
PENGENALAN POLA KEMAMPUAN PELANGGAN DALAM MEMBAYAR AIR PDAM MENGGUNAKAN ALGORITMA NAIVE BAYES

Putrama Alkhairi^[1]*, P.P.P.A.N.W Fikrul Ilmi R.H.Zer^[2], Ela Roza Batubara^[3], Fazli Nugraha Tambunan^[4], Rika Rosnelly^[5]

^{1,2,3,4,5}Program Studi Ilmu Komputer, Fakultas Teknik dan Ilmu Komputer Universitas Potensi Utama, Jl.K.L Yos Sudarso KM.6,5 No. 3A Tj.Mulia, Medan, Indonesia

email : putramaalkhairi97@gmail.com^[1], fikrulilmizer@gmail.com^[2], elabatubara22@gmail.com^[3], Fazlinugraha313@gmail.com^[4],rikarosnelly@gmail.com^[5]

Abstrak

Dengan meningkatnya jumlah MBR (Masyarakat Berpenghasilan Rendah) yang masuk setiap tahunnya dimasing-masing wilayah di Pematangsiantar, pihak PDAM Tirta Lihou berencana mencari alternatif solusi dalam menangani permasalahan kemampuan pelanggan dalam membayar tagihan air sehingga biaya operasional tetap bisa berjalan baik dan produksi dapat memenuhi kebutuhan masyarakat. Dalam menentukan alternatif untuk menentukan kemampuan masyarakat dalam membayar tagihan air digunakan metode datamining. Dengan menggunakan teknik datamining khususnya klasifikasi menggunakan algoritma Naive Bayes dapat dilakukan prediksi terhadap kemampuan pelanggan dalam membayar tagihan air bersih berdasarkan data yang ada. Naive Bayes adalah teknik prediksi probabilistik sederhana yang berdasarkan pada teorema Bayes dengan asumsi independensi (ketidak tergantungan) yang kuat. Berdasarkan hasil dari perhitungan menggunakan algoritma Naive Bayes, diperoleh hasil klasifikasi dari 30 alternatif yang digunakan, dimana terdapat 11 kelas mampu membayar tagihan dan 19 Tidak Mampu dengan total Accuracy yang diperoleh sebesar 70%. Dari hasil yang diperoleh, diharapkan penelitian ini dapat membantu pihak PDAM Tirta Lihou dalam menentukan lokasi yang layak dilakukan penaybungan sumber air untuk pelanggan yang memiliki prospek baik dengan kemampuan untuk membayar tagihan air, sehingga dapat meminimalisir kerugian PDAM dan dapat memenuhi kebutuhan masyarakat. Penelitian ini juga diharapkan dapat menjadi referensi bagi peneliti selanjutnya yang berkaitan dengan pengguna algoritma yang digunakan.

Keyword: *Datamining, Naive Bayes, RapidMiner, Kemampuan Pelanggan, PDAM Tirta Lihou*

1. Pendahuluan

Bagi manusia, air memiliki peranan yang sangat besar bukan hanya untuk kebutuhan biologisnya, yaitu bertahan hidup. “Air diperlukan manusia untuk keperluan industri dan lain sebagainya, sehingga tidak terpungkiri terkadang keterbatasan persediaan air untuk pemenuhan kebutuhan menjadi pemicu timbulnya konflik sosial di masyarakat”. Perusahaan PDAM Tirta Lihou merupakan salah satu perusahaan air minum yang terletak di Kabupaten Simalungun. PDAM Tirta Lihou terdiri dari 33 unit produksi. Unit produksi dibentuk untuk meningkatkan mutu pelayanan kepada masyarakat guna memenuhi kebutuhan masyarakat. Dari 33 unit produksi yang terletak di beberapa wilayah masing-masing memiliki satu atau dua lokasi sumber air yang dapat disalurkan pada masyarakat. Di tahun 2014 jumlah sambungan yang aktif dari seluruh unit/wilayah yang dapat dilayani sebesar 40.032. Namun dengan meningkatnya jumlah pertumbuhan penduduk setiap tahunnya membuat masyarakat khususnya di Kabupaten Simalungun sering mengalami kekurangan air bersih, ditambah dengan ketersediaan air di setiap lokasi untuk masing-masing unit produksi semakin lama semakin berkurang. Hal ini dibuktikan dengan meningkatnya jumlah masyarakat yang bergabung dalam program MBR (Masyarakat Berpenghasilan Rendah) yang meningkat dari setiap tahunnya, seperti pada salah satu unit produksi terbesar yaitu unit siantar kota yang mengalami peningkatan dalam jumlah sambungan baru.

Tetapi ada masalah baru muncul setelah meningkatnya jumlah kebutuhan sandang pangan untuk kehidupan sehari-hari dan semakin tingginya inflasi membuat daya beli masyarakat berkurang bahkan kemampuan untuk membayar Air PDAM Tirta Lihou sering kali terlambat dan ada yang memiliki tunggakan yang begitu besar, bahkan ada yang sama sekali tak sanggup lagi membayar hingga pihak PDAM dengan terpaksa sambungan air tersebut. Ini yang membuat masyarakat dalam membuat sambungan air kembali perlu di tinjau ulang dalam kemampuan membayar, agar pihak PDAM Tirta Lihou juga mengalami kerugian yang cukup besar karena biaya operasional yang juga cukup besar. Dengan meningkatnya jumlah MBR yang masuk setiap tahunnya dimasing-masing wilayah, pihak PDAM Tirta Lihou berencana mencari alternatif solusi dalam membuat sambungan baru kepada masyarakat dengan mengklasifikasi kemampuan membayar masyarakat, sehingga dapat

meminimlasi kerugian PDAM. Dalam menentukan kemampuan pelanggan dalam membayar atau tidak pada unit produksi sektor siantar kota digunakan metode *datamining*.

“Tugas utama pada *datamining* diantaranya yaitu klasifikasi yang merupakan suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan”[1]. Salah satu algoritma dalam klasifikasi yaitu algoritma *Naive Bayes*. Dalam hal ini penulis menggunakan algoritma *Naive Bayes* dalam melakukan prediksi terhadap kelayakan lokasi pembangunan sumber air bersih berdasarkan data yang diperoleh langsung dari pihak PDAM Tirta Lihou dengan tujuan untuk menentukan lokasi pembangunan sumber air bersih pada PDAM Tirta Lihou.

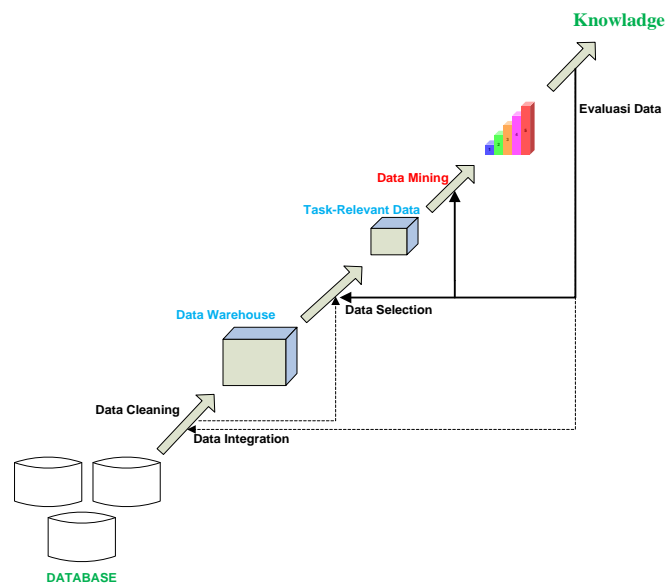
“*Naive Bayes* adalah salah satu metode klasifikasi yang berakar pada Teorema Bayes. Teorema Bayes dikombinasikan dengan “*Naive*” yang berarti setiap atribut/variabel bersifat bebas” (Brilian Argario, Hidayat and Kartika Dewi, 2018). “*Naive Bayes* memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan pada database dengan data yang besar”[2].

Dalam beberapa penelitian yang telah dilakukan oleh peneliti terkait metode *datamining* menggunakan algoritma *naive bayes*: [3] menjelaskan tentang algoritma *Naive Bayes* yang digunakan dalam Dalam Klasifikasi Lokasi Pembangunan Sumber Air memanfaatkan data training untuk menghasilkan probabilitas setiap kriteria untuk class yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk memprediksi kelayakan lokasi pembangunan sumber air bersih di PDAM Tirta Lihou berdasarkan proses klasifikasi yang dilakukan oleh metode *Naive Bayes* itu sendiri [4].

2. Metode Penelitian

Pengertian Data Mining

Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*mechine learning*) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis [5], [2]. Defenisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobervasi contoh-conroh spesifik dari konsep-konsep yang dipelajari. *Knowledge Discovery in Database* (KDD) adalah penerapan metode saintifik pada *data mining* [6].



Gambar 1. Proses KDD

Tahapan proses dalam penggunaan data mining yang merupakan proses *Knowledge Discovery in Database* (KDD) seperti yang terlihat pada gambar 2.1 dapat diuraikan sebagai berikut [7]:

1. Memahami domain aplikasi untuk mengetahui dan menggali pengetahuan awal serta apa sasaran pengguna
2. Membuat target data-set yang meliputi pemilihan data dan fokus pada sub-set data.
3. Pembersihan dan transformasi data meliputi eliminasi derau, *outliner*, *missing value* serta pemilihan fitur dan reduksi dimensi
4. Penggunaan algoritma *data mining* yang terdiri dari asosiasi, sekuensial, klasifikasi, klusterisasi, dll.
5. Interpretasi, evaluasi, dan visualisasi pola untuk melihat apakah ada sesuatu yang baru dan menarik dan dilakukan iterasi jika diperlukan.

Data Mining bertujuan untuk memanfaatkan data dalam database dengan mengolah data tersebut sehingga menghasilkan informasi baru yang berguna untuk kepentingan perusahaan atau instansi.

Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Model itu sendiri bisa berupa aturan “jika-maka”, berupa pohon keputusan, atau formula matematis[8].

Algoritma Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan[9]. Naive Bayes yang dikemukakan oleh ilmuwan inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian [10]. Untuk menyelesaikan metode Naive Bayes dapat dilakukan dengan persamaan-persamaan sebagai berikut:

$$P(H|X) = \frac{P(H|X) * P(H)}{P(X)}$$

Dimana:

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X): Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H): Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

Penjabaran lebih lanjut rumus Naive Bayes tersebut dilakukan dengan menjabarkan secara terperinci $(C|X_1, \dots, X_n)$ menggunakan aturan perkalian sebagai berikut.

$$\begin{aligned} P(C|x_1, \dots, x_n) &= P(C) P(x_1, \dots, x_n|C) \\ &= P(C) P(x_1|C) P(x_2, \dots, x_n|C, x_1) \\ &= P(C) P(x_1|C) P(x_2|C, x_1) P(x_3, \dots, x_n|C, x_1, x_2) P(x_1|C) \\ &= P(x_1|C) P(x_2|C, x_1) P(x_3|C, x_1, x_2) P(x_4, \dots, x_n|C, x_1, x_2, x_3) P(C) \\ &= P(x_1|C) P(x_2|C, x_1) P(x_3|C, x_1, x_2) \dots \\ &= P(x_n|C, x_1, x_2, x_3, \dots, x_{n-1}) \dots \end{aligned}$$

Dapat dilihat bahwa jika semakin banyak faktor-faktor yang semakin kompleks yang berpengaruh terhadap nilai probabilitas, maka semakin tidak mungkin untuk menghitung nilai tersebut satu persatu. Akibatnya proses perhitungan akan semakin susah untuk dilakukan, maka disinilah digunakan asumsi independensi yang sangat tinggi, bahwa masing-masing atribut dapat saling bebas. Dengan asumsi tersebut, diperlukan persamaan berikut:

$$P(X_i|X_j) = \frac{P(X_i)P(X_j)}{P(X_j)} = \frac{P(X_i \cap X_j)}{P(X_j)} = P(X_i)$$

Untuk $i \neq j$, sehingga

$$P(X_i|C, X_j) = P(X_i|C)$$

Dari persamaan (4) tersebut dapat di ambil kesimpulan bahwa asumsi independensi membuat syarat perhitungan menjadi lebih sederhana. Selanjutnya penjabaran $(P(C|X_1, \dots, X_n))$ dapat disederhanakan menjadi persamaan berikut

$$P(X_2|C) P(X_3|C) \dots P(C|X_1, \dots, X_n) = P(X_1|C) = \prod_i^n = 1 P(X_i|C) \quad (5)$$

Keterangan :

$$\prod_i^n = 1 P(X_i|C) = \text{perkalian ranting antar atribut.}$$

Persamaan (5) merupakan teorema bayes yang kemudian akan digunakan untuk melakukan perhitungan klasifikasi. Untuk klasifikasi dengan data continue atau data angka menggunakan rumus distribusi Gaussian dengan 2 parameter : mean μ dan varian σ :

$$p(X_i|C = c_j) = \sqrt{2\pi\sigma_{ij}} \exp \frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}}$$

Dimana :

P : Peluang

X_i : Atribut ke i

X_j : Nilai atribut ke i

C : Kelas yang dicari

C_i : Sub kelas Y yang dicari

μ : Menyatakan rata-rata dari seluruh atribut

σ : Deviasi standar, menyatakan varian dari seluruh atribut.

Dalam metode naive bayes diperlukan data latih dan data uji yang ingin diklasifikasikan, dalam naive bayes, semakin banyak data latih yang yang dilibatkan, semakin baik hasil yang prediksi yang diberikan. Menghitung P(C_i) yang merupakan probabilitas prior untuk setiap sub kelas C yang akan dihasilkan menggunakan persamaan:

$$P(c_i) = \frac{S_i}{S}$$

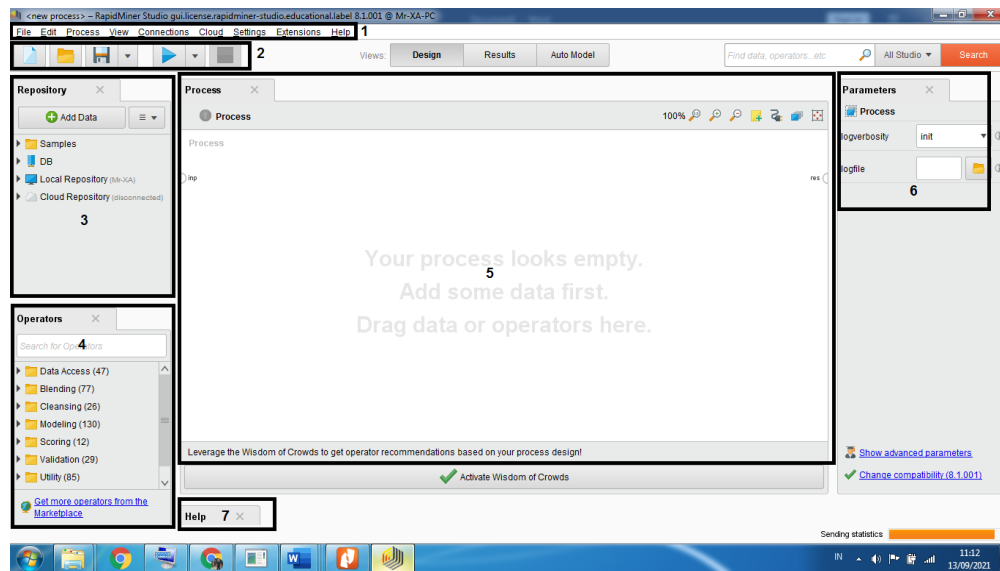
Dimana S_i adalah jumlah data training dari kategori C_i, dan S adalah jumlah total data training. Menghitung P(X_i|C_i) yang merupakan probabilitas posterior X_i dengan syarat C menggunakan persamaan diatas.

a. Rapid Miner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (Open Source), sebuah salah satu solusi untuk melakukan analisis terhadap *Data Mining*. *RapidMiner* adalah platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan bernama sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambahan teks, dan analisis prediktif [11], [12].

Hal ini digunakan untuk bisnis dan komersial, juga untuk penelitian, pendidikan, pelatihan, *rapid prototyping*, dan pengembangan aplikasi serta mendukung semua langkah dalam proses pembelajaran mesin termasuk persiapan data, hasil visualisasi, validasi model dan optimasi.

Penelitian yang dilakukan penulis menggunakan tool *RapidMiner Studio* versi 8.1, tampilan *RapidMiner Studio* versi 8.1 terdapat pada gambar 2.6 :



Gambar 2. Tampilan *RapidMiner Studio* versi 8.1

Keterangan pada Gambar 2 adalah :

1. *Menu Bar*, merupakan menu utama yang terletak dibawah Title Bar yang berfungsi sebagai menu perintah untuk mengoperasikan tool *RapidMiner Studio* versi 8.1.
2. *Repository View*, merupakan komponen penting dalam Design Perspective untuk mengolah dan menata proses serta dapat digunakan sebagai sumber data.
3. *Operator View*, merupakan komponen yang digunakan untuk operator dalam *RapidMiner* yang terdiri dari Data Access, Blending, Cleansing, dan lainnya untuk dalam hal proses analisis.
4. *Process View*, merupakan langkah-langkah dalam proses analisis, sebagai penghubung langkah-langkah proses dan sebagai media desain proses langkah-langkah analisis.
5. *Parameters View*, merupakan komponen dalam hal menentukan parameter dari operator yang digunakan, beberapa operator membutuhkan satu atau lebih parameter agar dapat diindikasikan sebagai fungsionalitas yang baik dan benar.
6. *Help dan Comment View*, merupakan komponen yang menunjukkan penjelasan setiap operator yang digunakan.
7. *Toolbox*, merupakan komponen yang digunakan untuk proses penyimpanan, membuat proses baru, proses eksekusi analisis, dan untuk mencari data, operator dan membuka proses yang sudah ada.

3. Metode Penelitian

Metode penelitian merupakan suatu teknik atau mencari, memperoleh, mengumpulkan data, baik berupa data primer maupun data sekunder yang digunakan untuk keperluan menyusun suatu karya ilmiah dan menganalisa faktor- faktor yang berhubungan dengan pokok-pokok permasalahan sehingga akan terdapat suatu kebenaran data-data yang akan diperoleh. Adapun metode penyelesaian yang digunakan dalam penelitian ini yaitu metode *data mining* dengan algoritma *naive bayes*. Penelitian ini dilakukan untuk mengimplementasikan algoritma *naive bayes* dalam klasifikasi kemampuan pelanggan membayar tagihan air di PDAM Kota Pematangsiantar. Dari hasil klasifikasi data Kemampuan Pelanggan Dalam Membayar PDAM Pada Siantar Kota dapat mengetahui informasi sehingga dapat dilakukan penanganan terhadap data Kemampuan Pelanggan Dalam Membayar PDAM.

a) Metode Pengumpulan Data

Dalam melakukan penelitian untuk mendapatkan data dan informasi, maka metode yang digunakan dalam proses pengumpulan data sebagai berikut:

1. Metode Observasi, melihat serta mempelajari permasalahan yang ada di lapangan yang berkaitan dengan

objek yang diteliti, yaitu informasi mengenai data kemampuan pelanggan dalam membayar PDAM untuk penanggulangan kerugian PDAM Kota Pemtangsiantar karena tidak balancenya biaya operasional dengan pemasukan iuran dari PDAM.

2. Metode Kajian Literatur, mencari bahan yang mendukung dalam pendefinisian masalah melalui buku-buku, makalah, *internet*, yang erat kaitannya dengan objek permasalahan.

b) Analisis Data

Analisis data yang digunakan penelitian ini menggunakan data kuantitatif dengan teknik analisis data yang menggunakan jenis statistik *deskriptif*. Statistik *deskriptif* ialah menganalisis data dengan cara menggambarkan atau mendeskripsikan data yang telah terkumpul tanpa bermaksud membuat kesimpulan yang berlaku untuk umum atau generalisasi. Data yang diperoleh kemudian diolah dengan *RapidMiner* menggunakan *Performance* yang berfungsi sebagai validasi dan reabilitas data untuk mencari keakuratan data.

c) Analisis dan Perancangan Sistem

Di dalam analisis dan metode perancangan sistem terdapat rancangan yang digunakan dalam membangun sistem atau perhitungan-perhitungan algoritma dari metode *Naive Bayes*.

d) Perancangan Penelitian

Rancangan atau model penelitian disajikan dalam rancangan *Flowchart* pada gambar 3.1



Gambar 3. Rancangan Penelitian

Gambar 3. menjelaskan rancangan penelitian yang dilakukan untuk mencari daerah dengan hasil produksi sawit yang belum optimal dengan Algoritma K-Means yang terdiri dari :

1. Analisis Masalah

Masalah yang terkait dengan minimnya hasil produksi sawit yang di panen di wilayah sumatera utara maka penulis berusaha mencari penyebab dengan memetakan daerah yang belum optimal.

2. Mempelajari Literatur

Penelitian ini harus didasari rujukan yang digunakan untuk mendapatkan rujukan yang digunakan untuk mendapatkan informasi dalam penelitian.

3. Analisa

Proses yang dilakukan untuk mencari daerah dengan hasil produksi sawit yang belum optimal dengan pemetaan.

4. Implementasi

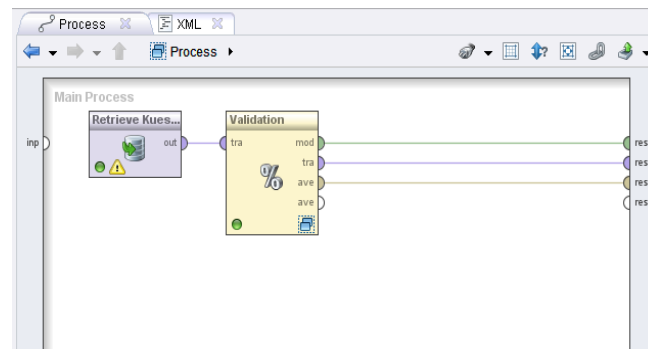
Implementasi dilakukan dengan menggunakan aplikasi *RapidMiner versi 8.1* sebagai sistem yang mencari keputusan dalam analisa daerah dengan hasil produksi sawit yang belum optimal dengan pemetaan.

5. Keputusan

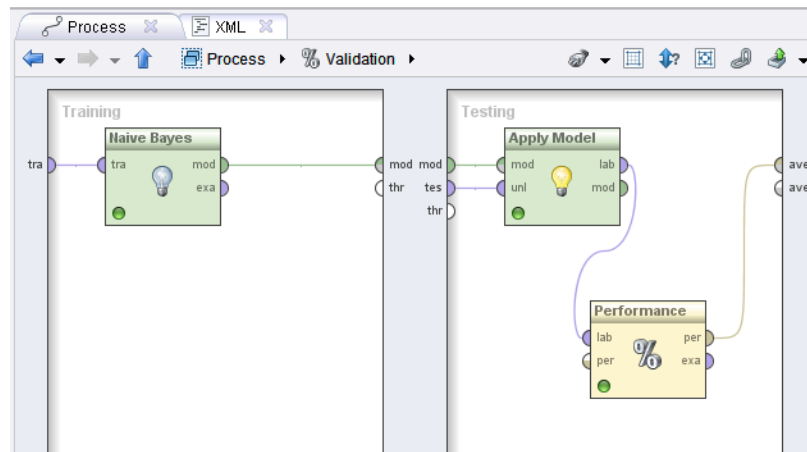
Hasil yang diberikan oleh sistem akan ditindak lanjuti oleh pihak Dinas Perkebunan Provinsi Sumatera Utara. Sehingga perlu mendapat perhatian dan penanganan efektif karena berkaitan dengan pengambilan kebijakan penyaluran bantuan yang dilakukan.

e) Pemodelan Rapidminer

Pengujian menggunakan algoritma *Naive Bayes* dengan menggunakan *RapidMiner* dapat dilihat pada Gambar 3.3 dan 3.4 :



Gambar 4. Model di RapidMiner Dengan Menggunakan Validasi



Gambar 5. Desain Model Algoritma *Naïve Bayes* dan Validasi Data

Gambar 4. dan 5. dapat dijelaskan permodelan yang digunakan menggunakan *X-Validation* yang digunakan untuk membagi dua area yaitu *training* dan *testing*. Area *Training* digunakan *Naïve Bayes* yang berfungsi untuk Klasifikasi data yang diolah untuk mencari kesimpulan. Area *testing* menggunakan *Apply Model* untuk hasil dari algoritma *Naïve Bayes* dan *Performance* untuk mendapatkan akurasi data yang digunakan.

4. Hasil Penelitian

a. Hasil

Pada penelitian ini menggunakan *tools rapidminer* sebagai pengujian data yang telah di analisis. Berikut adalah bagian-bagian dari *tools rapidminer* beserta isi dari data yang dibuat.

b. Data Set Penelitian

Untuk proses mengelompokkan data penulis menggunakan rumus pada metode *Naive Bayes*. Kemudian data Kemampuan Pelanggan Dalam Membayar PDAM Pematangsiantar tersebut akan diproses menggunakan aplikasi *Rapidminer 8.1* untuk mendapatkan hasil yang akurat. Data yang akan diolah adalah 80 Sampel Data yang ada di PDAM Kota Pematangsiantar sebagai data *Training* dan data *Testing* di ambil dari 20 data terakhir dari data *training*.

Tabel 1. Data *Testing*

No	Nama Pelanggan	Pekerjaan	Penghasilan	Jumlah anggota rumah tangga	Pengeluaran Keluarga	Status Pemakaian Air	Status Kemampuan Membayar
1	P51	Pedagang	Penghasilan ≥ 2500000 dan ≤ 4500000	3	Sedang	Sedang	Mampu
2	P52	Pedagang	Penghasilan ≥ 2500000 dan ≤ 4500000	4	Rendah	Sedang	Tidak Mampu
3	P53	Peternak	Penghasilan > 1500000	4	Sedang	Besar	Tidak Mampu

N o	Nama Pelanggan	Pekerjaan	Penghasilan	Jumlah anggota rumah tangga	Pengeluara n Keluarga	Status Pemakaia n Air	Status Kemampua n Membayar
4	P54	PNS	dan <=3000000 Penghasilan >= 3000000 dan	5	Besar	Kecil	Tidak Mampu
5	P55	Karyawan Swasta	<=5000000 Penghasilan >=2500000 dan <=4000000	4	Besar	Kecil	Tidak Mampu
.
.
30	P80	Pedagang	Penghasilan >= 2500000 dan <=4500000	4	Besar	Sedang	Tidak Mampu

Hasil belum diketahui klasifikasinya karena variabel tersebut merupakan hasil prediksi dari data yang dihitung dari setiap variabel. Berikut adalah proses perhitungannya :

Nama Anak (Status Kemampuan) = Penghasilan * Pengeluaran keluarga * Kualitas Air * Status Debit Air

$$P1 (\text{Mampu}) = \text{Penghasilan} \geq \text{Pedagang} * \text{Penghasilan} \geq 2500000 \text{ dan } \leq 4500000 * 3 * \text{Sedang} * \text{Sedang} \\ = 0,13 * 0,13 * 0,38 * 0,29 * 0,63 \\ = 0,0011$$

$$P1 (\text{Tidak Mampu}) = \text{Penghasilan} \geq \text{Pedagang} * \text{Penghasilan} \geq 2500000 \text{ dan } \leq 4500000 * 3 * \text{Sedang} * \text{Sedang} \\ = 0,2 * 0,2 * 0,29 * 0,38 * 0,68 \\ = 0,0029$$

.

.

.

$$P30 (\text{Mampu}) = 0,00007$$

$$P30 (\text{Tidak Mampu}) = 0,00257$$

c. Bandingkan Hasil Class Mampu dan Tidak Mampu

Dari hasil perhitungan tersebut, bandingkan nilai probabilitas antara kelas (P|mampu) dan kelas (P| tidak mampu) nilai probabilitas tertinggi dapat dilihat pada tabel berikut :

Tabel 2. Hasil Perbandingan

N o	Nama Pelanggan	Pekerjaan	Penghasilan	Jumlah anggota rumah tangga	Penge luara n Kelua rga	Status Pema kaian Air	Status Kemamp uan Membay ar	Class Predicti on	Ma mpu	Tidak mampu
1	P51	Pedagang	Penghasilan >= 2500000 dan <=4500000	3	Sedang	Sedang	Mampu	Tidak Mampu	0,0 01	0,003
			Penghasilan >= 2500000 dan <=4500000							
2	P52	Pedagang	Penghasilan >= 2500000 dan <=4500000	4	Rendah	Sedang	Tidak Mampu	Tidak Mampu	0,0 00	0,001
			Penghasilan >= 2500000 dan <=4500000							
3	P53	Peternak	Penghasilan >1500000 dan	4	Sedang	Besar	Tidak Mampu	Tidak Mampu	0,0 00	0,000

			<=300000 0							
	P54	PNS	Penghasila n >= 300000 dan <=500000 0	5	Besar	Kecil	Tidak Mampu	Mampu	0,0 00	0,000
4										
	P55	Karyaw an Swasta	Penghasila n >=250000 0 dan <=400000 0	4	Besar	Kecil	Tidak Mampu	Tidak Mampu	0,0 00	0,002
5										
.
.
.
	P80	Pedaga ng	Penghasila n >= 2500000 dan <=450000 0	4	Besar	Sedan g	Tidak Mampu	Tidak Mampu	0,0 00	0,003
3										
0										

Tabel 3. Confusion Table

Prediksi	Class	
	Mampu	Tidak mampu
Mampu	4	2
Tidak mampu	7	17

Hitungan Akurasi = $\frac{\text{Prediksi (Mampu,Class Mampu)} + \text{Prediksi (Tidak Mampu,Class Tidak Mampu)}}{\text{Jumlah Total Prediksi (Mampu,Class Mampu) dan Prediksi (Tidak Mampu,Class Tidak Mampu)}}$
 $= \frac{(4 + 17)}{30}$
 $= 0,70$ atau 70%

Dari hasil perhitungan tersebut, bandingkan nilai probabilitas antara kelas (P|Mampu) dan kelas (P|Tidak Mampu) nilai probabilitas tertinggi dapat disimpulkan bahwa Status Kemampuan Pelanggan tersebut memiliki Akurasi 70%.

d. Pembahasan

Berdasarkan dari penjelasan diatas mengenai tahap-tahap penggunaan serta hasil yang telah ditampilkan maka selanjutnya membahas mengenai keterkaitan dari hasil yang didapat antara perhitungan manual algoritma dengan hasil yang ditampilkan oleh *tools rapidminer*.

e. Pembuktian Dengan Menggunakan Tools Rapid Miner

Nilai probabilitas di atas akan diuji dengan data sebanyak 30 data dan diselesaikan dengan menggunakan *tools rapidminer* sehingga diperoleh hasil klasifikasi seperti pada gambar berikut :

Row No.	Status Kem...	prediction(S...	confidence(...	confidence(...	Pekerjaan	Penghasilan	Jumlah ang...	Pengetuara...	Stat
1	Mampu	Tidak Mampu	0.137	0.863	Pedagang	Penghasilan ...	3	Sedang	Sedi
2	Tidak Mampu	Tidak Mampu	0.149	0.851	Pedagang	Penghasilan ...	4	Rendah	Sedi
3	Tidak Mampu	Tidak Mampu	0.365	0.635	Peternak	Penghasilan ...	4	Sedang	Besi
4	Tidak Mampu	Mampu	0.716	0.284	PNS	Penghasilan ...	5	Besar	Keci
5	Tidak Mampu	Tidak Mampu	0.076	0.924	Karyawan Sw...	Penghasilan ...	4	Besar	Keci
6	Mampu	Mampu	0.584	0.416	PNS	Penghasilan ...	3	Besar	Sedi
7	Tidak Mampu	Tidak Mampu	0.070	0.930	Karyawan Sw...	Penghasilan ...	5	Besar	Sedi
8	Tidak Mampu	Tidak Mampu	0.130	0.870	Pedagang	Penghasilan ...	5	Besar	Besi
9	Tidak Mampu	Tidak Mampu	0.267	0.733	Petani	Penghasilan ...	4	Sedang	Besi
10	Mampu	Tidak Mampu	0.144	0.856	Karyawan Sw...	Penghasilan ...	6	Besar	Besi
11	Mampu	Tidak Mampu	0.099	0.901	Karyawan Sw...	Penghasilan ...	3	Rendah	Sedi
12	Mampu	Tidak Mampu	0.254	0.746	Petani	Penghasilan ...	3	Sedang	Sedi

Gambar 6. Hasil Perhitungan Dengan Menggunakan Tools RapidMiner

f. Validasi Data

Dalam melakukan validasi data terdapat hal-hal yang harus diperhatikan diantaranya: perhitungan manual algoritma harus telah menampilkan hasil akhir berupa pohon keputusan, serta data yang digunakan haruslah data yang *valid* dan sama dengan yang dipakai pada *tools*. Hasil pengujian Model Algoritma Naive Bayes Classifier ditunjukkan pada gambar berikut:

	true Mampu	true Tidak Mampu	class precision
pred. Mampu	4	2	66.67%
pred. Tidak Mampu	7	17	70.83%
class recall	36.36%	89.47%	

Gambar 7. Nilai Accuracy Performance

Keterangan :

1. Jumlah prediksi Mampu dan kenyataannya Mampu adalah 2 *record*.
2. Jumlah prediksi Mampu dan kenyataannya Tidak Mampu adalah 4 *record*.
3. Jumlah prediksi Tidak Mampu dan kenyataannya Mampu adalah 3 *record*.
4. Jumlah prediksi Tidak Mampu dan kenyataannya Tidak Mampu adalah 11 *record*.

5. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan, maka penulis mengambil beberapa kesimpulan:

1. Penerapan *Datamining* dengan menggunakan algoritma *naive bayes* pada klasifikasi kemampuan pelanggan dalam membayar tagihan air pada PDAM Tirta Lihou dapat diterapkan. Sumber data yang digunakan pada penelitian ini adalah data yang diperoleh langsung dari perusahaan. Jumlah data yang *Data Testing* sebanyak 30 data pelanggan dengan menggunakan dua kelas. Dari hasil perhitungan Algoritma *Naive Bayes* diperoleh klasifikasi dengan kelas mampu sebanyak 11 data dan kelas Tidak mampu sebanyak 19 data dan hasil dari pengujian menggunakan *tools rapidminer* diperoleh 11 kelas mampu dan 19 kelas tidak mampu.
2. Pengujian data pada *Rapidminer 8.1* menggunakan naive bayes dapat menampilkan dua kelas dari hasil klasifikasi dengan tingkat akurasi sebesar 70%.

6. Daftar Pustaka

- [1] Y. I. Kurniawan, "Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 455, 2018, doi: 10.25126/jtiik.201854803.
- [2] F. Handayani and S. Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110," *J. Tek. Elektro*, vol. 7, no. 1, pp. 19–24, 2015, doi: 10.15294/jte.v7i1.8585.
- [3] T. Imandasari, E. Irawan, A. P. Windarto, and A. Wanto, "Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 750, 2019, doi: 10.30645/senaris.v1i0.81.
- [4] D. Sartika and D. Indra, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *J. Tek. Inform. Dan Sist. Inf.*, vol. 1, no. 2, pp. 151–161, 2017.
- [5] N. Y. Septian, "Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan

- Mahasiswa Universitas Dian Nuswantoro,” *J. Semant.* 2013, pp. 1–11, 2009.
- [6] M. H. Rifqo and A. Wijaya, “Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit,” *Pseudocode*, vol. 4, no. 2, pp. 120–128, 2017, doi: 10.33369/pseudocode.4.2.120-128.
- [7] A. A. Rahman and Y. I. Kurniawan, “Aplikasi Klasifikasi Penerima Kartu Indonesia Sehat Menggunakan,” *Progr. Stud. Inform. Univ. Muhammadiyah Surakarta*, 2016.
- [8] A. Jananto, “Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa,” *Teknol. Inf. Din.*, vol. 18, no. 1, pp. 9–16, 2013.
- [9] H. Abnnur, “Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes,” *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [10] C. Fadlan, S. Ningsih, and A. P. Windarto, “Penerapan Metode Naive Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra,” *J. Tek. Inform. Musirawas*, vol. 3, no. 1, p. 1, 2018, doi: 10.32767/jutim.v3i1.286.
- [11] P. Alkhairi, I. S. Damanik, and A. P. Windarto, “Penerapan Jaringan Saraf Tiruan untuk Mengukur Korelasi Beban Kerja Dosen Terhadap Peningkatan Jumlah Publikasi,” *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 581, 2019, doi: 10.30645/senaris.v1i0.65.
- [12] A. D. I. Suradi, “Penerapan data mining untuk menentukan rekomendasi beasiswa dengan metode algoritma c4.5,” 2018.