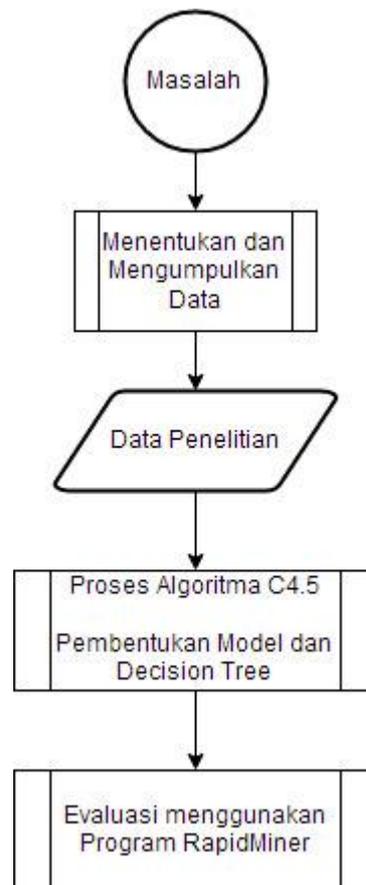


BAB III

METODOLOGI PENELITIAN

3.1. Desain Penelitian

Dalam sebuah penelitian, dibutuhkan sebuah desain penelitian agar penelitian dapat dilakukan dengan baik. Gambar 3.1 di bawah akan menjelaskan desain penelitian yang akan digunakan oleh peneliti.



Gambar 3.1 Desain Penelitian

Di bawah ini merupakan penjelasan dari desain penelitian:

1. Masalah, merupakan alasan utama penelitian dilakukan. Masalah harus melalui tahap Identifikasi Masalah untuk menentukan pokok masalah yang dibahas, lalu dilanjutkan ke tahap Perumusan Masalah.
2. Menentukan dan mengumpulkan data.

Untuk mendapatkan solusi dalam menyelesaikan sebuah masalah, diperlukan data-data yang berhubungan dengan masalah dan metode yang dibahas. Proses menentukan dan mengumpulkan data akan menghasilkan data penelitian yang akan dilanjutkan dalam proses selanjutnya.
3. Proses Algoritma C4.5

Dengan menggunakan algoritma C4.5, maka peneliti akan memproses data penelitian yang didapatkan dan menghasilkan sebuah pengetahuan yang baru yang diharapkan dapat menjadi solusi dalam memecahkan masalah.
4. Evaluasi

Evaluasi yang dilakukan dengan menggunakan program *RapidMiner* bertujuan untuk melakukan pengecekan ulang dan membandingkan nilai akurasi antara algoritma yang digunakan.

3.2. Teknik Pengumpulan Data

Pada penelitian ini, penulis telah mengumpulkan data dan informasi yang dapat memperkuat ketepatan penelitian yang telah dilakukan. Teknik pengumpulan data yang dilakukan oleh peneliti sebagai berikut :

1. Kajian Pustaka

Pada penelitian ini, kajian pustaka yang dilakukan adalah mengumpulkan data – data dari berbagai media seperti buku, jurnal, referensi skripsi dan juga artikel – artikel yang langsung maupun tidak langsung berhubungan dengan topik pembahasan yang dibahas dalam penelitian ini. Pengumpulan data yang dilakukan bertujuan untuk memahami algoritma yang digunakan serta memahami topik – topik yang dibahas dalam penelitian.

2. Observasi

Pada penelitian ini, penulis juga akan melakukan observasi langsung ke lapangan yang bertujuan untuk mengumpulkan data – data mengenai penjurusan yang telah dilakukan siswa-siswi sebelum dan setelah penjurusan. Pada observasi yang dilakukan, peneliti akan melakukan wawancara dengan guru yang bersangkutan dalam proses penjurusan siswa-siswi serta data mengenai penjurusan yang dilakukan. Data yang dikumpulkan akan dijadikan data yang akan diteliti.

3.3. Operasional Variabel

Terdapat 2 jenis variabel yang akan digunakan dalam penelitian ini, yaitu variabel bebas dan variabel terikat. Dari data yang telah terkumpul, data yang akan ditetapkan sebagai variabel terikat adalah data Jurusan yang terdiri dari 2 atribut isi yaitu IPA dan IPS. Atribut-atribut lainnya yang akan digunakan sebagai variabel bebas terdapat beberapa macam. Menurut Yayasan Konsultasi & Bimbingan “YACOBI”, keterangan singkat mengenai aspek-aspek yang akan menjadi variabel dalam penelitian ini adalah sebagai berikut :

1. Nilai Pemahaman Ruang (PR)

Kemampuan dalam meragakan suatu pola dalam bentuk ruang. Kemampuan ini dapat mendasari kesuksesan seorang dalam bidang arsitek, *home decoration* dan sejenisnya. Memiliki nilai yang baik pada kemampuan ini berdampak memiliki kemampuan yang tinggi untuk berimajinasi suatu pola dalam bentuk ruangan sedangkan memiliki nilai yang buruk pada kemampuan ini berdampak kurang memiliki daya imajinasi terhadap suatu pola berbentuk ruang.

2. Nilai Daya Abstraksi (PO)

Kemampuan dalam pemecahan suatu permasalahan dan yang abstrak. Dengan astraksi yang tinggi seseorang akan mempunyai daya tanggap yang peka sehingga akan mampu memberikan respon terhadap segala

jenis Stimuli dengan cepat dan tepat. Kemampuan ini sangat berguna bagi pemikir, *designer*, *programmer* dan sejenisnya. Memiliki nilai yang baik pada kemampuan ini berdampak memiliki kemampuan potensial yang tinggi dalam pemecahan suatu permasalahan dan bersifat abstrak, serta mampu memberikan respon dengan cepat dan akurat sedangkan memiliki nilai yang buruk pada kemampuan ini berdampak memiliki kemampuan potensial jauh di bawah rata-rata serta susah berpikir secara abstrak dan respon yang bersifat terlalu sederhana.

3. Nilai Kemampuan Dasar Ilmu Pasti (PB)

Kemampuan penalaran dalam berhitung dengan angka serta dasar ilmu pasti. Dengan kemampuan ini, pasti akan mempermudah dalam melanjutkan pendidikan yang lebih tinggi terutama dalam pengetahuan *exact* seperti MIPA. Memiliki nilai yang baik pada kemampuan ini berdampak memiliki kemampuan yang baik terhadap pengetahuan yang berhubungan dengan pengolahan angka-angka baik dalam penyusunan konsep ataupun menganalisa persamaan rumus sedangkan memiliki nilai yang buruk pada kemampuan ini berdampak kurang mampu menghadapi tugas yang berhubungan dengan pengolahan angka.

4. Nilai Potensi Verbal / Kapasitas Belajar (PV)

Kemampuan ini memiliki hubungan yang erat dengan prestasi belajar yang telah dicapai di sekolah. Dengan kemampuan verbal yang tinggi

(mencakup kemampuan bahasa, dasar ilmu pasti dan berpikir praktis) akan mempermudah dalam menerima segala pelajaran yang diberikan. Memiliki nilai yang baik pada kemampuan ini berdampak memiliki kemampuan berbahasa dan penalaran yang baik dan berpikiran praktis yang logis sedangkan memiliki nilai yang buruk pada kemampuan ini berdampak lemah dalam berbahasa dan penalaran terhadap angka serta berpikiran yang terlalu sederhana dan tidak praktis.

3.4. Metode Analisis dan Rancangan Sistem

Rancangan dalam penelitian ini menggunakan rancangan berikut :



Gambar 3.2 Rancangan Sistem

Dari gambar 3.2, dapat diperhatikan bahwa dalam rancangan penelitian ini proses diawali dengan melakukan seleksi pada data nilai IQ siswa yang telah diambil dari SMAK Yos Sudarso. Setelah dilakukan seleksi data, maka akan didapatkan sampel data yang akan diolah. Selanjutnya, sampel data akan dibersihkan atau *cleaning* yang bertujuan untuk menghilangkan *noise* atau atribut yang tidak perlu agar memudahkan perhitungan nantinya. Setelah dibersihkan, maka data akan dilakukan transformasi sehingga data dapat berbentuk format yang dapat diolah oleh algoritma. Dari hasil transformasi tersebut maka data akan diolah dengan menggunakan *data mining*.

Metode *data mining* yang akan dilakukan dalam penelitian ini adalah teknik prediksi dengan memanfaatkan algoritma C4.5. Proses algoritma meliputi pencarian nilai entropi, *gain*, *split info*, dan *gain ratio* yang merupakan proses kerja algoritma C4.5. Dari nilai-nilai tersebut, algoritma akan menghasilkan sebuah pohon keputusan dan dari pohon keputusan tersebut akan dibuat *rules* agar mudah dipahami oleh pembaca.

3.5. Lokasi dan Jadwal Penelitian

Pada subbab ini, peneliti akan mencantumkan lokasi dan waktu dalam pelaksanaan penelitian, yaitu sebagai berikut :

3.4.1. Lokasi Penelitian

Peneliti akan melakukan penelitian di lokasi SMAK Yos Sudarso Batam yang beralamat di Komplek Sekolah Yos Sudarso, Jl. Dang Merdu No. 2, Teluk Tering, Batam Kota, Kota Batam, Kepulauan Riau, Indonesia.

3.4.2. Waktu Penelitian

Pada tabel di bawah, dapat dilihat pembagian waktu dari proses persiapan hingga akhirnya penyusunan laporan.

Tabel 3.1 Tabel Waktu Penelitian

No	Uraian	Bulan																					
		2017												2018									
		Sept				Okt				Nov				Des				Jan				Feb	
1	Persiapan Penelitian																						
2	Perencanaan																						
3	Pelaksanaan Penelitian																						
4	Penyusunan Laporan																						
5	Bimbingan																						

Disebabkan oleh kebatasan lebar halaman, tabel waktu penelitian hanya mencakup hingga bulan Februari 2018. Namun, pada faktanya penyusunan laporan dan bimbingan memakan waktu hingga bulan Juli 2018.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Analisa Data

Analisa data yang dilakukan dalam penelitian ini dapat diuraikan dalam penjelasan secara mendetail sebagai berikut :

4.1.1. Seleksi Data

Data-data yang ada pada *dataset* merupakan hasil pengumpulan data dari SMAK Yos Sudarso yang berjumlah 236 data. Data tersebut merupakan populasi data yang akan diambil beberapa data yang ada sebagai sampel untuk diteliti. Jika jumlah populasi sebuah *dataset* yang akan diteliti kurang dari 100 data, maka keseluruhan populasi dijadikan sampel. Apabila populasi berjumlah lebih dari 100 data, maka sampel dapat diperoleh antara 10-15% atau 20-25%. Rumus perhitungan jumlah sampel yaitu : (Arikunto, 2010)

$$\text{Jumlah sampel} = 15\% \times \text{Jumlah populasi}$$

Rumus 4.1. Perhitungan Sampel

Untuk perhitungan total data populasi sebanyak 236 data, sehingga 15% dari total populasi yang akan dijadikan sampel berjumlah 36 data.

4.1.2. *Cleaning Data*

Berdasarkan sampel yang telah diseleksi, jumlah data yang akan diteliti adalah 36 data. Dari 36 data tersebut, akan dilakukan *cleaning* yang bertujuan untuk menghilangkan variabel yang tidak berguna. Penghapusan variabel meliputi :

1. Variabel nomor akan dihapus karena atribut ID yang dipakai untuk membedakan adalah variabel nama.
2. Variabel tanggal lahir akan dihapus karena tidak variabel ini tidak dapat diolah.

4.1.3. *Tranformasi Data*

Agar data dapat diolah oleh algoritma C4.5, data kuantitatif yang ada harus diolah menjadi data kualitatif. Oleh sebab itu, beberapa variabel harus melalui proses transformasi data. Berikut transformasi data yang dilakukan :

- a. Atribut Pemahaman Ruang (PR)

Data atribut PR berada pada jarak minimum 97 dan jarak maksimum 140.

Maka interval dapat dicari dengan menggunakan rumus berikut :

$$\text{Interval} = \frac{\text{nilai terbesar} - \text{nilai terkecil}}{\text{banyak kelas}}$$

Rumus 4.2. Perhitungan Interval

Atribut akan dibagi ke dalam 3 kelas, yaitu : KURANG, CUKUP, LEBIH sehingga perhitungan menjadi sebagai berikut :

$$\text{Interval} = \frac{140 - 97}{3} = 14.33 \approx 15$$

Interval untuk atribut PR adalah 14, sehingga dapat dibentuk dalam tabel sebagai berikut :

Tabel 4.1 Pembagian Kelas Nilai PR

Nilai	Kelas	Singkatan
97 s/d 111	KURANG	K
112 s/d 125	CUKUP	C
126 s/d 140	LEBIH	L

b. Atribut Daya Abstraksi (PO)

Data atribut PO berada pada jarak minimum 94 dan jarak maksimum 147.

Atribut akan dibagi ke dalam 3 kelas, yaitu : KURANG, CUKUP, LEBIH

sehingga perhitungan menjadi sebagai berikut :

$$Interval = \frac{147 - 94}{3} = 17$$

Interval untuk atribut PO adalah 17, sehingga dapat dibentuk dalam tabel sebagai berikut :

Tabel 4.2 Pembagian Kelas Nilai PO

Nilai	Kelas	Singkatan
94 s/d 111	KURANG	K
112 s/d 128	CUKUP	C
129 s/d 147	LEBIH	L

c. Atribut Kemampuan Dasar Ilmu Pasti (PB)

Data atribut PB berada pada jarak minimum 90 dan jarak maksimum 148.

Atribut akan dibagi ke dalam 3 kelas, yaitu : KURANG, CUKUP, LEBIH sehingga perhitungan menjadi sebagai berikut :

$$Interval = \frac{148 - 90}{3} = 19.33 \approx 20$$

Interval untuk atribut PB adalah 20, sehingga dapat dibentuk dalam tabel sebagai berikut :

Tabel 4.3 Pembagian Kelas Nilai PB

Nilai	Kelas	Singkatan
90 s/d 109	KURANG	K
110 s/d 128	CUKUP	C
129 s/d 148	LEBIH	L

d. Atribut Potensi Verbal / Kapasitas Belajar (PV)

Data atribut PB berada pada jarak minimum 92 dan jarak maksimum 148.

Atribut akan dibagi ke dalam 3 kelas, yaitu : KURANG, CUKUP, LEBIH sehingga perhitungan menjadi sebagai berikut :

$$Interval = \frac{148 - 92}{3} = 18.66 \approx 19$$

Interval untuk atribut PV adalah 19, sehingga dapat dibentuk dalam tabel sebagai berikut :

Tabel 4.4 Pembagian Kelas Nilai PV

Nilai	Kelas	Singkatan
92 s/d 110	KURANG	K
111 s/d 129	CUKUP	C
130 s/d 148	LEBIH	L

4.1.4. Perhitungan Algoritma

Setelah data melalui seluruh tahap di atas, maka data akan diproses dengan menggunakan algoritma yang di penelitian ini menggunakan algoritma C4.5. Perhitungan algoritma C4.5 dalam memprediksi penjurusan siswa diproses dengan mencari nilai entropi, *gain*, *split info*, dan *gain ratio* masing-masing variabel. Pencarian dilakukan berulang-ulang dalam sebuah iterasi hingga menghasilkan simpul yang tidak bisa diturunkan lagi. Berikut merupakan tabel data sampel penelitian yang akan diolah dengan menggunakan algoritma C4.5 :

Tabel 4.5 Sampel Penelitian

Nama	PR	PO	PB	PV	Jurusan
Adrian Hartanto	C	C	C	C	IPS
Andes Willi Wijaya	L	L	C	C	IPA
Andy Edyanto	L	C	L	C	IPA
Arron Constantin	K	C	C	C	IPS
Catrine Mega Putri	C	K	K	K	IPS
Cinthya Govianda	C	C	C	C	IPA
Daniel Yeoda	L	L	L	L	IPA
Deviyanti	K	C	K	K	IPS
Dicky Kurniawan	C	L	C	C	IPA
Ellis	K	K	C	C	IPS
Ervin	C	C	C	K	IPS

Felicia Handali	C	C	C	C	IPS
Franky August T.	C	C	C	C	IPS
Handoyo	C	C	C	C	IPS
Hendro	C	C	C	C	IPA
Ivan Giovanni	L	C	C	C	IPA
Jennifer Handali	L	L	L	C	IPA
Jonathan Wijaya	C	L	L	L	IPA
Kevin	L	C	C	C	IPA
Lidya Vinesa	K	C	C	K	IPS
Maria Naning H.	K	K	K	K	IPS
Michael Juan	L	L	L	C	IPA
Nicholas Josua L.	C	L	L	L	IPA
Novita Sari	C	C	C	C	IPS
Ricky	L	L	L	L	IPA
Santi	C	L	C	C	IPA
Shervi	C	C	C	C	IPS
Sofyana	C	C	K	K	IPS
Suryadi	C	C	C	C	IPS
Vania Lystia	C	C	C	C	IPS
Vie Vie Weng	L	L	L	L	IPA
Viola Chyntialie	C	C	C	C	IPA
Vivian Williana	C	C	C	C	IPS
Willy Tantoso W.	L	C	L	C	IPA
Yolanda Tjong	C	C	C	C	IPS
Yulandha Felicia S.	L	C	C	C	IPS

Pada data sampel diatas, akan dicari nilai entropi total dengan menggunakan rumus 2.1. Data dengan atribut IPA berjumlah 17 data sedangkan atribut IPS berjumlah 19 data yang berarti total keseluruhan sampel data berjumlah 36 data.

$$\text{Entropy (total)} = \left(-\frac{17}{36} * \log_2 \frac{17}{36}\right) + \left(-\frac{19}{36} * \log_2 \frac{19}{36}\right) = 0,9978$$

Hasil dari perhitungan entropi total data sampel sebesar 0,9977. Setelah menghitung nilai entropi total data, maka akan dilanjutkan dengan pencarian nilai entropi masing-masing atribut sebagai berikut :

a. Atribut PR

Perhitungan entropi atribut PR dimulai dengan menghitung atribut yang ada pada variabel dengan berbeda atribut. Dapat dilihat dari tabel data sampel penelitian, pada atribut PR dengan nilai KURANG dan berjurusan IPS terdapat 5 data sedangkan yang berjurusan IPA tidak ada. Nilai CUKUP dan berjurusan IPS terdapat 13 data sedangkan yang berjurusan IPS terdapat 7 data. Nilai LEBIH dan berjurusan IPS terdapat 1 data sedangkan yang berjurusan IPA terdapat 10 data. Dari data-data tersebut maka berikut nilai entropi masing-masing kasus :

$$\text{Entropy (K)} = \left(-\frac{5}{5} * \log_2 \frac{5}{5}\right) + \left(-\frac{0}{5} * \log_2 \frac{0}{5}\right) = 0$$

$$\text{Entropy (C)} = \left(-\frac{13}{20} * \log_2 \frac{13}{20}\right) + \left(-\frac{7}{20} * \log_2 \frac{7}{20}\right) = 0,9341$$

$$\text{Entropy (L)} = \left(-\frac{1}{11} * \log_2 \frac{1}{11}\right) + \left(-\frac{10}{11} * \log_2 \frac{10}{11}\right) = 0,4395$$

b. Atribut PO

Perhitungan entropi atribut PO dimulai dengan menghitung atribut yang ada pada variabel dengan berbeda atribut. Dapat dilihat dari tabel data sampel penelitian, pada atribut PO dengan nilai KURANG dan berjurusan IPS terdapat 3 data sedangkan yang berjurusan IPA tidak ada. Nilai CUKUP dan berjurusan IPS terdapat 16 data sedangkan yang berjurusan IPS terdapat 7 data. Nilai LEBIH dan berjurusan IPS tidak ada sedangkan yang berjurusan IPA terdapat 10 data. Dari data-data tersebut maka berikut nilai entropi masing-masing kasus :

$$\text{Entropy (K)} = \left(-\frac{3}{3} * \log_2 \frac{3}{3}\right) + \left(-\frac{0}{3} * \log_2 \frac{0}{3}\right) = 0$$

$$\text{Entropy (C)} = \left(-\frac{16}{23} * \log_2 \frac{16}{23}\right) + \left(-\frac{7}{23} * \log_2 \frac{7}{23}\right) = 0,8865$$

$$\text{Entropy (L)} = \left(-\frac{0}{10} * \log_2 \frac{0}{10}\right) + \left(-\frac{10}{10} * \log_2 \frac{10}{10}\right) = 0$$

c. Atribut PB

Perhitungan entropi atribut PB dimulai dengan menghitung atribut yang ada pada variabel dengan berbeda atribut. Dapat dilihat dari tabel data sampel penelitian, pada atribut PB dengan nilai KURANG dan berjurusan IPS terdapat 4 data sedangkan yang berjurusan IPA tidak ada. Nilai CUKUP dan berjurusan IPS terdapat 15 data sedangkan yang berjurusan IPS terdapat 8 data. Nilai LEBIH dan berjurusan IPS tidak ada sedangkan yang berjurusan IPA terdapat 9 data. Dari data-data tersebut maka berikut nilai entropi masing-masing kasus :

$$\text{Entropy (K)} = \left(-\frac{4}{4} * \log_2 \frac{4}{4}\right) + \left(-\frac{0}{4} * \log_2 \frac{0}{4}\right) = 0$$

$$\text{Entropy (C)} = \left(-\frac{15}{23} * \log_2 \frac{15}{23}\right) + \left(-\frac{8}{23} * \log_2 \frac{8}{23}\right) = 0,9321$$

$$\text{Entropy (L)} = \left(-\frac{0}{9} * \log_2 \frac{0}{9}\right) + \left(-\frac{9}{9} * \log_2 \frac{9}{9}\right) = 0$$

d. Atribut PV

Perhitungan entropi atribut PV dimulai dengan menghitung atribut yang ada pada variabel dengan berbeda atribut. Dapat dilihat dari tabel data sampel penelitian, pada atribut PV dengan nilai KURANG dan berjurusan IPS terdapat 6 data sedangkan yang berjurusan IPA tidak ada. Nilai

CUKUP dan berjurusan IPS terdapat 13 data sedangkan yang berjurusan IPS terdapat 12 data. Nilai LEBIH dan berjurusan IPS tidak ada sedangkan yang berjurusan IPA terdapat 5 data. Dari data-data tersebut maka berikut nilai entropi masing-masing kasus :

$$\text{Entropy (K)} = \left(-\frac{6}{6} * \log_2 \frac{6}{6}\right) + \left(-\frac{0}{6} * \log_2 \frac{0}{6}\right) = 0$$

$$\text{Entropy (C)} = \left(-\frac{13}{25} * \log_2 \frac{13}{25}\right) + \left(-\frac{12}{25} * \log_2 \frac{12}{25}\right) = 0,9988$$

$$\text{Entropy (L)} = \left(-\frac{0}{5} * \log_2 \frac{0}{5}\right) + \left(-\frac{5}{5} * \log_2 \frac{5}{5}\right) = 0$$

Setelah mencari nilai entropi masing-masing atribut, maka akan dilanjutkan dengan perhitungan *gain* masing-masing atribut dengan menggunakan rumus 2.2. Berikut perhitungan pada masing-masing atribut :

a. *Gain* (Total,PR)

Perhitungan *gain* dari atribut PR dapat dihitung dengan menggunakan nilai entropi yang bernilai 0,9978 yang lalu dikurangkan dengan total data per kategori atributnya dibagi dengan total data atributnya dan dikalikan dengan nilai entropi per kategori. Perhitungannya dapat ditulis sebagai berikut :

$$\begin{aligned} \text{Gain (PR)} &= 0,9978 - \left(\left(\frac{5}{36} * 0 \right) + \left(\frac{20}{36} * 0,9341 \right) + \left(\frac{11}{36} * 0,4395 \right) \right) \\ &= 0,3346 \end{aligned}$$

b. *Gain* (Total,PO)

Perhitungan *gain* dari atribut PO dapat dihitung dengan menggunakan nilai entropi total yang bernilai 0,9978 yang lalu dikurangkan dengan total data per kategori atributnya dibagi dengan total data atributnya dan dikalikan dengan nilai entropi per kategori. Perhitungannya dapat ditulis sebagai berikut :

$$\begin{aligned} \text{Gain (PO)} &= 0,9978 - \left(\left(\frac{3}{36} * 0 \right) + \left(\frac{23}{36} * 0,8865 \right) + \left(\frac{10}{36} * 0 \right) \right) \\ &= 0,4314 \end{aligned}$$

c. *Gain* (Total, PB)

Perhitungan *gain* dari atribut PB dapat dihitung dengan menggunakan nilai entropi yang bernilai 0,9978 yang lalu dikurangkan dengan total data per kategori atributnya dibagi dengan total data atributnya dan dikalikan dengan nilai entropi per kategori. Perhitungannya dapat ditulis sebagai berikut :

$$\begin{aligned} \text{Gain (PB)} &= 0,9978 - \left(\left(\frac{4}{36} * 0 \right) + \left(\frac{23}{36} * 0,9321 \right) + \left(\frac{5}{36} * 0 \right) \right) \\ &= 0,4023 \end{aligned}$$

d. *Gain* (Total, PV)

Perhitungan *gain* dari atribut PV dapat dihitung dengan menggunakan nilai entropi yang bernilai 0,9978 yang lalu dikurangkan dengan total data per kategori atributnya dibagi dengan total data atributnya dan dikalikan dengan nilai entropi per kategori. Perhitungannya dapat ditulis sebagai berikut :

$$\begin{aligned} \text{Gain (PV)} &= 0,9978 - \left(\left(\frac{6}{36} * 0 \right) + \left(\frac{25}{36} * 0,9988 \right) + \left(\frac{5}{36} * 0 \right) \right) \\ &= 0,3041 \end{aligned}$$

Setelah melakukan perhitungan nilai *gain*, maka proses algoritma dilanjutkan dengan perhitungan nilai *SplitInfo*. Berikut perhitungan *SplitInfo* pada masing-masing atribut :

a. Atribut PR

Perhitungan *SplitInfo* menggunakan rumus yang hampir sama dengan rumus hanya perbedaan terletak pada jumlah kasus per kategori dibagikan dengan jumlah kasus keseluruhan. Jumlah kasus sesuai kategori KURANG, CUKUP dan LEBIH secara berurut adalah 5, 20 dan 11 sehingga perhitungan yang terbentuk sebagai berikut :

$$\begin{aligned} \text{Split Info (PR)} &= \left(-\frac{5}{36} * \log_2 \left(\frac{5}{36} \right) \right) + \left(-\frac{20}{36} * \log_2 \left(\frac{20}{36} \right) \right) + \left(-\frac{11}{36} * \right. \\ &\quad \left. \log_2 \left(\frac{11}{36} \right) \right) = 1,3893 \end{aligned}$$

b. Atribut PO

Perhitungan *SplitInfo* menggunakan rumus yang hampir sama dengan rumus hanya perbedaan terletak pada jumlah kasus per kategori dibagikan dengan jumlah kasus keseluruhan. Jumlah kasus sesuai kategori KURANG, CUKUP dan LEBIH secara berurut adalah 3, 23 dan 10 sehingga perhitungan yang terbentuk sebagai berikut :

$$\begin{aligned} Split\ Info\ (PO) &= \left(-\frac{3}{36} * \log_2\left(\frac{3}{36}\right)\right) + \left(-\frac{23}{36} * \log_2\left(\frac{23}{36}\right)\right) + \left(-\frac{10}{36} * \right. \\ &\quad \left. \log_2\left(\frac{10}{36}\right)\right) = 1,225 \end{aligned}$$

c. Atribut PB

Perhitungan *SplitInfo* menggunakan rumus yang hampir sama dengan rumus hanya perbedaan terletak pada jumlah kasus per kategori dibagikan dengan jumlah kasus keseluruhan. Jumlah kasus sesuai kategori KURANG, CUKUP dan LEBIH secara berurut adalah 4, 23 dan 9 sehingga perhitungan yang terbentuk sebagai berikut :

$$\begin{aligned} Split\ Info\ (PB) &= \left(-\frac{4}{36} * \log_2\left(\frac{4}{36}\right)\right) + \left(-\frac{23}{36} * \log_2\left(\frac{23}{36}\right)\right) + \left(-\frac{9}{36} * \right. \\ &\quad \left. \log_2\left(\frac{9}{36}\right)\right) = 1,2652 \end{aligned}$$

d. Atribut PV

Perhitungan *SplitInfo* menggunakan rumus yang hampir sama dengan rumus hanya perbedaan terletak pada jumlah kasus per kategori dibagikan dengan jumlah kasus keseluruhan. Jumlah kasus sesuai kategori KURANG, CUKUP dan LEBIH secara berurut adalah 6, 25 dan 5 sehingga perhitungan yang terbentuk sebagai berikut :

$$Split\ Info\ (PV) = \left(-\frac{6}{36} * \log_2\left(\frac{6}{36}\right)\right) + \left(-\frac{25}{36} * \log_2\left(\frac{25}{36}\right)\right) + \left(-\frac{5}{36} * \log_2\left(\frac{5}{36}\right)\right) = 1,1917$$

Setelah mendapatkan nilai *SplitInfo* maka proses selanjutnya pada algoritma C4.5 adalah mencari nilai *gain ratio*. Perhitungan nilai *gain ratio* pada masing-masing atribut adalah sebagai berikut :

a. Atribut PR

Untuk menghitung nilai *gain ratio* pada atribut PR, dapat dihitung dengan menggunakan nilai *gain* atribut PR dibagikan dengan nilai *SplitInfo* atribut PR. Berikut hasil perhitungannya :

$$Gain\ Ratio\ (PR) = \frac{0,3446}{1,3893} = 0,248$$

b. Atribut PO

Untuk menghitung nilai *gain ratio* pada atribut PO, dapat dihitung dengan menggunakan nilai *gain* atribut PO dibagikan dengan nilai *SplitInfo* atribut PO. Berikut hasil perhitungannya :

$$Gain Ratio (PO) = \frac{0,4314}{1,225} = 0,3521$$

c. Atribut PB

Untuk menghitung nilai *gain ratio* pada atribut PB, dapat dihitung dengan menggunakan nilai *gain* atribut PB dibagikan dengan nilai *SplitInfo* atribut PB. Berikut hasil perhitungannya :

$$Gain Ratio (PB) = \frac{0,4023}{1,2652} = 0,3179$$

d. Atribut PV

Untuk menghitung nilai *gain ratio* pada atribut PV, dapat dihitung dengan menggunakan nilai *gain* atribut PV dibagikan dengan nilai *SplitInfo* atribut PV. Berikut hasil perhitungannya :

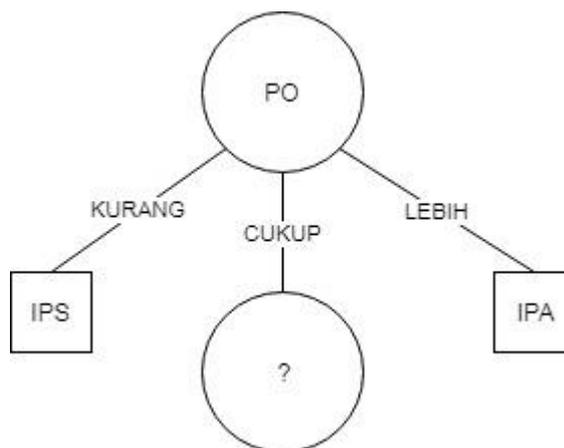
$$Gain Ratio (PV) = \frac{0,3041}{1,1917} = 0,2552$$

Hasil perhitungan nilai-nilai di atas kemudian dibentuk ke dalam tabel agar dapat dibaca secara jelas. Berikut penampilan tabel hasil perhitungan :

Tabel 4.6 Hasil Perhitungan *Node 1*

Atribut		Kasus	IPS	IPA	Entropi	Gain	Split Info	Gain Ratio
Total		36	19	17	0,9978			
PR						0,3446	1,3893	0,2480
	KURANG	5	5	0	0			
	CUKUP	20	13	7	0,9341			
	LEBIH	11	1	10	0,4395			
PO						0,4314	1,2250	0,3521
	KURANG	3	3	0	0			
	CUKUP	23	16	7	0,8865			
	LEBIH	10	0	10	0			
PB						0,4023	1,2652	0,3179
	KURANG	4	4	0	0			
	CUKUP	23	15	8	0,9321			
	LEBIH	9	0	9	0			
PV						0,3041	1,1917	0,2552
	KURANG	6	6	0	0			
	CUKUP	25	13	12	0,9988			
	LEBIH	5	0	5	0			

Berdasarkan hasil perhitungan yang telah disusun pada tabel di atas, dapat dilihat bahwa atribut dengan nilai *gain ratio* tertinggi adalah atribut PO yang bernilai 0,3521. Maka dari itu, atribut PO yang akan menjadi *node* akar, yang menghasilkan pohon keputusan sementara yang berbentuk seperti berikut :



Gambar 4.1 Pohon Keputusan *Node* 1

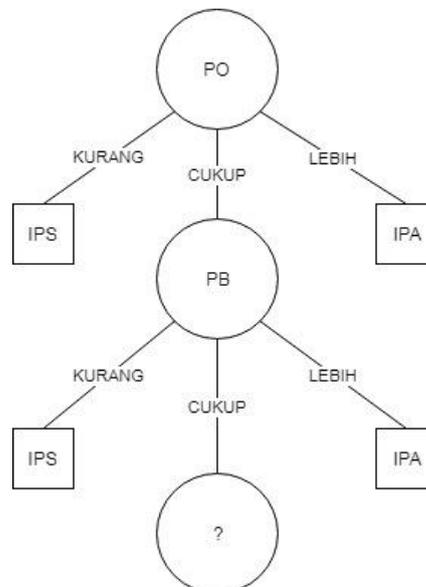
Arti dari pohon keputusan *node* 1 adalah akar dengan atribut PO memiliki 3 cabang dimana nilai pada cabang dengan kategori KURANG dan LEBIH berhenti dan menjadi *leaf* sedangkan pada kategori CUKUP masih perlu dilakukan perhitungan lebih lanjut lagi.

Untuk perhitungan yang akan dilakukan pada *node* 2, iterasi akan dilakukan dengan mengurangi jumlah kasus data. Kasus data yang dikurangi dengan cara penghapusan kasus yang memperoleh nilai *gain ratio* tertinggi, sehingga data yang dihapus berjumlah 13 kasus. Sisa 23 kasus akan dimasukkan ke dalam iterasi lagi untuk mencari variabel dengan nilai *gain ratio* tertinggi lagi. Berikut tabel hasil perhitungan *node* 2 :

Tabel 4.7 Hasil Perhitungan *Node 2*

Atribut		Kasus	IPS	IPA	Entropi	Gain	Split Info	Gain Ratio
Total		23	16	7	0,8865			
PR						0,2588	1,2641	0,2047
	KURANG	3	3	0	0			
	CUKUP	15	12	3	0,7219			
	LEBIH	5	1	4	0,7219			
PB						0,1997	0,8405	0,2376
	KURANG	2	2	0	0			
	CUKUP	19	14	5	0,8315			
	LEBIH	2	0	2	0			
PV						0,1022	0,6666	0,1533
	KURANG	4	4	0	0			
	CUKUP	19	12	7	0,9495			
	LEBIH	0	0	0	0			

Berdasarkan tabel data hasil perhitungan *node 2*, dapat diketahui atribut dengan nilai *gain ratio* tertinggi adalah atribut PB dengan nilai 0,2376. Maka dari itu, atribut PB akan menjadi *node 2* dalam pohon keputusan sementara. Gambaran pohon keputusan setelah *node 2* masuk menjadi seperti berikut :

**Gambar 4.2** Pohon Keputusan *Node 2*

Arti dari pohon keputusan *node 2* adalah akar dengan atribut PB memiliki 3 cabang dimana nilai pada cabang dengan kategori KURANG dan LEBIH berhenti dan menjadi *leaf* sedangkan pada kategori CUKUP masih perlu dilakukan perhitungan lebih lanjut lagi dengan keadaan yang hampir sama dengan *node 1*.

Untuk perhitungan yang akan dilakukan pada *node 3*, iterasi akan dilakukan dengan mengurangi jumlah kasus data. Kasus data yang dikurangi dengan cara penghapusan kasus yang memperoleh nilai *gain ratio* tertinggi pada perhitungan *node 2*, sehingga data yang dihapus berjumlah 4 kasus. Sisa 19 kasus akan dimasukkan ke dalam iterasi lagi untuk mencari variabel dengan nilai *gain ratio* tertinggi lagi. Berikut tabel hasil perhitungan *node 3* :

Tabel 4.8 Hasil Perhitungan *Node 3*

Atribut		Kasus	IPS	IPA	Entropi	Gain	Split Info	Gain Ratio
Total		19	14	5	0,8315			
PR						0,1341	1,0870	0,1234
	KURANG	2	2	0	0			
	CUKUP	14	11	3	0,7496			
	LEBIH	3	1	2	0,9183			
PV						0,0495	0,4855	0,1019
	KURANG	2	2	0	0			
	CUKUP	17	12	5	0,8740			
	LEBIH	0	0	0	0			

Berdasarkan tabel data hasil perhitungan *node 3*, dapat diketahui atribut dengan nilai *gain ratio* tertinggi adalah atribut PR dengan nilai 0,1234. Maka dari itu, atribut PR akan menjadi *node 3* dalam pohon keputusan sementara. Gambaran pohon keputusan setelah *node 3* masuk menjadi seperti berikut :