

MODEL DAN ANALISA FAKTOR EKSTERNAL AKTIFITAS SISWA KELAS X TKJ SMKN 1 PAKONG PAMEKASAN MENGGUNAKAN ALGORITMA *DECISION TREE*

Muhsi¹

¹Teknik Komputer dan Jaringan, SMKN 1 Pakong, Pamekasan

¹Sistem Informasi, Fakultas Teknik, Universitas Islam Madura, Pamekasan

¹muhsiy@gmail.com

ABSTRAK

Kegiatan pembelajaran yang interaktif antara siswa guru akan efektif apabila terdapat aktifitas yang baik dari siswa. Terdapat faktor eksternal yang mempengaruhinya sehingga perlu dimodelkan dan dianalisis apa saja faktor eksternal tersebut kemudian bagaimana hal tersebut bisa terjadi. Penelitian ini bermaksud memodelkan dan menganalisa faktor eksternal aktifitas siswa kelas X TKJ SMKN 1 Pakong menggunakan algoritma *decision tree* dalam hal ini adalah algoritma C4.5. karena di beberapa penelitian menunjukkan tingkat akurasi yang bagus dan mudah untuk diinterpretasikan. Model *decision tree* yang dihasilkan adalah faktor eksternal tidak berkumpulnya siswa dengan orang tua di rumah menjadi faktor utama rendahnya aktifitas siswa kelas X TKJ SMKN 1 Pakong. Sedangkan bagi siswa yang berkumpul dengan orang tua di rumah bisa rendah juga aktifitasnya karena asal sekolah4 dan mode transportasinya menggunakan motor sendiri namun bisa tinggi sekali apabila berboncengan dengan temannya. Dengan demikian, secara umum, siswa yang berkumpul dengan orang tuanya di rumah memiliki tingkat aktifitas tinggi bahkan sangat tinggi. Berdasarkan hal tersebut, siswa kelas X TKJ SMKN 1 Pakong memiliki tingkat aktifitas rendah sebanyak 31%, aktifitas tinggi sebanyak 45% dan aktifitas sangat tinggi sebanyak 24%.

Kata kunci: aktifitas siswa, TKJ SMKN 1 Pakong, *decision tree*, algoritma C4.5.

ABSTRACT

Interactive learning activities between students and teachers will be effective if there are good activities from students. There are external factors that influence it, so it needs to be modeled and analyzed what these external factors are and then how can this happen. This study intends to model and analyze external factors of class X TKJ student activities at SMKN 1 Pakong using a decision tree algorithm in this case the C4.5 algorithm because in several studies it shows a good level of accuracy and is easy to interpret. The resulting decision tree model is an external factor that does not gather students with parents at home to be the main factor in the low activity of class X students TKJ SMKN 1 Pakong. Meanwhile, for students who gather with their parents at home, their activities can also be low because they come from sekolah4 and the mode of transportation uses their own motorbike, but it can be very high when riding with their friends. Thus, in general, students who gather with their parents at home have a high or even very high activity level. Based on this, the students of class X TKJ SMKN 1 Pakong had a low activity level of 31%, high activity as much as 45% and very high activity as much as 24%.

Keywords: student activities, TKJ SMKN 1 Pakong, *decision tree*, C4.5 algorithm

1. PENDAHULUAN

Proses pembelajaran terjadi ketika terdapat kegiatan interaktif antara siswa dan guru [1], [2] baik dalam kelas tatap muka maupun pendidikan jarak jauh. Interaksi antara siswa dan guru menjadi kunci utama keberhasilan proses pembelajaran sebagai bagian dari ikhtiar mencerdaskan kehidupan bangsa. Interaksi tersebut ditandai dengan aktifitas yang selalu rutin dilakukan baik oleh guru maupun siswa baik dalam bentuk perkataan/berbicara, mendengarkan, membaca dan sejenisnya [3]–[5]. Keaktifan siswa

sangat penting untuk menjadikan siswa lebih cepat berhasil melalui prestasi yang akan dia dapatkan nanti [6]. Hal ini dikarenakan dengan belajar aktif akan menjadikan siswa lebih dapat mengoptimalkan potensi yang dimilikinya berdasarkan karakteristik yang dia dimiliki guna mencapai hasil belajar yang memuaskan [7]–[9].

Keaktifan siswa dalam mengikuti pembelajaran dapat dipengaruhi oleh factor internal dan eksternal [5], [10], [11]. Muhibbin Syah (2021) menjelaskan bahwa faktor yang dapat mempengaruhi terhadap keaktifan siswa dalam belajar dapat karena faktor internal (berupa pengaruh dari dalam diri siswa), faktor eksternal (berupa pengaruh dari luar siswa), dan faktor pendekatan dalam proses belajar (approach to learning) [1], [12]. Factor tersebut menjadi atensi bersama baik guru sebagai pendidik yang akan memberikan materi kemudian sekolah sebagai lembaga yang mengelola proses pembelajaran serta siswa itu sendiri sebagai pribadi pembelajar. Kegiatan berkesinambungan ini merupakan sistem pendidikan di tingkat sekolah yang satu dengan lainnya saling berkaitan dan harus saling mendukung. Pada penelitian ini akan melakukan pemodelan dan Analisa terhadap factor eksternal siswa di SMK Negeri 1 Pakong yang mempengaruhi terhadap keaktifan siswa dalam mengikuti kegiatan pembelajaran.

SMK Negeri 1 Pakong merupakan lembaga pendidikan tingkat menengah di bawah Dinas Pendidikan Jawa Timur yang berada di wilayah kecamatan, yaitu tepatnya di Desa Bandungan Kecamatan Pakong Kabupaten Pamekasan. Sebagai lembaga pendidikan yang berada di daerah kecamatan banyak faktor eksternal yang mempengaruhinya diantaranya adalah keberadaan siswa yang jauh dari orang tua karena merantau ke luar daerah. Selain hal tersebut masih ada beberapa faktor eksternal lainnya yang penulis modelkan untuk melihat bagaimana faktor eksternal tersebut mempengaruhi keaktifan siswa dalam proses pembelajaran. Model faktor eksternal keaktifan siswa dibuat menggunakan algoritma *decison tree* untuk kemudian dianalisa berdasarkan kecenderungannya sehingga menjadi acuan oleh guru, lembaga pendidikan dan siswa dalam menentukan kebijakan untuk mendukung keberhasilan belajar siswa.

Algoritma *decison tree* merupakan metode berbasis numerik namun dapat dikonversi menjadi *rule* (aturan) sebagai pengetahuan [13]. Pengetahuan yang didapat dari model menggunakan algoritma *decison tree* ini berupa estimasi tentang faktor eksternal keaktifan siswa dalam bentuk pohon keputusan yang penyajiannya nanti dapat berbentuk aturan [14]. Penggunaan algoritma *decison tree* ini sangat efektif dalam penyajian model estimasi karena nantinya akan menghitung nilai terkecil untuk dijadikan simpul awal dalam menentukan faktor eksternal keaktifan siswa [10]. Melalui simpul tersebut akan menghasilkan *node* sebagai pohon hubungan antar faktor eksternal tersebut [10], [14]–[16].

Algoritma *decision tree* merupakan teknik yang banyak digunakan untuk melakukan klasifikasi data pada teknik data mining [13], [17]. Implementasi teknik data mining pada dunia pendidikan dapat digunakan untuk mengidentifikasi kebutuhan dan preferensi siswa terhadap pilihan jenis kursus atau spesialisasi berdasarkan minat bakat, mengidentifikasi tren pola dari kondisi siswa, memprediksi tingkat pengetahuan siswa, mendukung eksplorasi data siswa secara otomatis, mempermudah memetakan siswa, membantu proses bisnis manajemen sekolah dan lain sebagainya [17]. Beberapa penelitian dengan penerapan algoritma *decision tree* diantaranya adalah tentang prediksi mahasiswa non aktif [18], penentuan keberhasilan akademik mahasiswa [19], pengolahan data siswa [20] dan klasifikasi peserta didik untuk mengelompokkan siswa dalam suatu kelas berdasarkan atribut yang digunakan [21].

2. METODOLOGI PENELITIAN

2.1 Metode, Data dan Waktu Penelitian

Penelitian ini merupakan terapan karena menerapkan teknik data mining berupa klasifikasi data siswa siswa kelas X TKJ SMKN 1 Pakong dengan algoritma *decision tree*. Klasifikasi data siswa berhubungan dengan keaktifan siswa dalam mengikuti kegiatan pembelajaran berdasarkan atribut faktor eksternal, yaitu jarak rumah siswa dengan sekolah, asal sekolah tingkat dasar sebelumnya, mode transportasi siswa, keberadaan orang tua saat ini dan di rumah berkumpul dengan siapa serta tingkat kehadiran siswa ke sekolah. Lokasi penelitian ini adalah SMKN 1 Pakong dan data yang digunakan adalah biodata dan

kegiatan siswa baru pelajaran 2019/2020, yaitu rentang bulan Juli 2019 sampai dengan Maret 2020. Hasil akhirnya berupa faktor yang mempengaruhi tingkat keaktifan siswa kelas X TKJ SMKN 1 Pakong.

2.2 Teknik Data Mining

Menurut beberapa ahli teknik data mining digunakan untuk eksplorasi dari hasil analisa data, metodologi dalam analisis data, proses yang digunakan untuk memperoleh pengetahuan [17], [22]–[24]. Selain itu teknik data mining merupakan proses dan metodologi untuk menerapkan alat dan teknik analisa [17], [25]. Namun ada beberapa ahli berpendapat bahwa teknik data mining merupakan seni untuk mengekstrak informasi dari suatu data, proses menemukan pola dalam data (Witten, 2005), dan teknik menemukan data yang bermakna (pengetahuan) dari data yang besar sehingga akan memperoleh hasil yang jelas dan lebih bermanfaat [17], [25]–[27].

Sesuai dengan teknik data mining yang dijelaskan beberapa ahli, terdapat metode dan algoritma yang digunakan dalam data mining [17], [28], [29], yaitu :

1. Estimasi, dapat menggunakan algoritma Linear Regression (LR), Neural Network (NN), Deep Learning (DL), Support Vector Machine (SVM), Generalized Linear Model (GLM)
2. Prediksi/Peramalan, dapat menggunakan algoritma Linear Regression (LR), Neural Network (NN), Deep Learning (DL), Support Vector Machine (SVM), Generalized Linear Model (GLM).
3. Klasifikasi, dapat menggunakan Decision Tree (CART, ID3, C4.5, Credal DT, Credal C4.5, Adaptive Credal C4.5), Naive Bayes (NB), K-Nearest Neighbor (kNN), Linear Discriminant Analysis (LDA), Logistic Regression (LogR).
4. Klastering, dapat menggunakan K-Means, K-Medoids, Self-Organizing Map (SOM), Fuzzy C-Means (FCM).
5. Asosiasi, dapat menggunakan FP-Growth, A Priori, Coefficient of Correlation, Chi Square.

Tahapan dalam penerapan teknik data mining menurut beberapa ahli berbeda. Giudici (2003) [30], membuat tahapan dalam proses data mining dalam 4 (empat) tahap, yaitu *Strategic, Training, Creation, dan Migration*. Kantardzic (2004) [31] menjelaskan bahwa terdapat 5 (lima) tahapan proses data mining, yaitu *State the Problem, Collect the data, Pre-Process the data, Estimate the model, dan Interpret model and draw conclusions*. Hsu (2006) [32], merumuskan tahapan dalam proses data mining ada 4 (empat), yaitu *Data Collection, Data Pre-processing, Data Mining dan Information interpretation and Visualisation*. Selain hal tersebut terdapat model tahapan dalam proses data mining yang digunakan oleh beberapa peneliti. Ali (2013) menjelaskan bahwa tahapan yang umum digunakan dalam proses data mining di industri adalah model CRISP-DM (Cross-Industry Standard Process Data Mining) yang memuat tahapan *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment* [17]. Giudici (2003) menjelaskan bahwa pada tahun 1995, peneliti Usama Fayaad pada kegiatan memperkenalkan istilah *knowledge discovery in databases (KDD)* yang terdiri dari *data selection, pre-processing/cleaning, transformation coding, data mining, dan interpretation/evaluation*. Komponen data mining yang terdapat di dalamnya tersebut kemudian *KDD* dijadikan tahapan proses data mining [30].

Penelitian ini menggunakan teknik data mining yaitu klasifikasi dengan algoritma decision tree melalui tahapan pada model CRISP-DM, yaitu:

1. *Business Understanding*

Pada tahap ini dilakukan objektivitas terhadap permasalahan berupa rendahnya aktifitas siswa kelas X TKJ di SMKN 1 Pakong untuk kemudian akan dilihat faktor eksternal apa saja yang mempengaruhinya.

2. *Data Understanding*

Pada tahap ini dilakukan pengumpulan data dari aktifitas siswa berupa absensi kehadiran yang dikumpulkan dari dokumen absen guru. Kemudian variabel dari siswa berupa

kondisi siswa secara eksternal. Data tersebut didapat dari kantor Tata Usaha berupa biodata siswa yang diisi sendiri oleh siswa dan dikumpulkan pada saat daftar ulang.

3. *Data Preparation*

Pada tahap ini data absensi dan biodata siswa ditabulasi untuk kemudian diseleksi. Proses ini dimaksudkan untuk menyiapkan data agar siap untuk diolah.

4. *Modelling*

Pada tahap ini dilakukan proses data mining berupa klasifikasi data aktifitas dan biodata siswa menggunakan metode *Decision Tree* yang dioleh melalui *software RapidMiner* dengan keluarannya berupa model klasifikasi.

5. *Evaluation*

Pada tahap ini hasil model pada tahapan sebelumnya dievaluasi dan dilihat performanya berupa tingkat akurasi model. Selanjutnya dianalisis mengenai faktor eksternal siswa yang memiliki pengaruh besar terhadap aktifitas belajar siswa kelas X TKJ Pakong.

6. *Deployment*

Pada tahap ini dilakukan pelaporan dan pemantuan visualisasi grafik model faktor eksternal siswa untuk kemudian dibuatkan keputusan yang relevan dengan kondisi siswa kelas X TKJ SMKN 1 Pakong.

2.3 *Decision Tree*

Decision tree (pohon keputusan) merupakan teknik yang membantu dalam pengambilan keputusan yang bentuk menyerupai pohon atau berhirarki. Model yang dihasilkan nantinya berupa prosedur rekursif, di mana dalam satu set n unit statistik secara progresif dibagi menjadi beberapa kelompok, berdasarkan aturan pembagian yang bertujuan untuk memaksimalkan homogenitas atau kemurnian ukuran variabel respon di setiap kelompok yang diperoleh [30]. Beberapa algoritma yang digunakan dalam metode *decision tree*, yaitu CHAID (*Chi-squared Automatic Interaction Detection*), CART (*Classification And Regression Tree*), ID3 (*Iterative Dichotomiser 3*), dan C4.5 yang merupakan pengembangan dari ID3 [30], [33]. Kemudian dikembangkan oleh beberapa peneliti menjadi Credal DT dan Credal C4.5 [28]–[30], [33]. Penelitian ini penulis menggunakan algoritma C4.5 karena di beberapa penelitian menunjukkan tingkat akurasi yang bagus dan mudah untuk diinterpretasikan.

Tahapan penggunaan algoritma C4.5 pada penelitian ini adalah (1) menghitung nilai *Entropy*, (2) menghitung nilai *Gain Ratio* untuk masing-masing atribut, (3) atribut yang memiliki *Gain Ratio* tertinggi dipilih menjadi akar (*root*) dan atribut yang memiliki nilai *Gain Ratio* lebih rendah dari akar (*root*) dipilih menjadi cabang (*branches*), (4) menghitung lagi nilai *Gain Ratio* tiap-tiap atribut dengan tidak mengikutsertakan atribut yang terpilih menjadi akar (*root*) di tahap sebelumnya, (5) atribut yang memiliki *Gain Ratio* tertinggi dipilih menjadi cabang (*branches*), (6) mengulangi langkah ke-4 dan ke-5 sampai dengan dihasilkan nilai *Gain* = 0 untuk semua atribut yang tersisa.

Berikut adalah formula yang digunakan [33]–[35][33], [34] :

1. Menghitung *Entropy*, yaitu nilai yang menyatakan ukuran ketidakpastian (*impurity*) atribut dari kumpulan objek data:

$$entropy(S) = \sum_{i=1}^n -p_i \times \log_2 p_i \quad (1)$$

Dimana :

S : himpunan data set

n : banyaknya partisi S

p_i : proporsi dari S_i terhadap S

2. Menghitung *Gain Ratio*, yaitu nilai pada atribut sebagai dasar dipilihnya sebagai simpul, akar maupun cabang suatu pohon keputusan.

$$gainratio(A) = \frac{gain(S,A)}{splitinfo(S,A)} \quad (2)$$

Dimana :

S : himpunan data set

A : atribut

$gain(S,A)$: *info gain* pada atribut A

$splitinfo(S,A)$: *split info* pada atribut A

3. Menghitung *Information Gain*, yaitu berupa ukuran efektifitas suatu atribut dalam mengklasifikasi data atau urutan atribut sehingga nilai terbesar yang akan dipilih.

$$Gain(S,A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times entropy(S_i) \quad (3)$$

Dimana :

S : himpunan data set

A : atribut

n : banyaknya partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke-i

$|S|$: jumlah kasus dalam S

4. Menghitung *split info* :

$$splitinfo(S,A) = \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (4)$$

Dimana :

S : himpunan data set

A : atribut

n : banyaknya partisi atribut A

S_i : jumlah sampel untuk atribut ke-i

Dari model yang telah dihasilkan nantinya akan dilihat performanya berupa akurasi (*accuracy*), model. Tingkat akurasi model didapat dari implementasi model terhadap data yang berbeda dengan atribut yang sama. Teknik melihat akurasi yang digunakan dalam penelitian ini adalah berupa perbandingan antara jumlah record yang tepat diprediksi dibandingkan dengan keseluruhan data. Data tersebut didapat melalui *Confussion Matrix* (**Tabel 1**) dengan menjumlahkan nilai secara diagonal. Semakin besar nilainya maka akurasi model akan semakin bagus.

Tabel 1 *Confussion Matrix*

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positive (TPs)	False Positive (FPs)
Predicted Negative (0)	False Negative (FNs)	True Negative (TNs)

3. HASIL DAN PEMBAHASAN

Pada bagian ini dilakukan pembahasan hasil sesuai dengan tahapan CRISP-DM. Adapun hasil dari setiap tahapan adalah sebagai berikut :

a. *Business Understanding*

Lokasi SMKN 1 Pakong berada di wilayah kecamatan tetapanya di Kecamatan Pakong. Sebagai lembaga pendidikan tingkat menengah yang berada di wilayah kecamatan memiliki tantangan tersendiri terutama berhubungan dengan aktifitas siswa yang notabene berasal dari daerah perdesaan.

Aktifitas siswa sangat dipengaruhi oleh motivasi yang dapat saja berasal dari dalam diri siswa itu sendiri yang dikenal dengan faktor internal maupun dari luar diri siswa atau faktor eksternal. Faktor eksternal tersebut yang menjadi konsen penelitian ini untuk mengetahui dan menganalisa faktor eksternal yang mempengaruhi aktifitas siswa kelas X TKJ SMKN 1 Pakong.

b. Data Understanding

Data yang dihimpun dalam penelitian ini adalah menyeleksi data siswa kelas X TKJ SMKN 1 Pakong yang merupakan variabel dari eksternal siswa seperti terlihat pada **Tabel 2**. Data diambil pada masa pembelajaran semester genap tahun pelajaran 2020/2021, yaitu rentang bulan Januari 2021 sampai dengan Juli 2021.

c. Data Preparation

Pada **Tabel 1** terlihat bahwa sebanyak 29 siswa yang dimodelkan dan dianalisa. Atributnya berupa jarak asal sekolah (asal sekolah tingkat dasar sebelumnya berupa MTs atau SMP dan status sekolah negeri atau swasta) dengan tipe data *polynomial*, (jarak rumah siswa dengan sekolah) dengan tipe data *integer*, orang tua (keberadaan orang tua saat ini, yaitu bekerja di dalam atau di luar wilayah Kabupaten Pamekasan) dengan tipe data *binomial*, kumpul (sehari-hari di rumah berkumpul dengan orang tua atau lainnya) dengan tipe data *polynomial*, transportasi (alat transportasi siswa ke sekolah) dengan tipe data *polynomial*, dan aktifitas (merupakan konversi dari jumlah absensi siswa seperti pada **Tabel 3**) dengan tipe data *polynomial*. Dari atribut-atribut tersebut terdapat atribut yang dijadikan *role* label/kelas, yaitu aktifitas berupa tingkat kehadiran siswa yang didapat dari buku absensi siswa dengan tipe data *polynomial* yang dikonversi dari tingkat kehadiran siswa seperti terlihat pada **Tabel 3**. Informasi nama siswa dan nama sekolah disamarkan untuk menghormati siswa dan sekolah asal siswa.

Tabel 2. Data dan atribut eksternal siswa

No	Nama	Atribut					
		Asal Sekolah	Jarak	Ortu	kumpul	Transport	Aktifitas
1	Siswa1	Sekolah3	10	Kumpul	Ortu	motor sendiri	Sangat Tinggi
2	Siswa2	Sekolah4	6	Kumpul	Ortu	bonceng	Tinggi
3	Siswa3	Sekolah4	12	Tidak Kumpul	Kakek/Nenek	motor sendiri	Rendah
4	Siswa4	Sekolah1	2	Kumpul	Ortu	bonceng	Rendah
5	Siswa5	Sekolah4	6	Kumpul	Ortu	motor sendiri	Rendah
6	Siswa6	Sekolah3	5	Kumpul	Ortu	motor sendiri	Rendah
7	Siswa7	Sekolah1	6	Kumpul	Ortu	motor sendiri	Rendah
8	Siswa8	Sekolah2	5	Tidak Kumpul	Saudara	motor sendiri	Rendah
9	Siswa9	Sekolah3	6	Kumpul	Ortu	bonceng	Rendah
10	Siswa10	Sekolah1	2	Kumpul	Ortu	motor sendiri	Sangat Tinggi
11	Siswa11	Sekolah1	1	Tidak Kumpul	Om/tante	jalan kaki	Sangat Tinggi
12	Siswa12	Sekolah3	7	Kumpul	Ortu	bonceng	Sangat Tinggi
13	Siswa13	Sekolah2	9	Tidak Kumpul	Nenek	motor sendiri	Sangat Tinggi
14	Siswa14	Sekolah3	2	Kumpul	Ortu	bonceng	Sangat Tinggi
15	Siswa15	Sekolah3	9	Tidak Kumpul	Om/tante	motor sendiri	Sangat Tinggi
16	Siswa16	Sekolah4	12	Kumpul	Ortu	bonceng	Tinggi
17	Siswa17	Sekolah1	2	Kumpul	Ortu	dianter	Tinggi
18	Siswa18	Sekolah1	2	Kumpul	Ortu	motor sendiri	Tinggi
19	Siswa19	Sekolah3	9	Kumpul	Ortu	bonceng	Tinggi
20	Siswa20	Sekolah3	6	Kumpul	Ortu	motor sendiri	Tinggi

No	Nama	Atribut					
		Asal Sekolah	Jarak	Ortu	kumpul	Transport	Aktifitas
21	Siswa21	Sekolah4	20	Kumpul	Ortu	motor sendiri	Tinggi
22	Siswa22	Sekolah4	6	Kumpul	Ortu	bonceng	Tinggi
23	Siswa23	Sekolah1	2	Tidak Kumpul	Kakek/Nenek	jalan kaki	Tinggi
24	Siswa24	Sekolah3	2	Kumpul	Ortu	motor sendiri	Tinggi
25	Siswa25	Sekolah4	6	Kumpul	Ortu	motor sendiri	Tinggi
26	Siswa26	Sekolah1	6	Kumpul	Ortu	bonceng	Sangat Rendah
27	Siswa27	Sekolah1	2	Kumpul	Ortu	motor sendiri	Sangat Rendah
28	Siswa28	Sekolah3	6	Kumpul	Ortu	motor sendiri	Rendah
29	Siswa29	Sekolah1	15	Kumpul	ortu	motor sendiri	Rendah

Tabel 3. Konversi tingkat keaktifan siswa

Jumlah absensi	Tingkat keaktifan
> 7	Sangat Rendah
5 – 7	Rendah
1 – 4	Tinggi
0	Sangat Tinggi

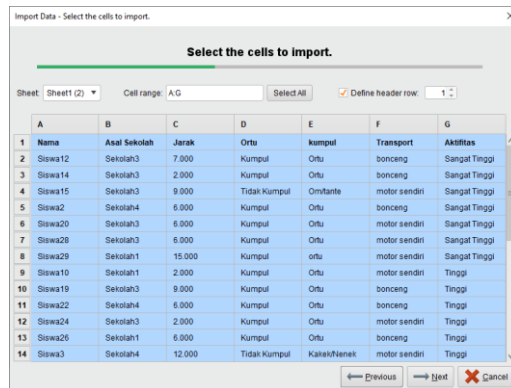
Data siswa pada **Tabel 2** dipecah menjadi dua bagian dengan pesentase. Data pertama digunakan untuk training menyusun model *Decision Tree* dengan persentase 75% atau 22 *record*. Sedangkan data kedua digunakan untuk kebutuhan implementasi model yang telah dihasilkan sebelumnya sekaligus untuk melihat performa dari model itu sendiri, yaitu 25% atau 7 *record*. Pembagian kedua data tersebut dilakukan secara *random* oleh sistem *software RapidMiner*. Data dari implementasi model ini menjadi nilai input nantinya ke dalam *Confussion Matrix*.

d. *Modelling*

Teknik dan metode data mining yang digunakan dalam penelitian ini ada klasifikasi menggunakan algoritma *decision tree*, yaitu algoritma algoritma C4.5. Proses pemodelan data faktor eksternal yang berhubungan dengan aktifitas siswa kelas X TKJ di SMKN 1 Pakong menggunakan *software RapidMiner Stodio version 9.6*. Adapun langkahnya sebagai berikut :

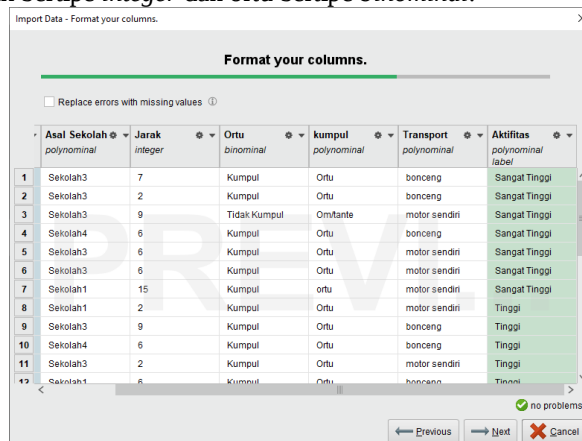
- 1) Memasukkan data siswa dan atribut faktor eksternal formal xls.

Pada tahap ini dilakukan untuk memasukkan data dalam bentuk file excel ke dalam aplikasi *RapidMiner Stodio*



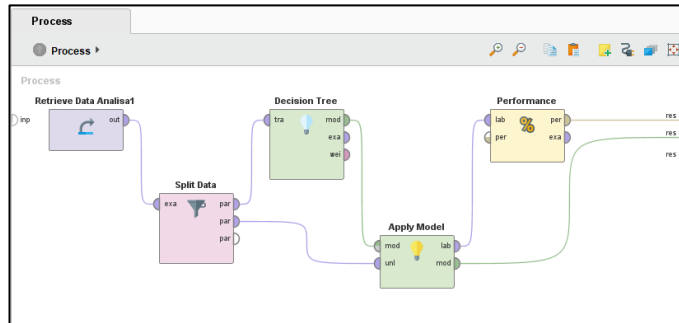
Gambar 1. Memasukkan data siswa dan atributnya

- 2) Menyesuaikan data atribut untuk *data type*, *role* dan *lable/class*
Pada tahap ini dilakukan penyesuaian beberapa atribut, yaitu nama siswa menjadi *role id*, aktifitas menjadi *role label*, asal sekolah, kumpul, transportasi dan aktifitas bertipe *polynomial*, jarak bertipe *integer* dan ortu bertipe *binominal*.



Gambar 2. Menyesuaikan data atribut

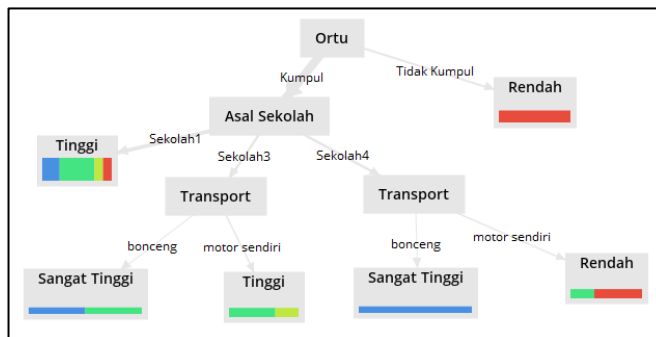
- 3) *Drag and drop dataset* dan *operator* ke dalam *process design*
Pada tahap ini memasukkan ke dalam *process design* berupa dataset, yaitu tabel siswa yang telah diinput sebelumnya dan operator berupa *split data* untuk membagi data, algoritma *decision tree* untuk membuat model, *apply model* untuk implementasi model, dan *performance* untuk melihat performa dari model,
- 4) Melakukan koneksi antara dataset dengan *operator*
Koneksi dilakukan untuk menghubungkan antara dataset dengan operator melalui output dari dataset ke bagian input operator. Demikian juga output dari operator dihubungkan dengan input operator lain seperti tampak pada Gambar 3.



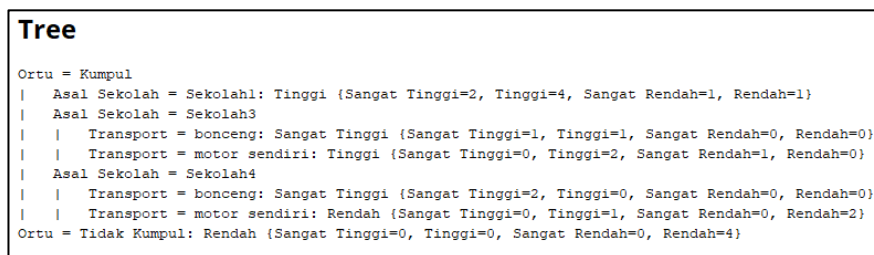
Gambar 3. Koneksi dataset dengan operator

- 5) Melakukan *running/start the execution* terhadap koneksi data dan *operator decision tree*

Pada tahap selanjutnya adalah melakukan *running* atau eksekusi terhadap dataset dan operator. Proses *running*, seperti terlihat pada Gambar 3, *output* dari dataset masuk ke operator *split data* untuk dipilah secara *random* antara data *training* dan data prediksi. Selanjutnya data training akan diproses oleh operator *decision tree* sedangkan data prediksi diproses di operator *apply model* setelah mendapat masukan model dari operator *decision tree*. Keluaran dari operator *apply model* menjadi masukan operator *performance* untuk menghitung nilai performa dari model setelah dilakukan implementasi terhadap data prediksi serta juga menjadi keluaran untuk menampilkan model *decision tree*. Hasil dari tahapan ini seperti tampak pada Gambar 4, Gambar 5, Gambar 6 dan Gambar 7.



Gambar 4. Model grafik *decision tree*



Gambar 5. Deskripsi *decision tree*

e. Evaluation

Model pada *decision tree* perlu dilihat performannya. Performa model pada penelitian ini menggunakan *confusion matrix* (**Gambar 6**) yang berisi jumlah dari *record* yang diprediksi benar dan salah melalui model *decision tree*. Performa model dilihat dari nilai akurasi implementasi model berupa persentase ketepatan memprediksi dengan membandingkan angka diagonal pada *confusion matrix*, berupa jumlah *record* yang diprediksi benar, dengan jumlah keseluruhan *record* dari data prediksi. Pada penelitian ini akurasi model adalah 85.71% memiliki kedekatan antara data sebenarnya dengan data prediksi. Selain presentase akurasi juga didapatkan presentase *recall*, sebesar 66.67% berupa performa model dalam menemukan informasi di dalam data prediksi dan presentase presisi sebesar 66.67% berupa ketepatan model dalam memprediksi data prediksi. Untuk lebih jelasnya seperti tampak pada **Gambar 7**.

	true Sangat Tinggi	true Tinggi	true Sangat Rendah	true Rendah	class precision
pred. Sangat Tinggi	2	1	0	0	66.67%
pred. Tinggi	0	2	0	0	100.00%
pred. Sangat Rend...	0	0	0	0	0.00%
pred. Rendah	0	0	0	2	100.00%
class recall	100.00%	66.67%	0.00%	100.00%	

Gambar 6. Confusion Matrix

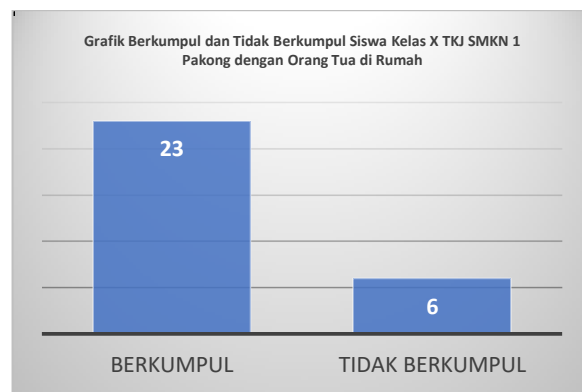
PerformanceVector				
PerformanceVector:				
accuracy: 85.71%				
ConfusionMatrix:				
True:	Sangat Tinggi	Tinggi	Sangat Rendah	Rendah
Sangat Tinggi:	2	1	0	0
Tinggi: 0	2	0	0	0
Sangat Rendah:	0	0	0	0
Rendah: 0	0	0	0	2
weighted_mean_recall: 66.67%, weights: 1, 1, 1, 1				
ConfusionMatrix:				
True:	Sangat Tinggi	Tinggi	Sangat Rendah	Rendah
Sangat Tinggi:	2	1	0	0
Tinggi: 0	2	0	0	0
Sangat Rendah:	0	0	0	0
Rendah: 0	0	0	0	2
weighted_mean_precision: 66.67%, weights: 1, 1, 1, 1				
ConfusionMatrix:				
True:	Sangat Tinggi	Tinggi	Sangat Rendah	Rendah
Sangat Tinggi:	2	1	0	0
Tinggi: 0	2	0	0	0
Sangat Rendah:	0	0	0	0
Rendah: 0	0	0	0	2

Gambar 7. Performance Model

f. Deployment

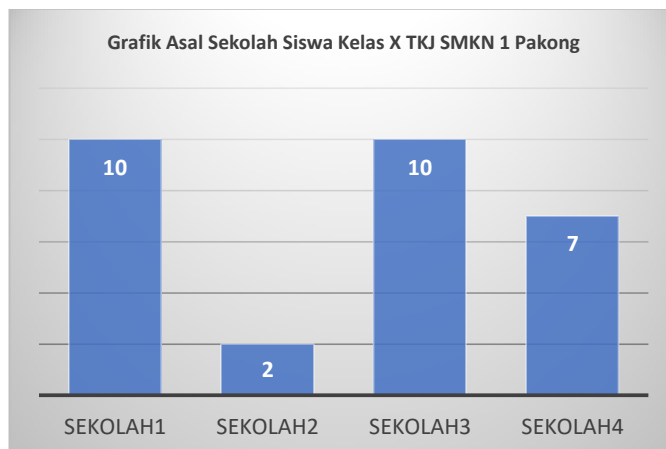
Pada **Gambar 4** terlihat bahwa faktor eksternal tidak berkumpul dengan orang tua menjadi faktor utama rendahnya aktifitas siswa kelas X TKJ SMKN 1 Pakong. Bagi siswa yang berkumpul dengan orang tua di rumah maka dipengaruhi oleh asal sekolah dan transportasi siswa saat berangkat ke sekolah. Asal sekolah siswa kelas X TKJ SMKN 1 Pakong baik sekolah1 dan sekolah2 dengan mode transportasi yang ada memiliki kecenderungan aktifitas tinggi bahkan terdapat sangat tinggi, yaitu transportasi bonceng. Asal sekolah sekolah4 pada mode transportasi motor sendiri maka aktifitas siswa rendah sedangkan pada mode transportasi bonceng maka aktifitas sangat tinggi. Deskripsi dari model grafik decision tree seperti terlihat pada **Gambar 5**.

Jumlah siswa kelas X TKJ SMKN 1 Pakong yang tidak berkumpul dengan orang tua lebih sedikit jika dibandingkan dengan siswa yang berkumpul dengan orang tuanya, yaitu sebanyak 6 siswa dari 29 siswa atau sebesar 20.69%. Hal ini seperti terlihat pada **Gambar 8** berupa grafik perbandingan antara jumlah siswa yang berkumpul ataupun tidak berkumpul dengan tingkat aktifitasnya. Dengan demikian berdasarkan model *decision tree* (**Gambar 4**) terdapat 20.69% siswa yang memiliki aktifitas rendah.



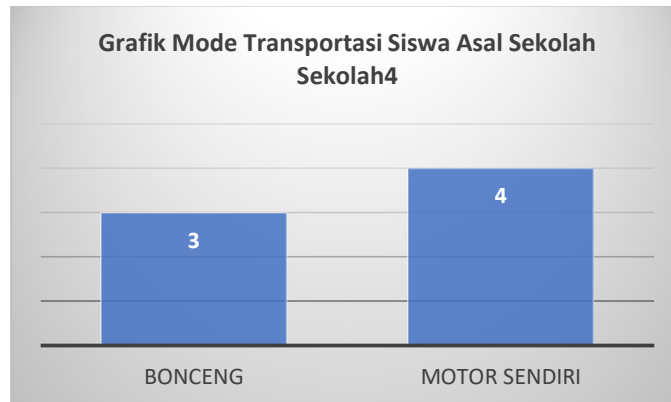
Gambar 8. Grafik berkumpul dan tidak berkumpul siswa dengan orang tua di rumah

Selanjutnya asal sekolah siswa kelas X TKJ SMKN 1 Pakong dari sekolah4 yang memiliki kemungkinan tingkat aktifitas rendah berdasarkan model *decision tree* (**Gambar 4**), yaitu sebanyak 7 siswa atau 24.14% dari keseluruhan siswa. Hal ini seperti terlihat pada **Gambar 9**.

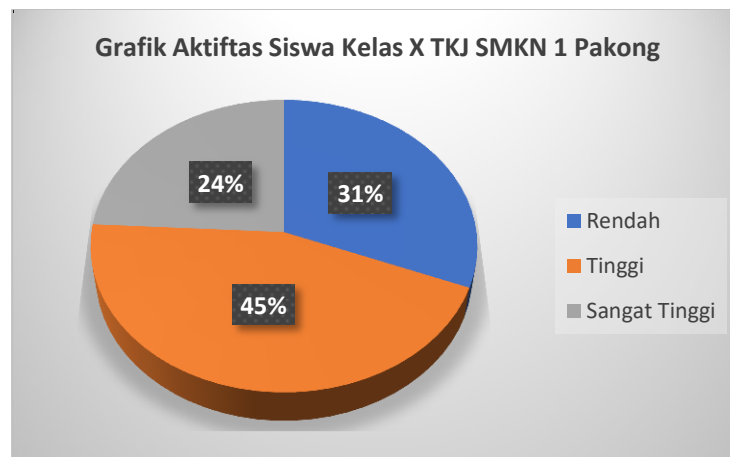


Gambar 9. Grafik asal sekolah siswa

Namun demikian dari 7 siswa tersebut mode transportasi menjadi faktor eksternal rendahnya aktifitas siswa dimana terdapat dua mode transportasi siswa asal sekolah sekolah4, yaitu bonceng dan motor sendiri. Dari dua mode transportasi tersebut yang memiliki tingkat aktifitas rendah adalah mode transportasi membawa motor sendiri, yaitu sebanyak 4 siswa atau sebesar 57.14% atau lebih besar dari siswa dengan mode transportasi bonceng seperti terlihat pada **Gambar 10**. Berdasarkan model *decision tree* dengan berbagai faktor eksternal yang dimiliki oleh siswa kelas X TKJ SMKN 1 Pakong maka tingkat aktifitasnya dapat di kelompokkan menjadi 3, yait rendah sebesar 31%, tinggi sebesar 45% dan sangat tinggi 24% seperti tampak pada **Gambar 11**.



Gambar 10. Grafik mode transportasi siswa dengan asal sekolah sekolah4



Gambar 10. Grafik aktifitas siswa

4. KESIMPULAN

Model yang dihasilkan dari implementasi algoritma *decision tree* untuk melihat faktor eksternal yang mempengaruhi aktifitas siswa kelas X TKJ SMKN 1 Pakong adalah tidak berkumpulnya siswa dengan orang tua di rumah menjadi faktor utama. Sementara bagi siswa yang berkumpul dengan orang tuanya di rumah aktifitasnya akan tinggi jika berasal dari sekolah1. Siswa yang berkumpul dengan orang tuanya di rumah aktifitasnya bisa rendah apabila berasal dari sekolah sekolah4 dan menggunakan mode transportasi motor sendiri namun bisa sangat tinggi apabila berboncengan dengan temannya. Kemudian untuk siswa yang berkumpul dengan orang tuanya di rumah yang berasal dari sekolah sekolah3 akan tinggi apabila menggunakan motor

sendiri bahkan sangat apabila berboncengan dengan temannya. Model tersebut memiliki nilai akurasi 85.71%, *recall* 66.67% dan presisi 66.67%.

Hasil analisa terhadap faktor eksternal siswa kelas X TKJ SMKN 1 Pakong didapat bahwa 31% siswa dengan aktifitas rendah, 45% siswa dengan aktifitas tinggi dan 24% siswa dengan aktifitas sangat tinggi.

DAFTAR PUSTAKA

- [1] J. A. Zaeni and F. F. Hidayah, "Analisis Keaktifan Siswa Melalui Penerapan Model Teams Game Tournamants (TGT) pada Materi Termokimia Kelas XI IPA 5 di SMA 15 Semarang," *Semin. Nas. Pendidikan, Sains dan Teknol. Fak. Mat. dan Ilmu Pengetah. Alam Univ. Muhammadiyah Semarang*, vol. 8, pp. 416–425, 2002, [Online]. Available: <https://jurnal.unimus.ac.id/index.php/psn120120/article/view/3086>.
- [2] M. Muhsi and S. Kamariyah, "Analisa Penggunaan Media Pembelajaran untuk Materi Mendiagnosa Permasalahan Perangkat yang Terhubung dengan Wide Area Network (WAN) Kelas XI-TKJ di SMK Negeri 1 Pakong Pamekasan," in *Seminar Nasional Humaniora dan Aplikasi Teknologi Informasi*, 2018, vol. 4, no. 1, pp. 589–594.
- [3] M. Achdiyat and K. D. Lestari, "Prestasi Belajar Matematika Ditinjau dari Kepercayaan Diri dan Keaktifan Siswa di Kelas," *Form. J. Ilm. Pendidik. MIPA*, vol. 6, no. 1, pp. 50–61, 2016, doi: 10.30998/formatif.v6i1.752.
- [4] L. E. Walker, M. Cross, and T. Barnett, "Students' Experiences and Perceptions of Interprofessional Education During Rural Placement: A Mixed Methods Study," *Nurse Educ. Today*, vol. 75, pp. 28–34, 2019, doi: 10.1016/j.nedt.2018.12.012.
- [5] S. Bilican, O. Kutlu, and O. Yildirim, "The Factors that Predict the Frequency of Activities Developing Students' Listening Comprehension Skills," *Procedia - Soc. Behav. Sci.*, vol. 46, pp. 5219–5224, 2012, doi: 10.1016/j.sbspro.2012.06.413.
- [6] Z. Amry and L. Badriah, "Pembelajaran Tematik Sebagai Upaya Meningkatkan Keaktifan Peserta Didik," *Elem. Islam. Teach. J.*, vol. 6, no. 2, pp. 254–270, 2018.
- [7] R. Mardiyani, "Peningkatan Keaktifan dan Hasil Belajar Siswa dalam Pembelajaran Akuntansi Materi Jurnal Penyesuaian Pada Siswa Kelas XI IPS 3 SMA Negeri 3 Bukittinggi Dengan Metode Bermain Peran (Role Playing)," *Pakar Pendidik.*, vol. 10, no. 2, pp. 151–162, 2012.
- [8] K. Rikawati and D. Sitinjak, "Peningkatan Keaktifan Belajar Siswa dengan Penggunaan Metode Ceramah Interaktif," *J. Educ. Chem.*, vol. 2, no. 2, p. 40, 2020, doi: 10.21580/jec.2020.2.2.6059.
- [9] M. Rizkina, "Upaya Meningkatkan Keaktifan Siswa Dalam Diskusi Kelompok Melalui Layanan Bimbingan Kelompok Pada Siswa Kelas Viii Di Smpn I9 Semarang Skripsi," Universitas Negeri Semarang, 2013.
- [10] A. Al-Ansari, F. Al-Harbi, W. AbdelAziz, M. AbdelSalam, M. M. El Tantawi, and I. ElRefae, "Factors affecting student participation in extra-curricular activities: A comparison between two Middle Eastern dental schools," *Saudi Dent. J.*, vol. 28, no. 1, pp. 36–43, 2016, doi: 10.1016/j.sdentj.2015.05.004.
- [11] H. Cakir, "Use of Blogs in Pre-Service Teacher Education to Improve Student Engagement," *Comput. Educ.*, vol. 68, pp. 244–252, 2013, doi: 10.1016/j.compedu.2013.05.013.
- [12] M. Syah, *Psikologi Belajar*. Jakarta: Raja Grafindo Persada, 2012.

- [13] S. Agarwal, *Data mining: Data mining concepts and techniques*, 3rd ed. Morgan Kaufmann, 2014.
- [14] F. E. Fahrudi and H. W. Wiryosutomo, "Analisis Faktor-Faktor Penyebab Perilaku Malas Belajar Daring Saat Pandemi Covid-19 Pada Siswa Kelas Xii Sma Negeri 1 Menganti Gresik," *J. BK UNESA*, vol. 12, no. 2, pp. 22–36, 2021.
- [15] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017, doi: 10.15575/join.v2i1.71.
- [16] F. Handayani, S. Sulistyarini, and O. Okianna, "PENGARUH FAKTOR LINGKUNGAN BELAJAR SEKOLAH TERHADAP MOTIVASI BELAJAR SISWA PADA MATA PELAJARAN EKONOMI DI SMA," *J. Pendidik. dan Pembelajaran Khatulistiwa*, vol. 7, no. 9, pp. 1–10, 2018.
- [17] M. M. Ali, "Role of Data Mining in Education Sector," *IJCSMC J.*, vol. 2, no. 4, pp. 374–383, 2013.
- [18] K. Hastuti and E. Y. Hidayat, "Analisis Algoritma Decision Tree untuk Prediksi Mahasiswa Non Aktif," *Semantik*, vol. 2013, no. November, pp. 211–216, 2013.
- [19] R. K. Niswatin and R. Wulanningrum, "Penerapan Algoritma Decision Tree Pada Penentuan Keberhasilan Akademik Mahasiswa," *Semin. Nas. Teknol. Inf. dan Multimed.* 2017, pp. 2–7, 2017.
- [20] Sianturi, Ariwisanto, and Fricles, "Analisa Decision Tree Dalam Pengolahan Data Siswa," *MEANS (Media Inf. Anal. dan Sist.*, vol. 3, no. 2, pp. 166–172, 2018, [Online]. Available: http://ejournal.ust.ac.id/index.php/Jurnal_Means/.
- [21] I. Sutoyo, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," *J. Pilar Nusa Mandiri*, vol. 14, no. 2, p. 217, 2018, doi: 10.33480/pilar.v14i2.926.
- [22] O. L. David and D. Delen, *Advanced Data Mining Technique*. Springer, 2008.
- [23] M. Tiwari, S. Randhir, and V. Neeraj, "An Empirical Study of Application of Data Mining Techniques for Predicting Student Performance in Higher Education," *Int. J. Comput. Sci. Mob. Comput.*, vol. 2, no. 2, pp. 53–57, 2013.
- [24] T.-C. Hsia, A.-J. Shie, and L.-C. Chen, "Course Planning of extension education to meet market demand by using data mining technique: an example of Chinkuo Technical University in Taiwan," in *Expert System with Applications*, 2008, pp. 596–602.
- [25] J. A. M. Berry and S. G. Linoff, *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management*, Second Edi. Wiley Publishing, 2004.
- [26] S. Tuffery, *Data Mining and Statistics for Decision making*. John Wiley and Sons Ltd., 2011.
- [27] H. ian Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edi. Morgan Kauffmann, 2005.
- [28] R. S. Wahono, "Materi 1 D : Data Mining," 2020.
- [29] R. S. Wahono, "Materi 1D : Knowledge Management," 2021.
- [30] P. Giudici, *Applied Data Mining: Statistical Methods for Business and Industry*. John Wiley and Sons Ltd., 2003.

- [31] M. Kantardzic, *Data Mining: Concepts, Models, Methods and Algorithm*, Second Edi. New Jersey: John Wiley and Sons Ltd., 2011.
- [32] H.-H. Hsu, *Advanced Data Mining Technologies in Bioinformatics*, Idea Group Inc, 2006, 1-12. United States: Ide Group Publishing, 2006.
- [33] E. S. Rahayu, R. S. Wahono, and C. Supriyanto, "Penerapan Metode Average Gain, Threshold Pruning dan Cost Complexity Pruning Untuk Split Atribut Pada Algoritma C4.5," *J. Intell. Syst.*, vol. 1, no. 2, pp. 91–97, 2015.
- [34] B. Novianti, T. Rismawan, and S. Bahri, "Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak)," *J. Coding, Sist. Komput. Untan*, vol. 04, no. 3, pp. 75–84, 2016.
- [35] M. K. Happe, Y. Salin, and P. Purnawansyah, "Perbandingan Hasil Prediksi Bandwidth Management menggunakan Metode Naive Bayes Classifier dan Decision Tree C4. 5," *Pros. SAKTI (Seminar Ilmu Komput. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 134–140, 2018.