BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penulis ingin menganalisis algoritma terbaik diantara Naïve bayes dan C4.5 pada klasifikasi produk Zam-Zam Time berdasarkan tingkat kepuasan pelanggan. Dalam klasifikasi produk Zam-Zam Time dilakukan pengelompokan produk tergolong Laris atau Kurang Laris. Penelitian sebelumnya terkait klasifikasi ditemukan Algoritma Naïve bayes dan C4.5 memiliki nilai akurasi yang baik tetapi hasil akhir belum merujuk pada salah satu Algoritma. Tabel 2.1 memperlihatkan penelitian terkait penjualan dari kepuasan pelanggan dan algoritma klasifikasi.

Tabel 2. 1 Penelitian Terkait Penjualan dan Algoritma Klasifikasi

No.	Judul Penelitian	Nama dan Tahun Peneliti	Masalah	Metode dan Hasil
1	Faktor –Faktor Yang Mempengaruhi Kualitas Pelayanan Terhadap Kepuasan Pelanggan	Fibria Anggraini Puji Lestari (2018)[26]	Kesesuaian pelayanan pemerintah yang dirasakan masyarakat dalam memenuhi kebutuhan air, dibutuhkan pengukuran pelayanan tersebut untuk mengetahui kepuasan pelanggan terhadap pelayanan yang diberikan	Metode penelitian kuantitatif menggunakan survey, wawancara dan observasi kepada responden, alat ukur pelayanan SERQUAL. Hasil *(reliability, tangibles, responsiveness, assurance, dan empaty)* memiliki pengaruh pada peningkatan kualitas kepuasan pelanggan
2	Penerapan Algoritma Naive Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Pulsa	Nawangsih dan Setyaningsih (2020) [27]	Permintaan pulsa yang meningkat membuat RA *Cell* sulit mengetahui stok produk yang tinggi pembelianya.	Metode klasifikasi penjualan laris dan tidak laris pada produk pulsa konter RA *cell* yaitu pulsa Telkomsel menggunakan Algoritma Naive Bayes mendapatkan hasil akurasi sebesar 97,50%, nilai *precision* 100 % dan nilai *recall* 93,48%.
3	Klasifikasi Penjualan Obat Pertanian Laris Dan Kurang Laris Pada UD Cahaya Tani menggunakan Metode Decission Tree	Nurhidayati dan Alimuddin (2019) [28]	UD. Cahaya Tani Sulit menentukan stok produk pestisida yang banyak diminati pelanggan.	Metode Klasifikasi dengan Algoritma *Decision Tree* C4,5 menerapkan 10 *cross validation* dan mendapatkan hasil Akurasi teringgi yaitu k=8 dengan nilai 97,43%.

No.	Judul Penelitian	Nama dan Tahun Peneliti	Masalah	Metode dan Hasil
4	Klasifikasi Nanas Layak Jual Dengan Metode Naïve Bayes Dan K-Nearest Neighbor	T. Jaya (2019) [23]	Buah nanas rentan terhdapat penyakit sehingga menyebabkan para petani harus melakukan proses analisis sesuai dengan kriteria bahwa nanas layak atau tidak untuk dikonsumsi.	Dari metode klasifikasi Algoritma Naïve Bayes memiliki nilai akurasi sebesar 73,3%, sedangkan Algoritma K-NN memiliki akurasi sebesar 53,3% pada proses klasifikasi layak atau tidak nanas yang dijual.
5	Perbandingan Klasifikasi Antara KNN Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation	F.Tempola dkk (2018) [29]	Gunung berapi di Indonesia hampir setiap tahun meletus, dibutuhkanya deteksi gempa yang menandakan jika gempa tersebut merupakan tanda gunung berapi akan meletus	Peneletian ini menggunakan metode klasifikasi pada Algoritma K-NN dan memperoleh nilai akurasi 63,68 % dan Naïve bayes sebesar 79,71 %.
6	Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5	Fida Hana (2020) [30]	Berkembangnya penderita diabetes menurut sumber data *Federasi Diabetes Internasional*, pengidap penyakit diabetes ada 10 juta jiwa di tahun 2015, tahun 2040 diprediksi mengalami peningkatan sebanyak 16.2 juta jiwa penduduk Indonesia.	Klasifikasi adalah metode yang diterapkan dalam melakukan penelitian dengan menggunakan Algoritma C4.5 dan mendapatkan hasil akurasi 97,12 % *Precision* 93,02% %, dan *Recall* 100,00%.

Penelitian [27] menunjukan bahwa faktor kualitas pelayanan memiliki pengaruh pada kepuasan pelanggan, penelitian [28][24][29] faktor Laris dan Kurang Laris menjadi hal yang penting dalam dunia bisnis untuk mendapatkan omset yang diinginkan. Karena hal itu peneliti melakukan penelitian klasifikasi produk Zam-Zam Time tergolong Laris atau Kurang Laris.

Berdasarkan tingkat kepuasan pelanggan. Klasifikasi dilakukan menggunakan 2 algoritma sehingga dilakukan komparasi dan mendapatkan algoritma dengan kinerja terbaik untuk menunjang proses evaluasi Zam-Zam Time. Perbedaan penelitian rujukan dengan penelitian ini adalah dataset yang bersifat privat atau primer pada Zam-Zam Time yang sudah tervalidasi kebenaranya. Berdasarkan penelitian [24][27][28][29][30] mengenai Algoritma C4.5 dan Naïve Bayes merupakan algoritma yang memiliki hasil akurasi yang baik. Algoritma C4.5 memiliki keunggulan dalam proses pembentukan pohon keputusan dan Naïve

Bayes memiliki keunggulan proses perhitungan yang sederhana dan bisa menangani pada data kecil. Pada penelitian [27][28] tidak menyertakan penilaian tingkat kepuasan pelanggan. Pada [24][27][28][29][30] tidak dilakukannya pengujian berdasarkan waktu komputasi pada Algoritma yang digunakan dalam proses klasifikasi. Proses analisis klasifikasi kinerja Algoritma C4.5 dan Naïve Bayes pada penelitian ini dilakukan dengan adanya waktu komputasi sehingga membuat proses validasi kinerja semakin akurat.

2.2 Dasar Teori

2.2.1 Produk Zam -Zam Time

Produk merupakan semua hal yang bisa ditawarkan, dimiliki, digunakan bahkan dikonsumsi untuk memenuhi sebuah keinginan pelanggan [31] . Produk memiliki beberapa ciri atribut yaitu ukuran, warna, rasa, kemasan serta pelayanan yang mendapat perhatian dari pelanggan [32] . Produk Zam-Zam Time merupakan sebuah produk minuman rasa yang dapat dinikmati dikala terik atau hujan. Varian rasa pada produk Zam-Zam Time terdiri dari 8 varian rasa yaitu *Coffee Latte, Hazelnut, Vanila Blue, Redvelvet, Choco Dark, Choco Avocado, Avocado, Taro*. Pada Gambar 2.1 merupakan kemasan produk Zam-Zam Time.

Gambar 2. 1 Varian Rasa Zam-Zam Time

2.2.2 Uji Validitas dan Uji Reliabilitas Tingkat Kepuasan Pelanggan

Validitas merupakan sebuah atribut yang dapat dibuktikan dengan data sebenarnya yang bertujuan mengukur kebenaran atribut yang digunakan [33] [34] . Pada penelitian ini menerapkan uji validitas atribut, dengan cara mengkorelasikan skor atribut dengan skor total selaras dengan penelitian dodi dkk [35] . Pengujian

dapat dilakukan dengan mencari koefisien korelasi hasil uji atribut dengan uji kriterianya dapat terlihat seperti pada persamaan korelasi Pearson (2.1).

$$r_{xy} = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{(n(\sum x_1^2) - (x_i)^2)(n(\sum y_1^2) - (y_i)^2)}} \quad (2.1)$$

r_{xy}= Koefisien korelasi atribut atau pertanyaan

$n =$ Jumlah Responden

$x_i =$ Skor atribut pada setiap percobaan pertama

y_i= Skor atribut pada setiap percobaan selanjutnya

Kemudian untuk menentukan valid tidaknya sebuah koefisien korelasi dapat dilakukan dengan membandingkan koefisien korelasi r menggunakan tabel momen produk, ketika nilai r hitung bandingkan dengan r tabel. Tabel 2. 2 merupakan tabel momen[36].

Tabel 2. 2 Tabel Momen Validitas

N	Taraf Signifikan		N	Taraf Signifikan		N	Taraf Signifikan	
	5%	1%		5%	1%		5%	1%
3	0.997	0.999	27	0.381	0.487	55	0.266	0.345
4	0.950	0.990	28	0.374	0.478	60	0.254	0.330
5	0.878	0.959	29	0.367	0.470	65	0.244	0.317
6	0.811	0.917	30	0.361	0.463	70	0.235	0.306
7	0.754	0.874	31	0.355	0.456	75	0.227	0.296
8	0.707	0.834	32	0.349	0.449	80	0.220	0.286
9	0.666	0.798	33	0.344	0.442	85	0.213	0.278
10	0.632	0.765	34	0.339	0.436	90	0.207	0.270
11	0.602	0.735	35	0.334	0.430	95	0.202	0.263
12	0.576	0.708	36	0.329	0.424	100	0.195	0.256
13	0.553	0.684	37	0.325	0.418	125	0.176	0.230
14	0.532	0.661	38	0.320	0.413	150	0.159	0.210
15	0.514	0.641	39	0.316	0.408	175	0.148	0.194
16	0.497	0.623	40	0.312	0.403	200	0.138	0.181
17	0.482	0.606	41	0.308	0.398	300	0.113	0.148
18	0.456	0.590	42	0.304	0.393	400	0.098	0.128
19	0.456	0.575	43	0.301	0.389	500	0.088	0.115
20	0.444	0.561	44	0.297	0.384	600	0.080	0.105
21	0.433	0.549	45	0.294	0.380	700	0.074	0.097
22	0.423	0.537	46	0.291	0.376	800	0.070	0.091
23	0.413	0.526	47	0.288	0.372	900	0.065	0.086

N	Taraf Signifikan		N	Taraf Signifikan		N	Taraf Signifikan	
	5%	1%		5%	1%		5%	1%
24	0.404	0.515	48	0.284	0.368	1000	0.062	0.081
25	0.396	0.505	49	0.281	0.364			
26	0.388	0.496	50	0.279	0.361			

Reliabilitas merupakan sebuah pengukuran dengan ketepatan dari sebuah atribut dari alat pengukuran yaitu kuesioner. Ketepatan yang dimaksud mengacu pada tingkat konsistensi hasil pengukuran meski sudah dilakukan berkali–kali [37] . Perhitungan reliabilitas dapat dilakukan jika atribut pada kuesioner sudah lolos pada tahap validasi dan dinyatakan valid. Uji reliabilitas dapat dihitung dengan persamaan *Cronbach's alpha (α)* [34] .

1. Menentukan nilai varian soal (2.2).

$$\sigma_i^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n} \quad (2.2)$$

2. Menentukan nilai varian total (2.3).

$$\sigma_T^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n} \quad (2.3)$$

3. Menentukan reliabilitas instrument (2.4).

$$r_{11} = \left[\frac{k}{(k-1)}\right]\left[1 - \frac{\sum \sigma_b^2}{\sigma_T^2}\right] \quad (2.4)$$

Ket :

n = jumlah sample

X_i= jawaban subjek masing-masing butir soal

$\sum X$= total jawaban subjek masing-masing butir soal

σ_T^2 = varial total

$\sum \sigma_b^2$= jumalah varian butir

k = jumlah butir soal

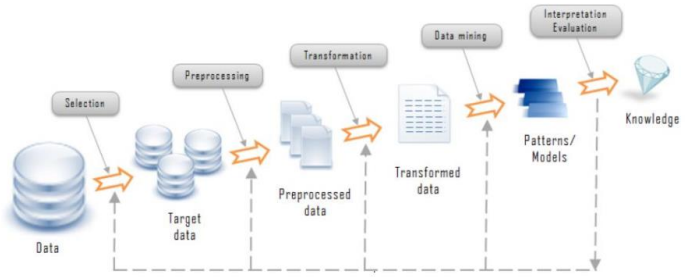
r_{11}= koefisien reliabilitas instrumen

Uji validitas dan uji reliabilitas diterapkan dalam perhitungan kuesioner. Kuesioner berisi daftar pernyataan yang terstruktur sehingga responden mampu menjawab sesuai dengan keadaan yang dialami [38]. Hasil kuesioner digunakan dalam perhitungan pada teknik pengumpulan data dalam menganalisis sikap,

pengetahuan, kepercayaan atau karakteristik seseorang yang dapat mempengaruhi sistem yang ada disekitarnya. Bentuk kuesioner yang diterapkan penelitian ini adalah mengaplikasikan *Skala Likert* dimana responden akan diberikan pilihan jawaban berupa tingkat kepuasan. Pilihan yang tersedia pada kuesioner yaitu Tidak Puas(1), Kurang Puas(2), Puas(3), Sangat Puas(4) pilihan pada penelitian ini selaras dengan penelitian [39] . Kuesioner yang disebar kepada responden akan mengeluarkan hasil data yang memiliki informasi terkait pada penelitian ini. Informasi yang dihasilkan menjadi sejalan, tercipta dari kuesioner yang memenuhi nilai standar validitas dan reliabilitas.

2.2.3 Data Mining

Data mining didefinisikan sebuah teknik mengolah data menggunakan pola tertentu, data yang diperoleh berdasarkan *database* yang tersedia [40] . Data mining dapat diimplementasikan untuk data komparasi [41] , pengolahan multidimensi dalam membuat informasi pohon keputusan untuk menghindari perulangan dan memutus model yang dibuat [42] . Kegiatan merangkum menganalisa *database* sebagai penghubung yang mudah dimengerti dan memiliki manfaat bagi pihak terkait [43] . KDD *(Knowledge Discovery In Databases)* ialah proses analisis data yang berukuran besar dari yang sulit dimengerti sehingga mudah untuk dipahami oleh manusia adapun tahapan KDD tertera pada Gambar 2.2.



Gambar 2. 2 Tahap KDD

Tahapan KDD diterangkan dalam penjelasan dibawah ini :

1. *Selection*

Pemilihan data sejenis dari data operasional sebelum proses analisis belum dimulai.

2. *Preprocessing*

Pada tahap ini data akan memasuki proses *cleansing* dimana data akan diperiksa yang tidak konsisten, tidak beraturan, memiliki kesalahan penulisan, duplikasi data akan dibuang.

3. Transformation

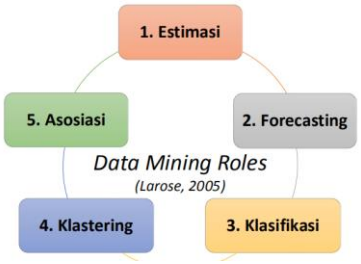
Data akan di transformasi dari tipe sebelumnya menjadi tipe data yang sesuai menggunakan *coding* yang ditentukan dalam proses data mining.

4. Data mining

Dilakukan proses mencari pola dengan informasi yang menarik memanfaatkan teknik , algoritma atau metode tertentu.

5. Evaluasi

Proses identifikasi pola yang dihasilkan pada tahap data mining yang ditampilkan sesuai pola yang menarik dan hasil pada masing–masing pihak terkait sehingga memperoleh sebuah visualisasi pengetahuan yang mudah dimengerti [44].



Gambar 2. 3 Bidang Data Mining

Gambar 2.3 merupakan pengelompokan bidang data mining [1] terbagi dalam masing–masing fungsi dan tujuan [45] :

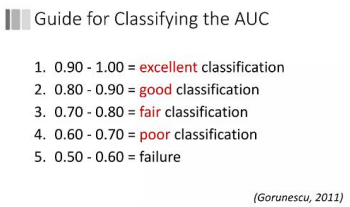
1. Estimasi atau disebut juga dengan perkiraan, perbedaan estimasi dengan klasifikasi adalah pada bentuk pengelompokan numerik bukan menggunakan kategorikal.
2. *Forecasting* atau prediksi, fungsi dan tujuan yang hampir sama dengan klasifikasi, prediksi ini digunakan untuk memprediksi nilai dan hasil dimasa depan.
3. Klasifikasi, pengelompokan yang berdasarkan pada jenis atau tipe yang sama.
4. Klastering, kelompok data dengan kemiripan nilai. Bentuk data pada pengklasteran ada hasil observasi, rekam data, kelas, dan objek yang memiliki kesamaan.

5. Asosiasi adalah sebuah pengelompokan, pengumpulan data yang muncul dalam waktu bersamaan.

2.2.4 Klasifikasi

Klasifikasi dapat didefinisikan sebagai salah satu cara pengelompokan dengan ciri yang dipunya objek klasifikasi [46] . Data yang telah mengalami pengelompokan dapat diselesaikan menggunakan beberapa algortima klasifikasi yaitu *Decision tree, K-Nearest Neighbour, Naïve Bayes, , Support Vector Machine* [47] . Klasifikasi menggunakan proses untuk menguraikan data penting serta menduga variabel dimasa depan dalam memvisualisasikan kelas atau konsep data tersebut [48].

Klasifikasi pada penelitian ini dilakukan untuk menentukan objek pada kategori tertentu salah satu pemecahan masalah dalam penelitian yaitu mengklasifikasikan produk Zam-Zam Time " Laris" atau " Kurang Laris" data yang digunakan bersumber dari penilaian yang dilakukan oleh pelanggan, data tersebut digunakan untuk melatih dan menguji sehingga mendapat hasil algoritma terbaik untuk menangani dataset yang digunakan. Pada Gambar 2.4 merupakan tingkatan klasifikasi akurasi [49].



Gambar 2. 4 AUC Classifying Accuracy

2.2.5 Preprocessing Data

Tahap preprocessing ialah tahap analisis kelengkapan data yang digunakan, data yang memiliki keunggulan tinggi maka menghasilkan hasil akhir dengan kualitas tinggi. Pada tahap ini dilakukan proses pelabelan status dengan kategori yang sudah ditentukan oleh pihak Zam-Zam Time pada seluruh data yang akan digunakan untuk meningkatkan keunggulan data dalam membantu menaikan nilai akurasi dan waktu komputasi. Setelah proses pelabelan selesai kemudian dilakukan

proses seleksi fitur. Seleksi fitur dapat dilakukan dengan penilaian kemampuan dengan metode pemilihan fitur (atribut) untuk menerapkan urutan penting dari fitur tersebut pada proses klasifikasi [50] .

2.2.6 Algoritma C4.5

Algoritma C4.5 banyak penerapan dalam membuat pohon keputusan memiliki tujuan untuk menambah nilai akurasi dari prediksi yang ada dengan hasil yang mudah dimengerti [51] . Algoritma C4.5 memiliki beberapa cabang dan masing–masing mewakili atribut hingga semua terpenuhi dan berakhir [52] . Langkah dalam membuat pohon keputusan Algoritma C4.5 yang dimaksud dengan memilih atribut yang akan digunakan sebagai akar pohon, membuat cabang dari masing–masing nilai yang akan dicari, dari nilai yang ada dibagi lagi sesuai dengan kasus masing–masing cabang yang diperlukan, dilakukan proses perulangan hingga hasil yang ditemukan memiliki kelas yang sama [53] . Tahap dalam membuat pohon keputusan pada Algoritma C4.5 yaitu:

1. Mempersiapkan data yang dibagi 2 yang diterapkan untuk data *Training* dan *Testing.*
2. Tentukan akar dari setiap pohon, menghitung nilai *Information Gain* tertinggi dari masing–masing atribut yang digunakan untuk ambil nilai *index entropy* terendah, seperti yang terlihat pada persamaan (2.5).

   $$\boldsymbol{Entropy(S) = \sum_{i=1}^{n} -pi * log_2 pi} \quad (2.5)$$

   Ket:

   S: kumpulan kasus

   pi: Atribut
3. Hitung nilai *Information Gain* dengan rumus, seperti yang terlihat pada persamaan (2.6):

   $$\boldsymbol{Information\ Gain(S, A) = Entropy(S) - \sum_{I-1}^{N} \frac{|si|}{|s|} * Entropy(si)} \quad (2.6)$$

   Ket:

   S: Himpunan kasus

   A: Atribut

N: Jumlah partisi atribut A

|si|: Jumlah kasus partisi ke-i

|s|: jumlah kasus pada S

Contoh kasus pada penggunaan Algoritma C4.5 [51]

Tabel 2.3 berisikan dataset yang digunakan hasil dari kuesioner konsumen Bengkel Zul Keluarga Jaya Pematangsiantar dan data sampel menggunakan penilaian skala yang terdiri dari : 1 = Sangat tidak puas, 2 = Kurang puas, 3 = Cukup puas, 4 = Puas, 5 = Sangat puas .

Tabel 2. 3 Dataset Kepuasan Konsumen Bengkel Zul Keluarga Jaya Pematangsiantar

No	Umur	Gender	Mahir	Tanggap	Jaminan	Empati	Wujud	Tanggapan
1	36	Wanita	4	5	5	4	4	Puas
2	27	Pria	5	4	5	5	4	Puas
3	32	Wanita	4	4	3	4	4	Puas
…	…	…	…	…	…	…	…	…
98	52	Pria	5	5	5	5	5	Puas
99	40	Pria	4	4	3	1	3	Tidak Puas
100	66	Pria	4	4	4	4	4	Puas

Algoritma C4.5 mendapatkan aturan pohon keputusan dengan cara :

Langkah 1 = Hitung jumlah *record* dengan tanggapan puas dan tidak puas.

Langkah 2 = Hitung nilai *Entropy* dari semua *record* yang dibagi berlandasakan kelas atribut dan persamaan, kemudian dilakukan perhitungan *Information Gain* untuk atribut dan persamaanya.

Perhitungan nilai *Entropy* dan *Information Gain* :

Node 1

1. Hitung *Entropy* total.

$$Entropy\ [Total] = \left(-\frac{83}{100} x\ log_2\left(\frac{83}{100}\right)\right) + \left(-\frac{17}{100} x\ log_2\left(\frac{17}{100}\right)\right) = 0.657704779$$

2. Hitung *Entropy* dan *Information Gain* dengan masing–masing atribut.
   a. Mahir

$$Entropy\ [Mahir - 1] = 0$$

$$Entropy\ [Mahir-2] = \left(-\frac{0}{11}x\ log_2\left(\frac{0}{11}\right)\right) + \left(-\frac{11}{11}x\ log_2\left(\frac{11}{11}\right)\right) = 0$$

$$Entropy\ [Mahir-3] = \left(-\frac{8}{13}x\ log_2\left(\frac{8}{13}\right)\right) + \left(-\frac{5}{13}x\ log_2\left(\frac{5}{13}\right)\right) = 0.961236605$$

$$Entropy\ [Mahir-4] = \left(-\frac{56}{57}x\ log_2\left(\frac{56}{57}\right)\right) + \left(-\frac{1}{57}x\ log_2\left(\frac{1}{57}\right)\right) = 0.127418512$$

$$Entropy\ [Mahir-5] = \left(-\frac{19}{19}x\ log_2\left(\frac{19}{19}\right)\right) + \left(-\frac{0}{19}x\ log_2\left(\frac{0}{19}\right)\right) = 0$$

$$Information\ Gain\ [Mahir]$$
$$= 0{,}657704779 - \left(\left(\frac{0}{100}x0\right) + \left(\frac{11}{100}x0\right) + \left(\frac{13}{100}x0.961236605\right) + \left(\frac{57}{100}x0.127418512\right) + \left(\frac{19}{100}x0\right)\right) = 0.460115468$$

b. Tanggap

$$Entropy\ [Tanggap-1] = 0$$

$$Entropy\ [Tanggap-2] = \left(-\frac{0}{8}x\ log_2\left(\frac{0}{8}\right)\right) + \left(-\frac{8}{8}x\ log_2\left(\frac{8}{8}\right)\right) = 0$$

$$Entropy\ [Tanggap-3] = \left(-\frac{16}{23}x\ log_2\left(\frac{16}{33}\right)\right) + \left(-\frac{7}{23}x\ log_2\left(\frac{7}{23}\right)\right) = 0.886540893$$

$$Entropy\ [Tanggap-4] = \left(-\frac{43}{45}x\ log_2\left(\frac{43}{45}\right)\right) + \left(-\frac{2}{45}x\ log_2\left(\frac{2}{45}\right)\right) = 0.262311220$$

$$Entropy\ [Tanggap-5] = \left(-\frac{24}{24}x\ log_2\left(\frac{24}{24}\right)\right) + \left(-\frac{0}{24}x\ log_2\left(\frac{0}{24}\right)\right) = 0$$

$$Information\ Gain\ [Tanggap]$$
$$= 0.657704779 - \left(\left(\frac{0}{100}x0\right) + \left(\frac{8}{100}x0\right) + \left(\frac{23}{100}x0.886540893\right) + \left(\frac{45}{100}x0.262311220\right) + \left(\frac{24}{100}x0\right)\right) = 0.335760325$$

c. Jaminan

$$Entropy\ [Jaminan-1] = \left(-\frac{0}{1}x\ log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1}x\ log_2\left(\frac{1}{1}\right)\right) = 0$$

$$Entropy\ [Jaminan-2] = \left(-\frac{2}{9}x\ log_2\left(\frac{2}{9}\right)\right) + \left(-\frac{7}{9}x\ log_2\left(\frac{7}{9}\right)\right) = 0.764204057$$

$$Entropy\ [Jaminan-3] = \left(-\frac{15}{24}x\ log_2\left(\frac{15}{24}\right)\right) + \left(-\frac{9}{24}x\ log_2\left(\frac{9}{24}\right)\right) = 0.954434003$$

$$Entropy\ [Jaminan-4] = \left(-\frac{46}{46}x\ log_2\left(\frac{46}{46}\right)\right) + \left(-\frac{0}{46}x\ log_2\left(\frac{0}{46}\right)\right) = 0$$

$$Entropy\ [Jaminan-5] = \left(-\frac{20}{20}x\ log_2\left(\frac{20}{20}\right)\right) + \left(-\frac{0}{20}x\ log_2\left(\frac{0}{20}\right)\right) = 0$$

$$Information\ Gain\ [Jaminan]$$
$$= 0.657704779 - \left(\left(\frac{1}{100}x0\right) + \left(\frac{9}{100}x0.764204057\right) + \left(\frac{24}{100}x0.954434003\right) + \left(\frac{46}{100}x0\right) + \left(\frac{20}{100}x0\right)\right) = 0.359862212$$

d. Empati

$$Entropy\ [Empati - 1] = \left(-\frac{0}{2}x\ log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2}x\ log_2\left(\frac{2}{2}\right)\right) = 0$$

$$Entropy\ [Empati - 2] = \left(-\frac{1}{10}x\ log_2\left(\frac{1}{10}\right)\right) + \left(-\frac{9}{10}x\ log_2\left(\frac{9}{10}\right)\right) = 0.46895594$$

$$Entropy\ [Empati - 3] = \left(-\frac{16}{22}x\ log_2\left(\frac{16}{22}\right)\right) + \left(-\frac{6}{22}x\ log_2\left(\frac{6}{22}\right)\right) = 0.845350937$$

$$Entropy\ [Empati - 4] = \left(-\frac{42}{42}x\ log_2\left(\frac{42}{42}\right)\right) + \left(-\frac{0}{42}x\ log_2\left(\frac{0}{42}\right)\right) = 0$$

$$Entropy\ [Empati - 5] = \left(-\frac{24}{24}x\ log_2\left(\frac{24}{24}\right)\right) + \left(-\frac{0}{24}x\ log_2\left(\frac{0}{24}\right)\right) = 0$$

$$Information\ Gain\ [Empati]$$
$$= 0{,}657704779 - \left(\left(\frac{2}{100}x0\right) + \left(\frac{10}{100}x0.468995594\right) + \left(\frac{22}{100}x0.845350937\right) + \left(\frac{42}{100}x0\right) + \left(\frac{24}{100}x0\right)\right) = 0.424828013$$

e. Wujud

$$Entropy\ [Wujud - 1] = \left(-\frac{0}{3}x\ log_2\left(\frac{0}{3}\right)\right) + \left(-\frac{3}{3}x\ log_2\left(\frac{3}{3}\right)\right) = 0$$

$$Entropy\ [Wujud - 2] = \left(-\frac{0}{6}x\ log_2\left(\frac{0}{6}\right)\right) + \left(-\frac{6}{6}x\ log_2\left(\frac{6}{6}\right)\right) = 0$$

$$Entropy\ [Wujud - 3] = \left(-\frac{18}{25}x\ log_2\left(\frac{18}{25}\right)\right) + \left(-\frac{7}{25}x\ log_2\left(\frac{7}{25}\right)\right) = 0.8554550811$$

$$Entropy\ [Wujud - 4] = \left(-\frac{44}{45}x\ log_2\left(\frac{44}{45}\right)\right) + \left(-\frac{1}{45}x\ log_2\left(\frac{1}{45}\right)\right) = 0.153742180$$

$$Entropy\ [Wujud - 5] = \left(-\frac{21}{21}x\ log_2\left(\frac{21}{21}\right)\right) + \left(-\frac{0}{21}x\ log_2\left(\frac{0}{21}\right)\right) = 0$$

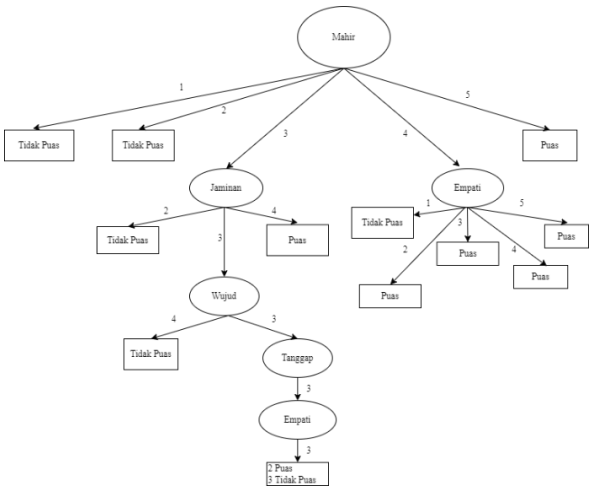
$$Information\ Gain\ [Wujud]$$
$$= 0.657704779 - \left(\left(\frac{3}{100}x0\right) + \left(\frac{6}{100}x0\right) + \left(\frac{25}{100}x0.855450811\right) + \left(\frac{45}{100}x0.153742180\right) + \left(\frac{21}{100}x0\right)\right) = 0.374658095$$

Dari perhitungan diatas dapat dilihat dengan hasil pada Tabel 2.4.

Tabel 2. 4 Hasil Perhitungan Node 1

Node 1		Jumlah record	Puas	Tidak Puas	*Entropy*	*Information Gain*
TOTAL		100	83	17	0.657704779	
Mahir						**0.460115468**
	1	0	0	0	0	
	2	11	0	11	0	
	3	13	8	5	0.961236605	
	4	57	56	1	0.127418512	
	5	19	19	0	0	
Tanggap						0.335760325
	1	0	0	0	0	
	2	8	0	8	0	
	3	23	16	7	0.886540893	
	4	45	43	2	0.26231122	
	5	24	24	0	0	
Jaminan						0.359862212
	1	1	0	1	0	
	2	9	2	7	0.764204507	
	3	24	15	9	0.954434003	
	4	46	6	0	0	
	5	20	20	0	0	
Empati						0.424828013
	1	2	0	2	0	
	2	10	1	9	0.468995594	
	3	22	16	6	0.468995594	
	4	42	42	0	0	
	5	24	24	0	0	
Wujud						0.374658095
	1	3	0	3	0	
	2	6	0	6	0	
	3	25	18	7	0.855450811	
	4	45	44	0	0.15374218	
	5	21	21	0	0	

Gambar 2.4 merupakan hasil pohon keputusan perhitungan manual dengan Algoritma C4.5.



Gambar 2. 5 Hasil Pohon Keputusan Perhitungan Manual Dengan Algoritma C4.5

Terdapat 12 aturan perhitungan yang bisa dibuat sebagai acuan untuk menentukan kepuasan pelanggan mengenai pelayanan *Service* di Bengkel Zul Family Jaya Pematangsiantar. Aturan yang terbentuk melihat dari pohon keputusan Gambar 2.4 adalah 6 aturan keputusan puas dan aturan keputusan tidak puas, diuraikan dengan teks narasi:

1. Bila mahir =1, lalu *output* tidak puas {Puas = 0, Tidak Puas =3}
2. Bila mahir = 2, lalu *output* tidak puas{Puas = 0, Tidak Puas = 8}
3. Bila mahir = 3 dan jaminan= 2, lalu *ouput* tidak puas {Puas = 0, Tidak Puas = 1}
4. Bila mahir = 3, jaminan = 3, wujud = 3, tanggap = 3, empati = 3 lalu *output* tidak puas {Puas = 2, Tidak Puas = 3}
5. Bila mahir = 3, jaminan = 3 dan wujud = 4, lalu *output* tidak puas {Puas = 0, Tidak Puas = 1}
6. Bila mahir = 3 dan jaminan = 4, lalu *output* puas {Puas = 6, Tidak Puas = 0}
7. Bila mahir = 4 dan empati = 1, lalu *output* tidak puas {Puas = 0, Tidak Puas = 1}
8. Bila mahir = 4 dan empati = 2, lalu *output* puas {Puas = 1, Tidak Puas = 0}
9. Bila mahir = 4 dan empati = 3, lalu *output* puas {Puas = 12, Tidak Puas = 0}
10. Bila mahir = 4 dan empati = 4, lalu *output* puas {Puas = 31, Tidak Puas = 0}
11. Bila mahir = 4 dan empati = 5, lalu *output* puas {Puas = 12, Tidak Puas = 0}

12. Bila mahir = 5, lalu output puas {Puas = 19, Tidak Puas = 0}

2.2.7 Algoritma Naïve Bayes

Naïve Bayes mengacu pada asumsi yang dibuat oleh metode, bahwa pengaruh nilai satu atribut pada probabilitas klasifikasi tertentu tidak bergantung pada nilai atribut lainnya. Kemampuan dan probabilitas bersyarat dalam satu rumus, yang dapat kita gunakan menghitung probabilitas dari setiap kemungkinan klasifikasi secara bergantian, kemudian memilih klasifikasi dengan nilai terbesar [54]. Naïve bayes menerapkan cabang matematika dan dikenal sebagai teori probabilitas dalam mencari peluang dari banyaknya kemungkinan yang ada, dan tetap memperhatikan frekuensi pada klasifikasi data *Traning* [55]. Tipe Naïve Bayes Bernoulli Naive Bayes memiliki pokok hasil ya/tidak, variabel digunakan tipe boolean. Multinomial Naive Bayes Sebagian besar diterapkan pada klasifikasi dokumen, menerapkan tipe data variabel kata. Gaussian Naive Bayes dapat menangani data diskrit atau kontinu dengan data akan diasumsikan Sebagai sampel distribusi gausian. Oleh karena itu, pada penelitian ini menggunakan Gaussian Naïve Bayes yang disesuaikan pada tipe data yang digunakan pada penelitian ini [56]. Konsep Naïve bayes menerapkan teorema kuno yang ditemukan *Thomas Bayes* pada abad ke 18 yang dinyatakan seperti yang terlihat pada persamaan (2.7).

$$\boldsymbol{P(X|H)} = \frac{\boldsymbol{P(H|X)P(X)}}{\boldsymbol{P(H)}}\textbf{;}\ \mathrm{P(H)}\neq 0 \qquad (2.7)$$

Ket:

H = Data kelas belum diketahui

X = Hipotesis dari B yang merupakan suatu class spesifik

P(A|H) = Probabilitas H maka A

P(B|H) = Probabilitas H maka B

P(B) = Probabilitas B

P(A)= Probabilitas A

Proses hitung P (B_I) dimana probabilitas *prior* pada masing–masing sub kelas B menghasilkan seperti yang terlihat pada persamaan(2.8).

$$P(B_i) = \frac{Si}{s} \qquad (2.8)$$

Keterangan:

Si: Total data *Training* dari kategori

s: Total data

Distribusi *Gaussian* digunakan untuk menunjukan kemungkinan bersyarat pada atribut numerik dari kelas $P(A_i|B)$ dengan ciri dua parameter yaitu mean μ dan varian σ^2. Dengan kemungkinan bersyarat kelas B_j untuk atribut A_j seperti pada persamaan (2.9):

$$P(A_i = a_i|B = b_j) = \frac{1}{\sigma_{ij}\sqrt{2\pi}} exp^{-\frac{(a_i-\mu_{ij})^2}{2\sigma_{ij}^2}} \quad (2.9)$$

σ= Standar Deviasi

$\mu \quad = mean$

Contoh kasus pada penggunaan Algoritma Naïve bayes [57] :

Tabel 2.5 merupakan contoh data dummy yang digunakan pada perhitungan Naïve Bayes klasifikasi menggunakan data *Traning* klasifikasi hewan.

Tabel 2. 5 Data Latih Klasifikasi Hewan

Nama Hewan	Penutup Kulit	Melahirkan	Berat	Kelas
Ular	Sisik	Ya	10	Reptil
Tikus	Bulu	Ya	0.8	Mamalia
Nama Hewan	Penutup Kulit	Melahirkan	Berat	Kelas
Kambing	Rambut	Ya	21	Mamalia
Sapi	Rambut	Ya	120	Mamalia
Kadal	Sisik	Tidak	0.4	Reptil
Kucing	Rambut	Ya	1.5	Mamalia
Bekicot	Cangkang	Tidak	0.3	Reptil
Harimau	Rambut	Ya	43	Mamalia
Rusa	Rambut	Ya	45	Mamalia
Kura - kura	Cangkang	tidak	7	Reptil

Atribut dengan tipe numerik adalah berat sedangkan untuk *mean* dan varian atribut masing–masing kelas dapat dihitung dengan beberapa langkah yaitu

Langkah 1:

A_i ($\bar{a}$) merupakan sampel dari total data latih milik B_j yang selanjutnya dijadikan sampel *mean* untuk digunakan sebagai parameter dari μ_{ij}.

$$\bar{a}\ mamalia = \frac{0.8+21+120+1.5+43+45}{6} = \frac{231.3}{6} = 38.55$$

$$\bar{a}\ reptil = \frac{10+0.4+0.3+7}{4} = \frac{17.7}{4} = 4.425$$

Langkah 2:

Selanjutnya pada varian sampel (s^2) didapat dari data latih yang akan digunakan untuk parameter σ_{ij}^2.

Mencari sampel mamalia :

$$s_{mamalia}^2 = \frac{(0.8-38.55)^2 + (21-38.55)^2 + (120-38.55)^2 + (1.5-38.55)^2 + (43-38.55)^2 + (45-38.55)^2}{6-1} = \frac{9801.275}{5} = 1960.25$$

$$s_{mamalia} = \sqrt{1960.25} = 44.275$$

Mencari sampel *reptile*:

$$s_{reptil}^2 = \frac{(10-4.425)^2 + (0.4-4.425)^2 + (0.3-4.425)^2 + (7-4.425)^2}{4-1} = \frac{70.9275}{3} = 23.6425$$

$$s_{reptil} = \sqrt{23.6425} = 4.8624$$

Langkah 3:

Setelah didapatkan nilai dari parameter yang diinginkan selanjutnya dilakukan perhitungan nilai probabilitas dengan sampel dan data uji hewan musang dengan atribut penutup kulit = rambut, melahirkan = ya, berat = 15, yaitu :

$$P(Berat = 15|Mamalia) = \frac{1}{44.275\sqrt{2\pi}} exp^{-\frac{(15-38.55)^2}{2x1960.255}} = 0.0104$$

$$P(Berat = 15|Reptil) = \frac{1}{4.8624\sqrt{2\pi}} exp^{-\frac{(15-4.425)^2}{2x23.6425}} = 0.8733$$

Pada Tabel 2.6 merupakan probabilitas atribut dengan kelas penutup kulit.

Tabel 2. 6 Probabilitas Atribut Dan Kelas Penutup Kulit

Penutup Kulit	
Mamalia	Reptil
Sisik =0 Bulu =1 Rambut = 5	Sisik =2 Bulu =0 Rambut = 0

Penutup Kulit	
Mamalia	Reptil
Cangkang = 0	Cangkang = 2
P (kulit = sisik |Mamalia) = 0 P (kulit = bulu |Mamalia) = 1/6 P (kulit = rambut|Mamalia) = 5/6 P (kulit = cangkang |Mamalia) =0	P (kulit = sisik |Reptil) = 0.5 P (kulit = bulu |Reptil) = 0 P (kulit = rambut| Reptil) = 0 P (kulit = cangkang | Reptil) = 0.5

Pada Tabel 2.7 merupakan atribut dengan kelas melahirkan.

Tabel 2. 7 Probabilitas Atribut Dan Kelas Melahirkan

Melahirkan	
Mamalia	Reptil
Ya =6 Tidak =0	Ya = 1 Tidak = 0
P (Lahir = Ya |Mamalia) = 1 P (Lahir = Tidak |Mamalia) = 0	P (Lahir = Ya|Reptil) = 0.25 P (Lahir = Tidak|Reptil) = 0.75

Tabel 2.8 merupakan probabilitas atribut berat dengan kelas.

Tabel 2. 8 Probabilitas Atribut Berat Dan Kelas

Berat		Kelas	
Mamalias	Reptil	Mamalia	Reptil
$\bar{a}_{\text{mamalia}} = 38.55$	$\bar{a}_{\text{reptil}} = 4.425$	Mamalia = 6	Reptil = 4
$s^2_{mamalia} = 1960.255$	$s^2_{reptil} = 23.6425$	P(Mamalia) = 6/10 = 0.6	P(Reptil)=4/10 = 0.4
$S_{\text{mamalia}} = 44.275$	$S_{\text{reptil}} = 4.8624$		

Langkah 4:

Proses dilakukanya perhitungan semua probabilitas untuk masing–masing kelas:

P (A | Mamalia) = P (Kulit = Rambut | Mamalia) x P (Lahir = Ya | Mamalia) x P(Berat = 15|Mamalia)

P(A|Mamalia) = $\frac{5}{6}$ x 1 x 0.0104 = 0.0087

P (A | Reptil) = P (Kulit = Rambut | Reptil) x P (Lahir = Ya | Reptil) x P (Berat = 15 | Reptil)

P (A | Reptil) = 0 x 0.25 x 0.87333 = 0

Langkah 5:

Dilakukan proses perhitungan probabilitas akhir.

P (Mamalia | A) = α x 0.6 x 0.0087 = 0.0052α

P (Reptil | A) = α x 0 x 0.4 = 0

Dari hasil α = 1/P(A) adalah memiliki tetap maka tidak perlu dicari karena tidak akan dipengaruhi oleh P(A). Hasil yang sudah diperoleh adalah nilai probabilitas terbesar dengan terdapat pada kelas mamalia maka data uji musang diperkirakan sebagai kelas mamalia.

2.2.8 K – Fold Cross Validation

K – Fold Cross Validation didefinisikan teknik dalam *Training Testing* yang sering digunakan ketika objek kecil yang berfungsi menilai kinerja algoritma [58]. Jika dataset terdiri N objek dibagi menjadi bagian K sama besar, K terdiri dari nilai 1 – 10 jika N tidak habis dibagi K maka bagian akhir akan memiliki nilai lebih sedikit daripada K lainnya. K run dilakukan masing–masing bagian secara bergantian digunakan sebagai data *Training* dan K lainya digunakan sebagai data uji. Kemudian jumlah objeknya diklasifikasikan dengan benar yang akan dibagi dengan jumlah N yang memberikan tingkat seluruh akurasi P, atau persamaan kesalahan standar [54]. Nilai *Cross Validation* diterapkan 10-fold, dimana *10-fold validation* akan melakukan perulangan. Tabel 2.9 contoh penggunaan *10–fold cross validation.*

Tabel 2. 9 Penggunaan 10-Fold Cross Validation

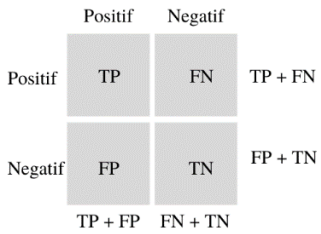
Validation	Dataset									
1	**1**	2	3	4	5	6	7	8	9	10
2	1	**2**	3	4	5	6	7	8	9	10
3	1	2	**3**	4	5	6	7	8	9	10
4	1	2	3	**4**	5	6	7	8	9	10
5	1	2	3	4	**5**	6	7	8	9	10
6	1	2	3	4	5	**6**	7	8	9	10
7	1	2	3	4	5	6	**7**	8	9	10
8	1	2	3	4	5	6	7	**8**	9	10
9	1	2	3	4	5	6	7	8	**9**	10
10	1	2	3	4	5	6	7	8	9	**10**

Training	*Testing*

2.2.9 Confusion Matrix

Confusion matrix diterapkan model klasifikasi yang memiliki nilai terbaik melihat dari perhitungan akurasi, *precision* dan *recall* [59]. *Confusion matrix*

merupakan instrumen tabel yang digunakan menganalisis seberapa besar ketepatan atau kinerja suatu model algoritma dapat digunakan untuk memahami *tuple* data yang berbeda dalam sebuah klasifikasi.



Gambar 2. 6 Model *Confusion Matrix*

Menurut Gambar 2.5 patokan dalam *Confusion Matrix* serta persamaan yang diterapkan yaitu [60] [61]:

1. *True Positive* (TP) adalah seluruh data sebenarnya bernilai Kurang Laris yang berhasil diklasifikasikan sebagai nilai Kurang Laris. Terlihat pada persamaan (2.10).

   $$\text{TPR} = \frac{TP}{TP+FN} \quad (2.10)$$

   Ket:

   TPR = *True Positive Rate*

   TP = *True Positive*

   FN = *False Negative*

2. *False Positive* (FP) adalah seluruh data yang sebenarnya bernilai Laris yang diklasifikasikan sebagai nilai Kurang Laris, terlihat pada persamaan (2.11).

   $$FPR = \frac{FP}{FP+TN} \quad (2.11)$$

   Ket:

   FPR = *False Positive Rate*

   FP = *False Positive*

   TN = *True Negative*

3. *False Negative (FN)* adalah seluruh data yang sebenarnya bernilai Kurang Laris yang diklasifikasikan sebagai nilai Laris, terlihat pada persamaan (2.12).

   $$FNR = \frac{FN}{FN+TP} \quad (2.12)$$

   Ket:

FNR = *False Negative Rate*

FN = *False Negative*

TP = *True Positive*

4. *True Negative* (TN) adalah seluruh data sebenarnya bernilai Laris yang diklasifikasikan sebagai nilai Laris, terlihat pada persamaan (2.13).

$$TNR = \frac{TN}{TN+FP} \quad (2.13)$$

TNR = *True Negative Rate*

TN = *True Negative*

FP = *False Positive*

Nilai yang dikeluarkan dengan *Confusion Matrix* [60][61] [62]adalah

1. Akurasi merupakan sebuah nilai yang melambangkan suatu ketepatan, ketelitian dan keakuratan pada suatu hasil proses klasifikasi secara benar menggunakan algoritma, terlihat pada persamaan (2.14).

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.14)$$

2. *Precision* merupakan tingkat ketepatan hasil sebuah model algoritma yang akan dibandingkan nilai Laris / Kurang Laris dengan total data label Laris / Kurang Laris, seperti yang terlihat pada persamaan (2.15).

$$Precision = \frac{TP}{TP+FP} \quad (2.15)$$

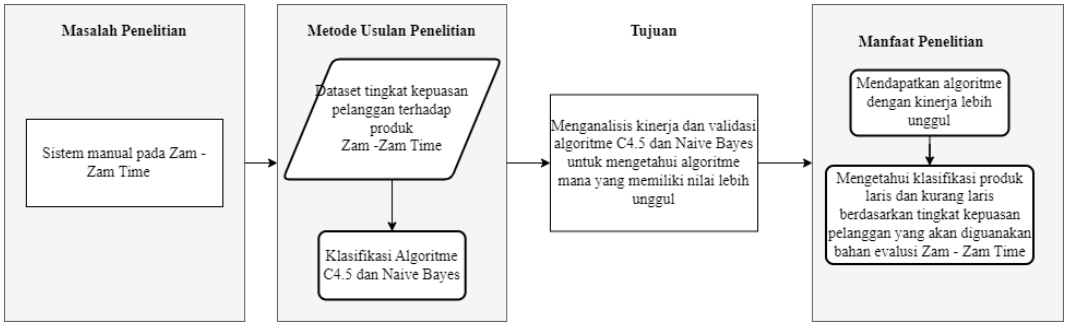
3. *Recall* merupakan parameter kelengkapan suatu model algoritma yang membandingkan total data yang benar – benar bernilai Laris / Kurang Laris (2.16).

$$Recall = \frac{TP}{TP+FN} \quad (2.16)$$

2.2.10 Kerangka Berpikir

Pada penelitian ini akan menghitung akurasi dan waktu komputasi pada klasifikasi produk Zam-Zam Time berdasarkan tingkat kepuasan pelanggan. Data yang digunakan diambil dari dataset privat atau primer hasil kuesioner tingkat kepuasan pelanggan. Dataset yang memiliki atribut sesuai dengan yang dibutuhkan seperti yang terlampir pada Tabel 3.1 Atribut dan deskripsi pada dataset Zam-Zam Time. Metode penelitian ini menggunakan klasifikasi dengan komparasi kinerja

akurasi dan waktu komputasi dari dua Algoritma yaitu C4.5 dan Naïve Bayes. Gambar 2. 6 adalah kerangka pemikiran peneliti.



Gambar 2. 7 Kerangka Berpikir Peneliti