernama Agha Zaki Athaya Samudra dan Azhar Bahari Zain.

Pendidikan yang ditempuh berawal dari SD/MI Pucang Sidoarjo sejak tahun 2006-2012, kemudian melanjutkan ke jenjang SMP di Bording School Ar-Rohmah Putri Malang pada tahun 2012-2014. Kemudian penulis melanjutkan sekolah ke jenjang SMA di SMAS Muhammadiyah 2 Sidoarjo pada tahun 2014-2017. Setelah itu melanjutkan pendidikan di Universitas Islam Negeri Maulana Malik Ibrahim Malang dengan program studi Matematika. Selama duduk dibangku kuliah, ia aktif berorganisasi di UKM Unior serta beberapa kali menjadi panitia di beberapa acara besar. Diantaranya pernah menjadi lines judge pada pertandingan bulutangkis pada acara PIONER IX di UIN Malang, serta berpartisipasi sebagai HUMAS pada pertandingan BUC selama 3 periode yang merupakan pertandingan bulutangkis se-Jawa Bali, dan masih banyak lagi.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

BUKTI KONSULTASI SKRIPSI

Nama : Larazani Arum Baruna Putri
NIM : 17610068
Fakultas / Program Studi : Sains dan Teknologi / Matematika
Judul Skripsi : Klasifikasi Faktor Penyebab Siswa Putus Sekolah
Menggunakan Metode *Decision Tree ID3*
Pembimbing I : Angga Dwi Mulyanto, M.Si
Pembimbing II : Erna Herawati, M.Pd

No	Tanggal	Hal	Tanda Tangan
1.	01 Juni 2021	Konsultasi BAB I	1.
2.	11 November 2021	Konsultasi Revisi BAB I	2.
3.	02 Februari 2022	Konsultasi BAB II dan III	3.
4.	30 Mei 2022	Konsultasi Revisi BAB II dan III	4.
5.	31 Mei 2022	Konsultasi Ayat Integrasi Islam	5.
6.	02 Juni 2022	ACC untuk Seminar Proposal	6.
7.	22 November 2022	Konsultasi Keagamaan	7.
8.	28 November 2022	Konsultasi BAB IV dan V	8.
9.	29 November 2022	Konsultasi Revisi Keagamaan	9.
10.	30 November 2022	Konsultasi Revisi BAB IV dan V	10.
11.	02 Desember 2022	Konsultasi Revisi Kajian Agama TTD Untuk Seminar Hasil	11.
12.	02 Desember 2022	ACC untuk Seminar Hasil	12.
13.	16 Desember 2022	Konsultasi Revisi Pasca Seminar Hasil	13.
14.	17 Desember 2022	Konsultasi Integrasi Agama dengan Hasil Penelitian	14.
15.	22 Desember 2022	Konsultasi Keseluruhan Revisi Seminar Hasil	15.
16.	23 Desember 2022	ACC Keseluruhan untuk Sidang	16.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

No	Tanggal	Hal	Tanda Tangan
17.	27 Desember 2022	ACC Skripsi untuk Syarat Yudisium	17.

Malang, 27 Desember 2022

Mengetahui,
Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc

NIP. 19741129 200012 2 005