

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1 Analisis Data**

Analisis data dalam penelitian ini dilakukan untuk mengetahui hubungan antara pola belajar siswa dengan hasil belajar Bahasa Inggris menggunakan pendekatan data mining melalui algoritma Decision Tree C4.5. Data yang digunakan berisi informasi siswa kelas X SMK Swasta Al-Washliyah 2 Marbau dengan variabel Absensi (X1), Kelas Tambahan (X2), Penggunaan Media Digital (X3), dan Nilai (Y). Variabel absensi dikategorikan menjadi  $\geq 7$  dan  $< 7$ , sedangkan nilai diklasifikasikan ke dalam kategori KKM dan Diatas KKM. Sebelum proses analisis dilakukan, data terlebih dahulu melalui tahap preprocessing seperti pemeriksaan kelengkapan data, transformasi nilai ke dalam bentuk kategorikal, serta pemisahan data menjadi data training dan data testing. Selanjutnya, data training digunakan untuk membangun model pohon keputusan dengan menghitung nilai entropy, information gain, split information, dan gain ratio guna menentukan atribut yang paling berpengaruh terhadap hasil belajar. Melalui proses ini, diperoleh pola hubungan antar variabel yang kemudian digunakan untuk mengklasifikasikan dan memprediksi hasil belajar siswa secara lebih objektif dan berbasis data.

#### **4.2 Pengumpulan Data**

Pengumpulan data dalam penelitian ini dilakukan dengan mengambil data akademik siswa kelas X SMK Swasta Al-Washliyah 2 Marbau yang berkaitan dengan proses pembelajaran Bahasa Inggris. Data yang dikumpulkan meliputi informasi mengenai tingkat kehadiran siswa (absensi), keikutsertaan dalam kelas

tambahan, penggunaan media digital dalam belajar, serta nilai akhir Bahasa Inggris sebagai indikator hasil belajar. Data tersebut diperoleh dari dokumentasi sekolah dan catatan pembelajaran yang telah tersedia, sehingga termasuk dalam jenis data sekunder. Seluruh data yang digunakan merupakan data yang lengkap dan relevan dengan variabel penelitian, sehingga dapat mendukung proses analisis menggunakan algoritma Decision Tree C4.5.

**Tabel 4.1 Data Sampel Penelitian**

Nama	Absensi (X1)	Kelas Tambahan (X2)	Media Digital (X3)	Nilai (Y)
S01	3	Ya	Tidak	80
S02	8	Tidak	Ya	77
S03	5	Ya	Tidak	77
S04	7	Ya	Tidak	77
S05	2	Ya	Ya	79
S06	4	Tidak	Tidak	77
S07	8	Tidak	Ya	77
S08	4	Ya	Tidak	77
S09	7	Ya	Ya	77
S10	2	Tidak	Tidak	80
S11	3	Ya	Tidak	77
S12	7	Tidak	Tidak	77
S13	5	Ya	Tidak	77
S14	7	Ya	Ya	77
S15	3	Ya	Tidak	81
S16	3	Ya	Ya	79
S17	8	Ya	Tidak	77
S18	4	Ya	Ya	78
S19	9	Tidak	Ya	77
S20	3	Tidak	Tidak	77
S21	4	Ya	Ya	80
S22	8	Ya	Tidak	77
S23	2	Tidak	Ya	77
S24	7	Ya	Ya	77
S25	5	Ya	Tidak	77
S26	8	Tidak	Tidak	77
S27	2	Ya	Ya	79
S28	4	Tidak	Tidak	77
S29	8	Ya	Ya	77
S30	3	Tidak	Tidak	77
S31	0	Ya	Tidak	78
S32	3	Ya	Ya	82
S33	4	Ya	Ya	82
S34	0	Ya	Ya	82

S35	11	Tidak	Ya	77
S36	3	Ya	Ya	80
S37	6	Tidak	Ya	79
S38	3	Ya	Ya	80
S39	4	Ya	Ya	80
S40	8	Tidak	Ya	77
S41	10	Tidak	Ya	77
S42	0	Tidak	Ya	78
S43	1	Ya	Ya	81
S44	1	Tidak	Tidak	77
S45	4	Tidak	Ya	77
S46	4	Tidak	Ya	78
S47	5	Tidak	Ya	79
S48	3	Ya	Ya	80
S49	6	Ya	Ya	80
S50	6	Tidak	Ya	78
S51	11	Ya	Tidak	77
S52	1	Ya	Tidak	80
S53	5	Tidak	Ya	78
S54	0	Tidak	Ya	78
S55	2	Tidak	Ya	78
S56	2	Ya	Ya	79
S57	0	Ya	Ya	80
S58	5	Tidak	Ya	77
S59	3	Ya	Ya	80
S60	0	Ya	Ya	83
S61	15	Tidak	Tidak	77
S62	6	Tidak	Ya	77
S63	6	Tidak	Ya	77
S64	8	Ya	Ya	80
S65	4	Ya	Ya	79
S66	2	Tidak	Ya	78
S67	6	Ya	Ya	79
S68	7	Ya	Ya	79
S69	0	Tidak	Ya	81
S70	12	Tidak	Ya	77
S71	3	Tidak	Ya	79
S72	3	Ya	Ya	80
S73	5	Ya	Ya	80
S74	0	Tidak	Ya	79
S75	4	Ya	Tidak	78
S76	4	Ya	Ya	80
S77	2	Ya	Ya	82
S78	0	Tidak	Ya	79
S79	2	Tidak	Ya	79
S80	4	Ya	Ya	79
S81	5	Ya	Ya	80
S82	2	Ya	Ya	82
S83	1	Ya	Ya	82
S84	3	Ya	Ya	84
S85	3	Ya	Ya	81

S86	21	Tidak	Ya	77
S87	4	Ya	Ya	81
S88	4	Ya	Ya	82
S89	5	Ya	Tidak	84
S90	4	Ya	Ya	81
S91	4	Ya	Ya	81
S92	5	Ya	Ya	80
S93	8	Ya	Ya	80
S94	2	Ya	Ya	81
S95	3	Ya	Ya	80
S96	4	Ya	Ya	81
S97	6	Ya	Ya	81
S98	3	Ya	Ya	81
S99	5	Ya	Ya	80
S100	4	Ya	Ya	82
S101	3	Ya	Ya	83
S102	5	Tidak	Tidak	77
S103	1	Ya	Ya	82
S104	3	Ya	Tidak	79
S105	4	Tidak	Ya	79
S106	9	Tidak	Tidak	77
S107	3	Ya	Tidak	78
S108	2	Tidak	Ya	79
S109	2	Ya	Tidak	80
S110	2	Ya	Ya	80
S111	2	Tidak	Ya	78
S112	4	Ya	Tidak	79
S113	3	Ya	Ya	81
S114	4	Ya	Ya	80
S115	5	Tidak	Ya	78

Tabel tersebut berisi data 115 siswa yang mencakup beberapa variabel, yaitu absensi (X1), kelas tambahan (X2), media digital (X3), dan nilai (Y). Variabel absensi menunjukkan jumlah ketidakhadiran siswa yang berkisar antara 0 hingga 21 kali, yang berarti ada siswa yang selalu hadir dan ada juga yang cukup sering tidak hadir. Variabel kelas tambahan dan media digital masing-masing menunjukkan apakah siswa mengikuti kelas tambahan dan menggunakan media digital dalam proses belajar dengan pilihan jawaban “Ya” atau “Tidak”. Sementara itu, nilai siswa berada pada rentang 77 hingga 84, dengan sebagian besar nilai berada di kisaran 77 sampai 80. Secara umum, terlihat bahwa banyak siswa yang mengikuti kelas tambahan dan menggunakan media digital

memperoleh nilai yang cukup baik. Selain itu, siswa dengan tingkat absensi yang rendah cenderung memiliki nilai yang lebih tinggi dibandingkan siswa dengan absensi yang tinggi. Hal ini menunjukkan bahwa kehadiran, keikutsertaan dalam kelas tambahan, dan penggunaan media digital memiliki hubungan dengan hasil belajar siswa.

### 4.3 Preprocessing Data

#### Tahap 1 Preprocessing Data dan Transformasi Data

**Tabel 4.2 Data Preprocessing dan Transformasi**

Nama	Absensi (X1)	Kelas Tambahan (X2)	Media Digital (X3)	Nilai (Y)
S01	<7	Ya	Tidak	Diatas KKM
S02	≥7	Tidak	Ya	KKM
S03	<7	Ya	Tidak	KKM
S04	≥7	Ya	Tidak	KKM
S05	<7	Ya	Ya	Diatas KKM
S06	<7	Tidak	Tidak	KKM
S07	≥7	Tidak	Ya	KKM
S08	<7	Ya	Tidak	KKM
S09	7	Ya	Ya	KKM
S10	<7	Tidak	Tidak	Diatas KKM
S11	<7	Ya	Tidak	KKM
S12	≥7	Tidak	Tidak	KKM
S13	<7	Ya	Tidak	KKM
S14	≥7	Ya	Ya	KKM
S15	<7	Ya	Tidak	Diatas KKM
S16	<7	Ya	Ya	Diatas KKM
S17	≥7	Ya	Tidak	KKM
S18	<7	Ya	Ya	Diatas KKM
S19	≥7	Tidak	Ya	KKM
S20	<7	Tidak	Tidak	KKM
S21	<7	Ya	Ya	Diatas KKM
S22	≥7	Ya	Tidak	KKM
S23	<7	Tidak	Ya	KKM
S24	≥7	Ya	Ya	KKM
S25	<7	Ya	Tidak	KKM
S26	≥7	Tidak	Tidak	KKM
S27	<7	Ya	Ya	Diatas KKM
S28	<7	Tidak	Tidak	KKM
S29	≥7	Ya	Ya	KKM
S30	<7	Tidak	Tidak	KKM

S31	<7	Ya	Tidak	Diatas KKM
S32	<7	Ya	Ya	Diatas KKM
S33	<7	Ya	Ya	Diatas KKM
S34	<7	Ya	Ya	Diatas KKM
S35	$\geq 7$	Tidak	Ya	KKM
S36	<7	Ya	Ya	Diatas KKM
S37	<7	Tidak	Ya	Diatas KKM
S38	<7	Ya	Ya	Diatas KKM
S39	<7	Ya	Ya	Diatas KKM
S40	$\geq 7$	Tidak	Ya	KKM
S41	$\geq 7$	Tidak	Ya	KKM
S42	<7	Tidak	Ya	Diatas KKM
S43	<7	Ya	Ya	Diatas KKM
S44	<7	Tidak	Tidak	KKM
S45	<7	Tidak	Ya	KKM
S46	<7	Tidak	Ya	Diatas KKM
S47	<7	Tidak	Ya	Diatas KKM
S48	<7	Ya	Ya	Diatas KKM
S49	<7	Ya	Ya	Diatas KKM
S50	<7	Tidak	Ya	Diatas KKM
S51	$\geq 7$	Ya	Tidak	KKM
S52	<7	Ya	Tidak	Diatas KKM
S53	<7	Tidak	Ya	Diatas KKM
S54	<7	Tidak	Ya	Diatas KKM
S55	<7	Tidak	Ya	Diatas KKM
S56	<7	Ya	Ya	Diatas KKM
S57	<7	Ya	Ya	Diatas KKM
S58	<7	Tidak	Ya	KKM
S59	<7	Ya	Ya	Diatas KKM
S60	<7	Ya	Ya	Diatas KKM
S61	$\geq 7$	Tidak	Tidak	KKM
S62	<7	Tidak	Ya	KKM
S63	<7	Tidak	Ya	KKM
S64	$\geq 7$	Ya	Ya	Diatas KKM
S65	<7	Ya	Ya	Diatas KKM
S66	<7	Tidak	Ya	Diatas KKM
S67	<7	Ya	Ya	Diatas KKM
S68	$\geq 7$	Ya	Ya	Diatas KKM
S69	<7	Tidak	Ya	Diatas KKM
S70	$\geq 7$	Tidak	Ya	KKM
S71	<7	Tidak	Ya	Diatas KKM
S72	<7	Ya	Ya	Diatas KKM
S73	<7	Ya	Ya	Diatas KKM
S74	<7	Tidak	Ya	Diatas KKM
S75	<7	Ya	Tidak	Diatas KKM
S76	<7	Ya	Ya	Diatas KKM
S77	<7	Ya	Ya	Diatas KKM

S78	<7	Tidak	Ya	Diatas KKM
S79	<7	Tidak	Ya	Diatas KKM
S80	<7	Ya	Ya	Diatas KKM
S81	<7	Ya	Ya	Diatas KKM
S82	<7	Ya	Ya	Diatas KKM
S83	<7	Ya	Ya	Diatas KKM
S84	<7	Ya	Ya	Diatas KKM
S85	<7	Ya	Ya	Diatas KKM
S86	$\geq 7$	Tidak	Ya	KKM
S87	<7	Ya	Ya	Diatas KKM
S88	<7	Ya	Ya	Diatas KKM
S89	<7	Ya	Tidak	Diatas KKM
S90	<7	Ya	Ya	Diatas KKM
S91	<7	Ya	Ya	Diatas KKM
S92	<7	Ya	Ya	Diatas KKM
S93	$\geq 7$	Ya	Ya	Diatas KKM
S94	<7	Ya	Ya	Diatas KKM
S95	<7	Ya	Ya	Diatas KKM
S96	<7	Ya	Ya	Diatas KKM
S97	<7	Ya	Ya	Diatas KKM
S98	<7	Ya	Ya	Diatas KKM
S99	<7	Ya	Ya	Diatas KKM
S100	<7	Ya	Ya	Diatas KKM
S101	<7	Ya	Ya	Diatas KKM
S102	<7	Tidak	Tidak	KKM
S103	<7	Ya	Ya	Diatas KKM
S104	<7	Ya	Tidak	Diatas KKM
S105	<7	Tidak	Ya	Diatas KKM
S106	$\geq 7$	Tidak	Tidak	KKM
S107	<7	Ya	Tidak	Diatas KKM
S108	<7	Tidak	Ya	Diatas KKM
S109	<7	Ya	Tidak	Diatas KKM
S110	<7	Ya	Ya	Diatas KKM
S111	<7	Tidak	Ya	Diatas KKM
S112	<7	Ya	Tidak	Diatas KKM
S113	<7	Ya	Ya	Diatas KKM
S114	<7	Ya	Ya	Diatas KKM
S115	<7	Tidak	Ya	Diatas KKM

Tabel tersebut berisi data siswa yang diklasifikasikan berdasarkan absensi (X1), kelas tambahan (X2), media digital (X3), dan hasil belajar (Y), di mana absensi dibagi menjadi dua kategori yaitu kurang dari 7 kali (<7) dan lebih dari atau sama dengan 7 kali ( $\geq 7$ ), sementara kelas tambahan dan media digital menunjukkan apakah siswa mengikuti atau menggunakan (Ya/Tidak), serta hasil

belajar dibedakan menjadi “KKM” dan “Di atas KKM”. Secara umum, mayoritas siswa memiliki absensi kurang dari 7 kali yang menunjukkan kehadiran yang baik, dan banyak di antara mereka yang mengikuti kelas tambahan serta menggunakan media digital cenderung memperoleh hasil “Di atas KKM”. Sebaliknya, siswa dengan absensi tinggi ( $\geq 7$ ) dan yang tidak mengikuti kelas tambahan lebih banyak berada pada kategori “KKM”, sehingga dapat disimpulkan bahwa kehadiran, keikutsertaan dalam kelas tambahan, dan penggunaan media digital memiliki hubungan terhadap hasil belajar siswa.

## Tahap 2 Split Data

**Tabel 4.2 Data Training**

Nama	Absensi (X1)	Kelas Tambahan (X2)	Media Digital (X3)	Nilai (Y)
S01	<7	Ya	Tidak	Diatas KKM
S02	$\geq 7$	Tidak	Ya	KKM
S03	<7	Ya	Tidak	KKM
S04	$\geq 7$	Ya	Tidak	KKM
S05	<7	Ya	Ya	Diatas KKM
S06	<7	Tidak	Tidak	KKM
S07	$\geq 7$	Tidak	Ya	KKM
S08	<7	Ya	Tidak	KKM
S09	$\geq 7$	Ya	Ya	KKM
S10	<7	Tidak	Tidak	Diatas KKM
S11	<7	Ya	Tidak	KKM
S12	$\geq 7$	Tidak	Tidak	KKM
S13	<7	Ya	Tidak	KKM
S14	$\geq 7$	Ya	Ya	KKM
S15	<7	Ya	Tidak	Diatas KKM
S16	<7	Ya	Ya	Diatas KKM
S17	$\geq 7$	Ya	Tidak	KKM
S18	<7	Ya	Ya	Diatas KKM
S19	$\geq 7$	Tidak	Ya	KKM
S20	<7	Tidak	Tidak	KKM
S21	<7	Ya	Ya	Diatas KKM
S22	$\geq 7$	Ya	Tidak	KKM
S23	<7	Tidak	Ya	KKM
S24	$\geq 7$	Ya	Ya	KKM
S25	<7	Ya	Tidak	KKM

S26	$\geq 7$	Tidak	Tidak	KKM
S27	$< 7$	Ya	Ya	Diatas KKM
S28	$< 7$	Tidak	Tidak	KKM
S29	$\geq 7$	Ya	Ya	KKM
S30	$< 7$	Tidak	Tidak	KKM
S31	$< 7$	Ya	Tidak	Diatas KKM
S32	$< 7$	Ya	Ya	Diatas KKM
S33	$< 7$	Ya	Ya	Diatas KKM
S34	$< 7$	Ya	Ya	Diatas KKM
S35	$\geq 7$	Tidak	Ya	KKM
S36	$< 7$	Ya	Ya	Diatas KKM
S37	$< 7$	Tidak	Ya	Diatas KKM
S38	$< 7$	Ya	Ya	Diatas KKM
S39	$< 7$	Ya	Ya	Diatas KKM
S40	$\geq 7$	Tidak	Ya	KKM
S41	$\geq 7$	Tidak	Ya	KKM
S42	$< 7$	Tidak	Ya	Diatas KKM
S43	$< 7$	Ya	Ya	Diatas KKM
S44	$< 7$	Tidak	Tidak	KKM
S45	$< 7$	Tidak	Ya	KKM
S46	$< 7$	Tidak	Ya	Diatas KKM
S47	$< 7$	Tidak	Ya	Diatas KKM
S48	$< 7$	Ya	Ya	Diatas KKM
S49	$< 7$	Ya	Ya	Diatas KKM
S50	$< 7$	Tidak	Ya	Diatas KKM
S51	$\geq 7$	Ya	Tidak	KKM
S52	$< 7$	Ya	Tidak	Diatas KKM
S53	$< 7$	Tidak	Ya	Diatas KKM
S54	$< 7$	Tidak	Ya	Diatas KKM
S55	$< 7$	Tidak	Ya	Diatas KKM
S56	$< 7$	Ya	Ya	Diatas KKM
S57	$< 7$	Ya	Ya	Diatas KKM
S58	$< 7$	Tidak	Ya	KKM
S59	$< 7$	Ya	Ya	Diatas KKM
S60	$< 7$	Ya	Ya	Diatas KKM
S61	$\geq 7$	Tidak	Tidak	KKM
S62	$< 7$	Tidak	Ya	KKM
S63	$< 7$	Tidak	Ya	KKM
S64	$\geq 7$	Ya	Ya	Diatas KKM
S65	$< 7$	Ya	Ya	Diatas KKM
S66	$< 7$	Tidak	Ya	Diatas KKM
S67	$< 7$	Ya	Ya	Diatas KKM
S68	$\geq 7$	Ya	Ya	Diatas KKM
S69	$< 7$	Tidak	Ya	Diatas KKM
S70	$\geq 7$	Tidak	Ya	KKM

S71	<7	Tidak	Ya	Diatas KKM
S72	<7	Ya	Ya	Diatas KKM
S73	<7	Ya	Ya	Diatas KKM
S74	<7	Tidak	Ya	Diatas KKM
S75	<7	Ya	Tidak	Diatas KKM
S76	<7	Ya	Ya	Diatas KKM
S77	<7	Ya	Ya	Diatas KKM
S78	<7	Tidak	Ya	Diatas KKM
S79	<7	Tidak	Ya	Diatas KKM
S80	<7	Ya	Ya	Diatas KKM
S81	<7	Ya	Ya	Diatas KKM
S82	<7	Ya	Ya	Diatas KKM
S83	<7	Ya	Ya	Diatas KKM
S84	<7	Ya	Ya	Diatas KKM
S85	<7	Ya	Ya	Diatas KKM
S86	$\geq 7$	Tidak	Ya	KKM
S87	<7	Ya	Ya	Diatas KKM
S88	<7	Ya	Ya	Diatas KKM
S89	<7	Ya	Tidak	Diatas KKM
S90	<7	Ya	Ya	Diatas KKM
S91	<7	Ya	Ya	Diatas KKM
S92	<7	Ya	Ya	Diatas KKM

Tabel data training ini terdiri dari 92 data siswa (S01–S92) yang digunakan untuk menganalisis hubungan antara absensi (X1), kelas tambahan (X2), dan penggunaan media digital (X3) terhadap hasil nilai siswa (Y). Variabel absensi dibedakan menjadi kurang dari 7 (<7) dan lebih dari atau sama dengan 7 ( $\geq 7$ ), kelas tambahan dan media digital masing-masing memiliki kategori Ya dan Tidak, sedangkan variabel nilai diklasifikasikan menjadi KKM dan Di atas KKM. Data ini menunjukkan bahwa sebagian besar siswa berada pada kategori nilai KKM, sementara siswa dengan nilai Di atas KKM cenderung mengikuti kelas tambahan dan memanfaatkan media digital dalam proses pembelajaran. Data training ini digunakan sebagai dasar pembentukan model klasifikasi untuk menemukan pola dan aturan yang memengaruhi pencapaian nilai siswa.

**Tabel 4.3 Data Testing**

Nama	Absensi (X1)	Kelas Tambahan (X2)	Media Digital (X3)	Nilai (Y)
S93	$\geq 7$	Ya	Ya	Diatas KKM
S94	$< 7$	Ya	Ya	Diatas KKM
S95	$< 7$	Ya	Ya	Diatas KKM
S96	$< 7$	Ya	Ya	Diatas KKM
S97	$< 7$	Ya	Ya	Diatas KKM
S98	$< 7$	Ya	Ya	Diatas KKM
S99	$< 7$	Ya	Ya	Diatas KKM
S100	$< 7$	Ya	Ya	Diatas KKM
S101	$< 7$	Ya	Ya	Diatas KKM
S102	$< 7$	Tidak	Tidak	KKM
S103	$< 7$	Ya	Ya	Diatas KKM
S104	$< 7$	Ya	Tidak	Diatas KKM
S105	$< 7$	Tidak	Ya	Diatas KKM
S106	$\geq 7$	Tidak	Tidak	KKM
S107	$< 7$	Ya	Tidak	Diatas KKM
S108	$< 7$	Tidak	Ya	Diatas KKM
S109	$< 7$	Ya	Tidak	Diatas KKM
S110	$< 7$	Ya	Ya	Diatas KKM
S111	$< 7$	Tidak	Ya	Diatas KKM
S112	$< 7$	Ya	Tidak	Diatas KKM
S113	$< 7$	Ya	Ya	Diatas KKM
S114	$< 7$	Ya	Ya	Diatas KKM
S115	$< 7$	Tidak	Ya	Diatas KKM

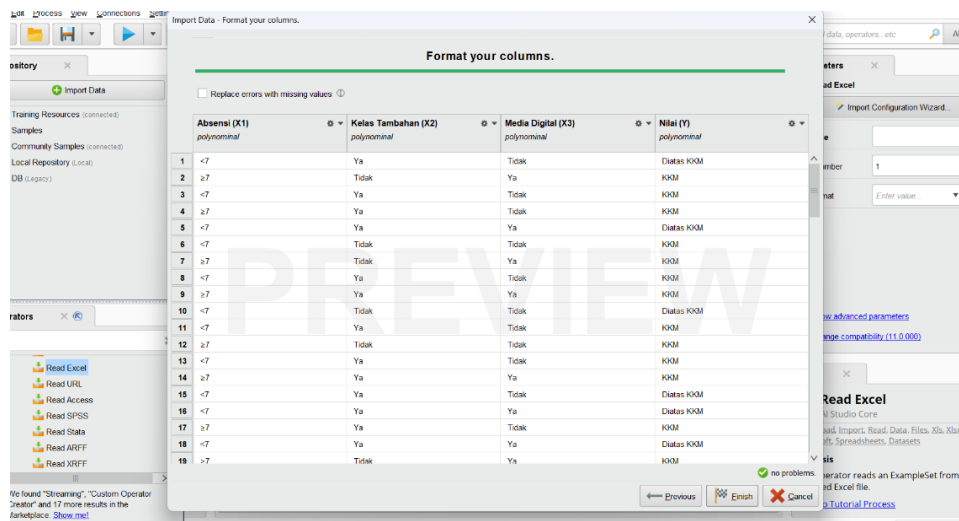
Tabel ini merupakan data testing yang terdiri dari 23 siswa (S93–S115) dan digunakan untuk menguji model klasifikasi yang telah dibangun sebelumnya. Variabel yang digunakan meliputi absensi (X1) dengan kategori  $< 7$  dan  $\geq 7$ , kelas tambahan (X2) dengan kategori Ya dan Tidak, serta media digital (X3) dengan kategori Ya dan Tidak, sedangkan variabel hasil (Y) adalah nilai siswa yang diklasifikasikan menjadi KKM dan Di atas KKM. Data testing ini berfungsi untuk mengevaluasi kemampuan model dalam memprediksi hasil nilai siswa berdasarkan pola yang diperoleh dari data training.

## 4.4 Pengolahan Data Menggunakan Rapidminer

Pada tahap ini, data yang telah melalui proses preprocessing selanjutnya diolah menggunakan aplikasi RapidMiner. RapidMiner digunakan untuk membangun model klasifikasi berdasarkan algoritma Decision Tree C4.5 guna mengetahui pola hubungan antara absensi, kelas tambahan, penggunaan media digital, dan nilai siswa.

### 4.4.1 Input Data

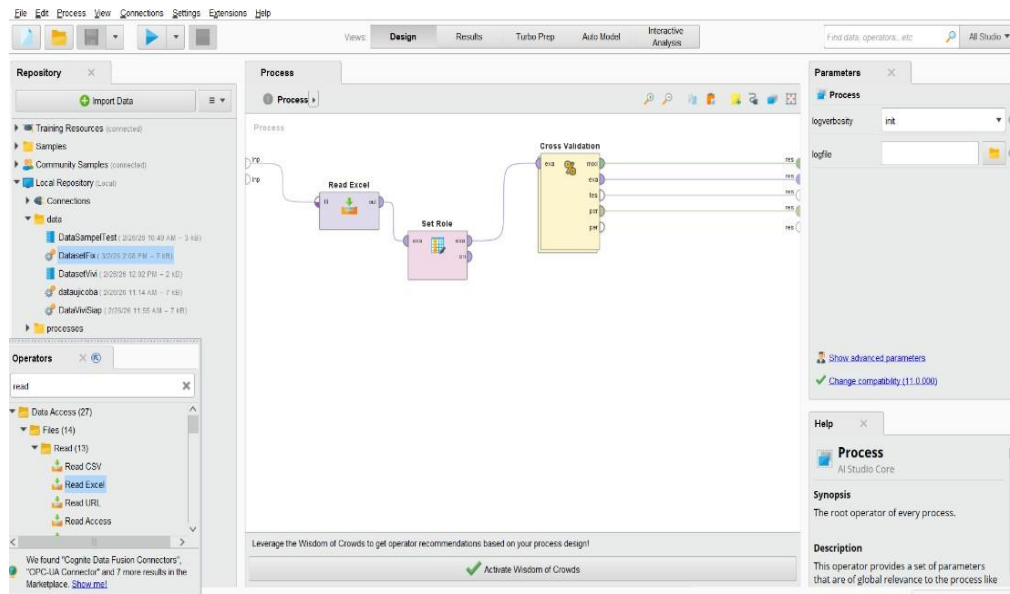
Pada tahap awal pengolahan data menggunakan RapidMiner, langkah pertama yang dilakukan adalah mengimpor data ke dalam sistem. Data yang telah disiapkan sebelumnya dalam format Excel atau CSV dipilih melalui menu Import Data, kemudian sistem akan menampilkan pratinjau data. Pada tahap ini dilakukan pengecekan setiap kolom agar sesuai, seperti Nama, Absensi, Kelas Tambahan, Penggunaan Media Digital, dan Nilai. Selanjutnya ditentukan tipe data masing-masing atribut, di mana Nilai dijadikan sebagai label atau target yang akan diprediksi, sedangkan atribut lainnya sebagai variabel pendukung.



Gambar 4.1 Input Data

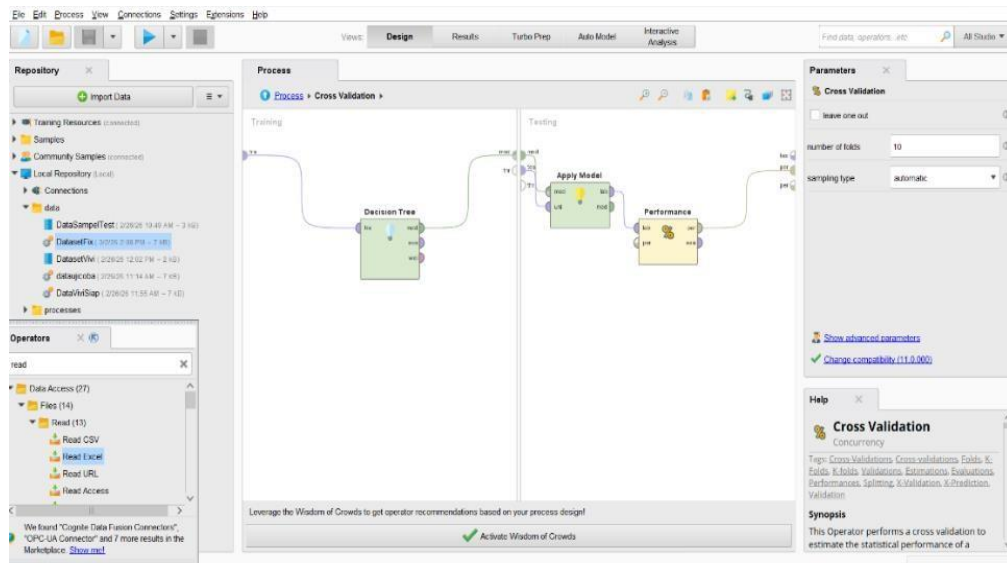
#### 4.4.2 Perancangan dan Penambahan Operator pada RapidMiner

Pada tahap ini dilakukan perancangan alur proses pengolahan data pada aplikasi RapidMiner Studio dengan menambahkan beberapa operator ke halaman *Process*. Perancangan ini bertujuan untuk membentuk model klasifikasi secara sistematis sesuai dengan kebutuhan peneliti.



**Gambar 4.2 Penambahan Operator *Read Excel*, *Set Role* dan *Cross Validation***

Operator pertama yang ditambahkan adalah *Read Excel*, yang berfungsi untuk membaca dan memasukkan data penelitian ke dalam proses analisis. Setelah data berhasil ditampilkan, langkah berikutnya adalah menambahkan operator *Set Role* untuk menentukan atribut label atau target, yaitu variabel Nilai sebagai variabel yang akan diprediksi. Selanjutnya, digunakan operator *Cross Validation* untuk membagi data menjadi data training dan data testing secara otomatis. Metode ini dipilih agar proses evaluasi model menjadi lebih akurat karena dilakukan pengujian berulang pada beberapa partisi data.

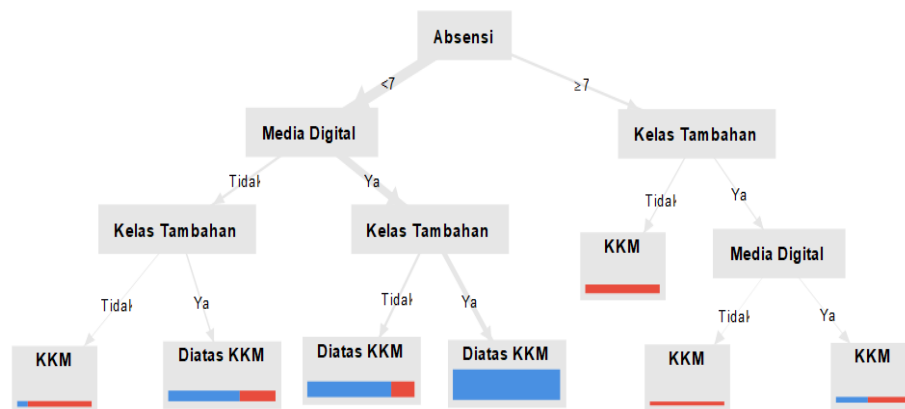


**Gambar 4.3 Penambahan Operator Decision Tree, Apply Model dan Performance**

Di dalam bagian training pada *Cross Validation*, ditambahkan operator *Decision Tree* untuk membentuk model klasifikasi berdasarkan perhitungan entropy dan gain ratio. Pada bagian testing, ditambahkan operator *Apply Model* untuk menerapkan model yang telah dibentuk ke data uji, kemudian dilanjutkan dengan operator *Performance (Classification)* untuk mengukur kinerja model. Hasil evaluasi yang diperoleh berupa nilai akurasi, precision, recall, dan confusion matrix. Seluruh rangkaian operator tersebut kemudian dijalankan (*Run*) untuk menghasilkan model pohon keputusan dan nilai performa klasifikasi yang akan dianalisis pada tahap selanjutnya.

#### 4.5 Pohon Keputusan

Pohon keputusan merupakan hasil model klasifikasi yang dibentuk menggunakan algoritma *Decision Tree* pada RapidMiner. Model ini menampilkan pola hubungan antar variabel dalam bentuk struktur bercabang yang terdiri dari akar (root), cabang (node), dan daun (leaf). Setiap cabang menunjukkan kondisi berdasarkan atribut tertentu, sedangkan daun menunjukkan hasil klasifikasi nilai siswa.



**Gambar 4.4 Pohon Keputusan**

Pohon keputusan pada gambar ini menjelaskan bagaimana hasil belajar siswa ditentukan berdasarkan beberapa faktor, yaitu absensi, penggunaan media digital, dan keikutsertaan dalam kelas tambahan.

Proses dimulai dari absensi sebagai faktor utama. Jika absensi siswa kurang dari 7 (<7), maka penentuan dilanjutkan ke media digital. Jika siswa tidak menggunakan media digital, maka dilihat lagi apakah ia mengikuti kelas tambahan atau tidak. Jika tidak ikut, hasilnya cenderung KKM, sedangkan jika ikut, hasilnya bisa di atas KKM. Jika siswa menggunakan media digital, maka meskipun tidak ikut kelas tambahan, hasilnya tetap bisa di atas KKM, dan jika

ikut kelas tambahan, hasilnya semakin baik (di atas KKM).

Sementara itu, jika absensi siswa  $\geq 7$ , maka langsung dilihat dari kelas tambahan. Jika siswa tidak mengikuti kelas tambahan, maka hasilnya KKM. Namun jika siswa mengikuti kelas tambahan, maka dilihat lagi penggunaan media digital. Baik menggunakan atau tidak, hasilnya tetap KKM.

Secara sederhana, pohon ini menunjukkan bahwa absensi adalah faktor paling penting, kemudian diikuti oleh media digital dan kelas tambahan. Siswa dengan absensi rendah masih bisa mendapatkan nilai baik jika memanfaatkan media digital dan mengikuti kelas tambahan, sedangkan siswa dengan absensi tinggi belum tentu mendapatkan hasil tinggi jika tidak didukung faktor lainnya.

#### **4.6 Analisis Faktor-Faktor yang Mempengaruhi Hasil Belajar**

Berdasarkan pohon keputusan yang telah dibuat, terdapat tiga faktor utama yang memengaruhi hasil belajar siswa, yaitu absensi, penggunaan media digital, dan keikutsertaan dalam kelas tambahan. Dari ketiga faktor tersebut, absensi menjadi faktor paling dominan karena berada pada akar pohon keputusan, yang berarti menjadi penentu awal dalam proses klasifikasi hasil belajar.

Siswa dengan absensi kurang dari 7 cenderung memiliki hasil belajar yang lebih rendah. Namun, kondisi ini masih dapat ditingkatkan jika siswa menggunakan media digital dan mengikuti kelas tambahan. Penggunaan media digital terbukti membantu siswa memahami materi dengan lebih baik, sehingga dapat meningkatkan peluang untuk mencapai nilai di atas KKM, terutama jika didukung dengan keikutsertaan dalam kelas tambahan.

Di sisi lain, siswa dengan absensi  $\geq 7$  tidak selalu mendapatkan hasil belajar yang tinggi. Jika siswa tidak mengikuti kelas tambahan, maka hasilnya cenderung hanya mencapai KKM. Bahkan ketika siswa mengikuti kelas tambahan, tanpa dukungan

penggunaan media digital yang optimal, hasil belajar tetap tidak menunjukkan peningkatan yang signifikan.

Secara keseluruhan, dapat disimpulkan bahwa hasil belajar siswa tidak hanya dipengaruhi oleh tingkat kehadiran, tetapi juga sangat dipengaruhi oleh strategi belajar yang digunakan, seperti pemanfaatan media digital dan keaktifan dalam mengikuti kelas tambahan. Kombinasi dari ketiga faktor ini menjadi kunci dalam meningkatkan hasil belajar siswa secara optimal.

#### **4.7 Pemanfaatan Hasil Analisis Untuk Strategi Pembelajaran**

Berdasarkan hasil analisis menggunakan algoritma Decision Tree C4.5, ditemukan bahwa faktor seperti absensi, keikutsertaan dalam kelas tambahan, dan penggunaan media digital memiliki pengaruh terhadap hasil belajar siswa. Informasi ini dapat dimanfaatkan untuk merancang strategi pembelajaran yang lebih efektif dan tepat sasaran.

Salah satu strategi yang dapat diterapkan adalah meningkatkan kehadiran siswa di kelas. Siswa dengan tingkat absensi yang baik cenderung memiliki pemahaman materi yang lebih optimal. Oleh karena itu, guru dapat memberikan motivasi, pengawasan, atau aturan yang mendorong siswa untuk lebih disiplin dalam mengikuti pembelajaran.

Selain itu, kelas tambahan dapat dimanfaatkan sebagai sarana pendukung bagi siswa yang membutuhkan penguatan materi. Program ini dapat difokuskan pada siswa yang memiliki nilai di bawah KKM agar mereka mendapatkan bimbingan lebih intensif.

Penggunaan media digital juga dapat menjadi strategi yang efektif dalam meningkatkan minat belajar siswa. Guru dapat memanfaatkan video pembelajaran, aplikasi edukasi, atau platform online untuk membuat proses

belajar menjadi lebih menarik dan mudah dipahami.

Dengan memanfaatkan hasil analisis ini, diharapkan guru dapat menentukan metode pembelajaran yang sesuai dengan kebutuhan siswa, sehingga hasil belajar dapat meningkat secara maksimal.

#### 4.8 Hasil Evaluasi

untuk mengetahui tingkat keakuratan model dalam mengklasifikasikan data siswa. Berdasarkan hasil pengujian, diperoleh nilai akurasi yang menunjukkan persentase ketepatan prediksi sesuai dengan kategori Nilai. Selain itu, ditampilkan juga nilai precision dan recall untuk melihat kemampuan model dalam mengidentifikasi setiap kelas secara lebih rinci. Confusion matrix memberikan gambaran jumlah data yang terklasifikasi benar maupun salah. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa model Decision Tree mampu memberikan performa yang baik dalam menganalisis pola belajar siswa.

accuracy: 85.23% +/- 10.01% (micro average: 85.22%)

	true Diatas KKM	true KKM	class precision
pred. Diatas KKM	75	13	85.23%
pred. KKM	4	23	85.19%
class recall	94.94%	63.89%	

**Gambar 4.5 Hasil Evaluasi**

Berdasarkan hasil evaluasi model yang ditampilkan pada tabel, diperoleh nilai akurasi sebesar 85,23% dengan toleransi kesalahan  $\pm 10,01\%$ . Hal ini menunjukkan bahwa model yang digunakan sudah mampu melakukan klasifikasi dengan cukup baik, karena sebagian besar data berhasil diprediksi dengan benar.

Jika dilihat lebih rinci, pada kelas Diatas KKM, terdapat 75 data yang berhasil diprediksi dengan benar. Namun, masih ada 4 data yang seharusnya termasuk Diatas KKM tetapi salah diprediksi sebagai KKM. Nilai recall sebesar 94,94% pada kelas ini menunjukkan bahwa model sangat baik dalam mengenali siswa yang memiliki nilai tinggi atau sudah melampaui KKM.

Sementara itu, pada kelas KKM, terdapat 23 data yang berhasil diprediksi dengan benar, tetapi terdapat 13 data yang sebenarnya termasuk KKM namun salah diprediksi sebagai Diatas KKM. Nilai recall sebesar 63,89% menunjukkan bahwa kemampuan model dalam mengenali siswa yang berada pada kategori KKM masih tergolong cukup, namun belum optimal.

Dari sisi precision, kedua kelas memiliki nilai yang hampir sama, yaitu sekitar 85%, yang berarti prediksi yang dihasilkan model cukup konsisten dan tidak terlalu banyak menghasilkan kesalahan.

Secara keseluruhan, model ini sudah cukup efektif dalam mengklasifikasikan hasil belajar siswa, terutama dalam mengenali siswa yang berada pada kategori Diatas KKM. Namun, model masih perlu ditingkatkan agar lebih akurat dalam mengidentifikasi siswa pada kategori KKM, sehingga hasil prediksi menjadi lebih seimbang dan dapat digunakan sebagai dasar pengambilan keputusan yang lebih tepat.