Prosiding Seminar Nasional Riset Information Science (SENARIS)

ISSN: 2686-0260
September 2019
Hal:750-761

Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air

Tia Imandasari[1], Eka Irawan[2], Agus Perdana Windarto[3], Anjar Wanto[4]
STIKOM Tunas Bangsa Pematangsiantar
Jln. Jendral Sudirman Blok A No. 1,2,3 Pematangsiantar
tiaimandasari@gmail.com

Abstract-*The purpose of this study can predict the feasibility of the location of the development of clean water sources in the Tirta Lihou PDAM using the Naive Bayes algorithm. With the increasing number of MBR (Low-Income Communities) that enter each year in each region, the Tirta Lihou PDAM plans to find alternative springs solutions for several production units so that they can meet the needs of the community. In determining the appropriate alternative sources of clean water in several production units, the datamining method is used. By using data mining techniques specifically classification using the Naive Bayes algorithm, predictions can be made on the feasibility of the location of the construction of clean water sources based on existing data. Naive bayes is a simple probabilistic prediction technique based on the Bayes theorem with a strong assumption of independence (dependence). Based on the results of calculations using algoritma naive bayes, the clasification results from 19 alternatives used, where there are 8 feasible classes and 11 classes are not feasible with the number of accuracy obtained at 78,95%. From the results obtained, it is expected that this research can help the PDAM Tirta Lihou in determining the location that is feasible to develop water sources so that it can meet the needs of the community. This research is also expected to be a reference for further researchers relating to the user algorithm used.*

Keywords: *Data Mining, Naive Bayes, RapidMiner, Location of Water Source, PDAM Tirta Lihou*

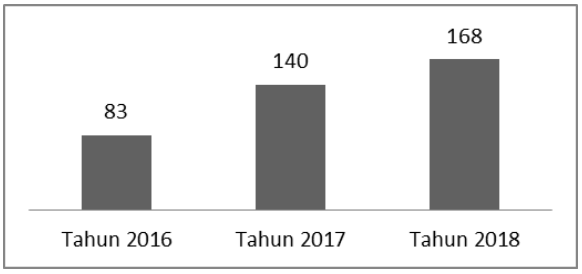
Abstrak*- Tujuan dari penelitian ini dapat memprediksi kelayakan lokasi pembangunan sumber air bersih di PDAM Tirta Lihou menggunakan algoritma Naive Bayes. Dengan meningkatnya jumlah MBR(Masyarakat Berpenghasilan Rendah) yang masuk setiap tahunnya dimasing-masing wilayah, pihak PDAM Tirta Lihou berencana mencari alternatif solusi sumber mata air untuk beberapa unit produksi sehingga dapat memenuhi kebutuhan masyarakat. Dalam menentukan alternatif sumber air bersih yang layak pada beberapa unit produksi digunakan metode datamining. Dengan menggunakan teknik datamining khususnya klasifikasi menggunakan algoritma Naive Bayes dapat dilakukan prediksi terhadap kelayakan lokasi pembangunan sumber 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 19 alternatif yang digunakan, dimana terdapat 8 kelas Layak dan 11 Tidak Layak dengan total Accuracy yang diperoleh sebesar 78,95%. Dari hasil yang diperoleh,diharapkan penelitian ini dapat membantu pihak PDAM Tirta Lihou dalam menentukan lokasi yang layak dilakukan pembangunan sumber air sehingga dapat memenuhi kebutuhan masyarakat. Penelitian ini juga diharapkan dapat menjadi referensi bagi peneliti selanjutnya yang berkaitan dengan pengguna algoritma yang digunakan.*

Kata Kunci: *Datamining, Naive Bayes, RapidMiner, Lokasi Sumber Air, PDAM Tirta Lihou*

ISSN: 2686-0260
September 2019
Hal:750-761

1. PENDAHULUAN

"Datamining adalah salah satu cabang ilmu kecerdasan buatan dalam penggalian pola-pola untuk mengubah data menjadi informasi. *Datamining* merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran *computer* untuk menganalisa dan mengekstraksi pengetahuan *(knowladge)* secara otomatis" [1]. "*Datamining* mampu mengolah data dengan jumlah yang besar dan dapat melakukan pencarian data secara otomatis, oleh karena itu *datamining* memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri, keuangan, cuaca, ilmu dan teknologi" [2]–[6].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 konflk sosial di masyarakat" [7]. Perusahaan PDAM Tirta Lihou merupakan salah satu perusahaan air minum yang terletak di Kabupaten Simalungun. PDAM Tirta Lihou terdiri dari 33 unit produksi. Dari 33 unit produksi yang terletak dibeberapa wilayah masing-masing memiliki satu atau dua lokasi sumber air yang dapat disalurkan pada masyarakat. Ditahun 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 semangkin 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 sinaksak yang mengalami peningkatan dalam jumlah sambungan baru seperti yang tertera pada Gambar 1.



(Sumber: PDAM Tirta Lihou, 2018)

Gambar 1. Grafik Jumlah MBR Unit Sinaksak Pertahun

Berdasarkan Gambar 1.1 dapat dilihat bahwa jumlah MBR untuk setiap tahun semakin meningkat. Pada tahun 2016 jumlah MBR sebanyak 83 rumah tangga. Tahun 2017 jumlah MBR sebanyak 140 rumah tangga, dan ditahun 2018 jumlah MBR sudah mencapai 168 rumah tangga. Dengan meningkatnya jumlah MBR yang masuk setiap tahunnya dimasing-masing wilayah, pihak PDAM Tirta Lihou berencana mencari alternatif solusi sumber mata air untuk bebrapa unit produksi sehingga dapat memenuhi kebutuhan masyarakat. Dalam menentukan alternatif

sumber air bersih yang layak pada beberapa unit produksi 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" [8]. Salah satu algoritma dalam klasifikasi yaitu algoritma *Naive Bayes*. Dalam hal ini penulis mengguankan 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" [9]. "*Naive Bayes* memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan pada database dengan data yang besar" [10]. Dalam beberapa penelitian yang telah dilakukan oleh peneliti terkait metode *datamining* menggunakan algoritma *naive bayes*: [11] menjelaskan tentang algoritma *Naive Bayes* yang digunakan dalam memprediksi penggunaan listrik dengan hasil perhitungan algoritma *Naive Bayes* memanfaatkan data training untuk menghasilkan probabilitas setiap kriteria untuk class yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk memprediksi penggunaan listrik berdasarkan proses klasifikasi yang dilakukan oleh metode *Naive Bayes* itu sendiri. Berdasarkan latar belakang masalah di atas, penulis menggunakan Algoritma *Naive Bayes* Dalam Klasifikasi Lokasi Pembangunan Sumber Air. Diharapkan penelitian ini dapat membantu pihak PDAM Tirta Lihou dalam menentukan lokasi yang layak dilakukan pembangunan sumber air sehingga dapat memenuhi kebutuhan masyarakat. Penelitian ini juga diharapkan dapat menjadi referensi bagi peneliti selanjutnya yang berkaitan dengan pengguna algoritma yang digunakan.

2. METODOLOGI PENELITIAN

2.1 Analisa Data

Penelitian dilakukan di PDAM Tirta Lihou, Kab. Simalungun. Analisis data yang digunakan penelitian ini menggunakan data kuantitatif dengan teknik analisis data yang menggunakan jenis statistik *deskriptif* . data yang digunakan diperoleh langsung dari pihak PDAM Tirta Lihou dengan melakukan wawancara terlebih dahulu kepada pihak PDAM Tirta Lihou. Data yang diperoleh selanjutnya diolah menggunakan algoritma *naive bayes.* Kemudian diuji dengan *tools RapidMiner* menggunakan *Performance* yang berfungsi sebagai *validasi*. Pada penelitian ini data yang digunakan terdiri dari data *training* dan data *testing,* dimana data *testing* akan diolah menggunakan algoritma *naive bayes* agar dapat diperoleh hasil klasifikasi. Berikut adalah data testing yang digunakan:

ISSN: 2686-0260
September 2019
Hal:750-761

Tabel 1. Data *Testing*

No	Lokasi	Kapasitas (Liter/detik)			Kapasitas Produksi (M3)			Kapasitas Terpasang (M3)		Klasifikasi
		Sumber	Terpasang	Produksi	Per-Jam	Per-Hari	Per-Bulan	Per-Hari	Per-Bulan	
1	Saribu Jawa	55	27	24,5	88,2	1499,4	46841,4	232,8	72316,8	?
2	Saribu Asih	50	35	28	100,8	1713,6	53121,6	3024	93744	?
3	Mual Code	10	7,5	6	21,6	54	1674	648	20088	?
4	Pancur Napitu	15	8,8	8,8	31,68	380,16	11784,96	760,32	23569,92	?
5	Bah Timuran	70	7,5	7,5	27	491,4	15233,4	648	20,088	?
6	Sampan Taon	175	10	10	36	799,2	24775,2	864	26784	?
7	Rambung	40	10	10	36	608,4	18860,4	864	26784	?
8	Sinaksak	47	20	20	72	1368	42408	1728	53568	?
9	Bah Pamujian	130	10	10	36	637,56	19764,36	864	26784	?
10	Tambun Sari	18	7,5	7,5	27	330,75	10253,25	648	20088	?
11	Serbelawan	6	10	6,5	23,4	135,72	4207,32	691,2	26784	?
12	Marihat HT 2	15	9	7	25,2	428,4	13280,4	777,6	24105,6	?
13	Bah Gunung	28	12,5	8	28,8	489,6	15177	1080	33480	?
14	Bah Sibagari	45	18	15	54	918	28458	1555,2	48211,2	?
15	Sumur Bor	15	7,5	7,5	27	364	11288	501,2	30215	?
16	Bah Ram	14,52	8,3	8,3	29,88	188,24	15277	575,52	22145	?
17	Sirongit	9	7	5	18	306	9486	804	18200	?
18	Sinaman	10	5	2	7,2	122,4	3794,4	432	12396	?
19	Bah Bukare	6	4	1,5	5,4	91,8	2845,8	345,6	10785	?

2.2 Data Mining

Data mining adalah suatu metode untuk menemukan pengetahuan dalam suatu tumpukan data yang cukup besar dengan proses menggali dan menganalisa sejumlah untuk memperoleh sesuatu yang benar, baru dan bermanfaat sehingga dapat ditemukan suatu corak atau pola dalam data tersebut [12]. *Datamining* adalah salah satu cabang ilmu kecerdasan buatan dalam penggalian pola-pola untuk mengubah data menjadi informasi. *Datamining* merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran *computer* untuk menganalisa dan mengekstraksi pengetahuan *(knowladge)* secara otomatis [1]

2.3 Algoritma Naive Bayes

Naïve Bayes merupakan suatu algoritma yang dapat mengklasifikasikan suatu variable tertentu dengan menggunakan metode probabilitas dan statistik [13]. *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada data training [14]. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan *(Training Data)* yang kecil untuk menentukan estimasi paremeter yang diperlukan dalam proses pengklasifikasian dan dapat bekerja jauh lebih baik dalam situasi dunia nyata yang kompleks [11].

Berikut adalah Alur penyelesaian dari metode *Naïve Bayes* [15]:

1. Baca data training
2. Hitung Jumlah dan probabilitas
   a. Jika terdapat data numerik, maka temukan nilai *mean* dan standar deviasi dari masing-masing parameter yang menggambarkan data angka. Berikut adalah

Rumus yang digunakan untuk menghitung nilai rata – rata (*mean*) dapat dilihat sebagai berikut :

$$\mu = \sum_{i=1}^{n} x_i \text{ atau } \mu = \frac{x_1+x_2+x_3+\cdots+x_n}{n} \quad (1)$$

Keterangan :
μ : rata – rata hitung (mean)
xi : nilai sample ke –i
n : jumlah sampel

Dan berikut rumus yang digunakan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat dibawah ini:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}} \quad (2)$$

Keterangan:
σ : standar deviasi
xi : nilai x ke –i
μ : rata-rata hitung
n : jumlah sampel

b. Jika data tidak numerik, hitung nilai probabilitas tiap kategori yang sama, dengan jumlah data dari kategori yang sama lalu dibagi dengan data pada kategori tersebut.

3. Nilai Probabilitas Fitur Setiap Kelas
Untuk mengetahui nilai probabilitas setiap fitur pada kelas kita harus hitung jumlah data yang sesuai dari kategori yang sama lalu dibagi dengan jumlah data pada kategori tersebut.

4. Nilai Distribusi *Gaussian*
Langkah selanjutnya adalah menghitung nilai probabilitas untuk fitur Data *Testing* yang mempunyai data numerik / angka. Berikut adalah persamaan dalam mencari nilai distribusi *gaussian.*

$$P = (X_i = x_i \mid Y = y_j) = \frac{1}{\sqrt{2\pi\sigma ij}} \times e^{-\frac{(x_i-\mu ij)^2}{2\sigma^2 ij}} \quad (3)$$

5. Probabilitas Akhir Setiap Kelas
Menghitung probabilitas akhir untuk setiap kelas artinya memasukkan semua data nilai distribusi *gaussian* yang ada ke dalam satu kelas yang sama.

P(X|Kelas)=P(V1|Kelas)×P(V2|Kelas)×P(V3|Kelas)×P(V4|Kelas)×P(V5|Kelas)×P(V6|Kelas)×P(V7|Kelas)×P(V8|Kelas)

ISSN: 2686-0260
September 2019
Hal:750-761

6. Probabilitas Akhir

Probabilitas akhir didapat melalui perhitungan nilai probabilitas akhir kelas ke dalam rumus *Naïve Bayes Classifier*. Perhitungan probabilitas akhir adalah sebagai berikut.

$$P(Kelas|X) = P(Kelas) * P(X) \quad (4)$$

Setelah mendapatkan probabilitas akhir, langkah terakhir dilakukan normalisasi dengan cara membagikan nilai probabilitas satu kategori dengan jumlah nilai semua kategori.

$$P(Kelas)= P(Kelas|X) / (P(X|Kelas)+ P(X|Kelas)) \quad (5)$$

3. HASIL DAN PEMBAHASAN

3.1 Analisa Algoritma Naive Bayes

Dalam penelitian ini digunakan data numerik. Berdasarkan data yang digunakan maka tahap pertama yang harus dilakukan yaitu menghitung nilai *mean* dan starndar deviasi pada data *training*. Berikut adalah data training yang digunakan:

Tabel 2. Data *Training*

No	Lokasi	Kapasitas (Liter/detik)			Kapasitas Produksi (M3)			Kapasitas Terpasang (M3)		Klasifikasi
		Sumber	Terpasang	Produksi	Per-Jam	Per-Hari	Per-Bulan	Per-Hari	Per-Bulan	
1	Sitahuan	65	40	31	111,6	1897,6	58813,2	3456	107336	Layak
2	Bah Timuran	70	7,5	7,5	27	491,4	15233,4	648	20,088	T. Layak
3	Bronbos	30	6,5	3,5	12,6	214,2	6640,2	691,2	17409,6	T. Layak
4	Sibunga-Bunga	20	8	7	25,2	428,4	13280,4	864	21427,2	Layak
5	Mual Jenggot	10	8	7	25,2	428,4	13280,4	691,2	21427,2	T. Layak
6	Sibual-Bual	25	10	6,5	23,4	397,8	12331,8	864	26784	T. Layak
7	Seribu Jawa	20	8	4	14,4	244,8	7588,8	691,2	21427,2	T. Layak
8	Marihat HT 1	20	10	8,5	30,6	520,2	16126,2	864	26784	T. Layak
9	Dusun	30	12	5	18	306	9486	1036,8	32140,8	Layak
10	Bah Bulian	20	11	6	21,6	367,2	11382,2	950,4	29462,4	T. Layak
11	Bah Sinurpi	10	8,5	5	18	306	9486	734,4	22766,4	T. Layak
12	Simarjarunjung	8	5	4	14	234	7078	565	20088,2	T. Layak
13	Silangit	10	7	5,5	19	254	15654	797	26784	T. Layak
14	Sinilon 1	13	10	9	22	256	17874	814	21427,2	T. Layak
15	Sinilon 2	15	12	7	25	438	22260	1054	33480	Layak
16	Aek Siboro	6	4	2,5	9	153	6545	290	20045	T. Layak
17	Bandar Saribu	5	3	1,5	5,4	93	2605	250	8077	T. Layak
18	Batu Saung	10	9	8	28	480	15379	557,2	33702	T. Layak
19	Pondok X	23	8	6,5	23,4	397	2873	382,4	27205	T. Layak
20	Dao	25	8	6,5	23,4	397	2873	382,4	27205	Layak
21	Bah Hisir	13	3,5	3,5	12,6	93,24	2893,44	366,4	8278	T. Layak
22	Dam Atok	8	7	6	21,6	367,2	11383,3	664	28704	T. Layak
23	Bah Bonaran	35	23	18	64,8	1101,2	36348	1505,25	22230	Layak
24	Tanduk Ni Huta	14	12	8	28,8	489,6	18972	1576	34451	Layak
25	Slmt Datang	15	13	10	36	612	16972	2523	54520	Layak
26	Sihole	46	43,5	40	144	2448	75888	3755	108522	Layak

3.1.1 Mencari Nilai *Mean* dan Standar Deviasi

Dalam penelitian ini digunakan dua kelas yaitu layak dan tidak layak. Sebelum mencari nilai *mean* dan standar deviasi, kita harus mengelompokkan terlebih dahulu jumlah kelas layak dan tidak layak berdasarkan data training yang digunakan.

Kelas Layak = 9

Kelas Tidak Layak = 17

a. Nilai *Mean*

μ(Layak) = (65+20+30+15+25+35+14+15+46)/9

μ(Layak) = (265)/9

μ(Layak) = 29,4444444

Data diatas diambil dari data pada variabel Sumber dengan hasil di kelas Layak, sehingga didapatkan nilai μ nya. Untuk perhitungan nilai μ dengan hasil Tidak Layak pada variabel Sumber langkah-langkah nya sama. Sedangkan untuk menghitung nilai μ pada variabel lainnya langkah-langkah yang digunakan juga sama seperti menghitung nilai *mean* pada variabel Sumber. Setelah mendapatkan nilai μ pada masingmasing variabel, maka selanjutnya kita mencari nilai standar deviasi, Untuk menghitung nilai standar deviasi klasifikasi Layak dan Tidak Layak.

b. Nilai Standar Deviasi

$$\sigma(\text{Layak})=\sqrt{\begin{array}{l}(((65-29{,}4444444)^2)+((20-29{,}4444444)^2)+\\ ((30-29{,}4444444)^2)+((15-29{,}4444444)^2)+\\ ((25-29{,}4444444)^2)+((35-29{,}4444444)^2)+\\ ((14-29{,}4444444)^2)+((15-29{,}4444444)^2)+\\ ((46-29{,}4444444)^2))/(9-1)\end{array}}$$

σ(Layak) = $\sqrt{2334{,}22222/8}$

σ(Layak) =$\sqrt{291{,}777778}$

σ(Layak) = 17,081504

Data diatas diambil dari data pada variabel Sumber dengan hasil dikelas Layak dan Tidak Layak, sehingga didapatkan nilai σ nya. Untuk perhitungan nilai σ dengan hasil Layak dan Tidak Layak pada varibel lainnya langkah-langkah nya sama.

Setelah menghitung nilai *mean* dan standar deviasi untuk data angka, maka selanjutnya menentukan nilai probabilitas setiap fitur pada setiap kelasnya. Untuk mengetahui nilai probabilitas setiap fitur pada kelas kita harus hitung jumlah data yang sesuai dari kategori. Jumlah data yang digunakan adalah sebanyak 26 data. Data Layak ada 9, data Tidak Layak ada 17. Sehingga untuk menghitung nilai probabilitas nya adalah sebagai berikut.

P(Layak)=9/26

P(Layak)=0,34615

P(Tidak Layak)=17/26

P(Tidak Layak)=0,65385

ISSN: 2686-0260
September 2019
Hal:750-761

Setelah mencari nilia probabilitas tiap fitur, maka didapat tabel untuk nilai, *mean*, standar deviasi dan probabilitas untuk masing-masing variabel, bisa dilihat pada tabel berikut.

Tabel 3. Nilai Probabilitas Kapasitas

Nilai	Kapasitas (Liter/detik)					
	Sumber		Terpasang		Produksi	
	Layak	T.Layak	Layak	T.Layak	Layak	T.Layak
μMean	29,4444444	17,7058824	19,05556	7,41176	14,72222	5,55882
σDeviasi^2	291,77778	234,47059	185,27778	5,69485	157,94444	4,62132
σDeviasi	17,08150	15,31243	13,61168	2,38639	12,56760	2,14973

3.1.2 Nilai Distribusi Gaussian

Langkah selanjutnya adalah menghitung nilai probabilitas untuk fitur Data *Testing*. Dikarenakan semua kelas menggunakan tipe data numerik, maka nilai probabilitas harus dihitung semua sesuai dengan Data *Testing* yang digunakan.

Tabel 4. Data *Testing*

No	Lokasi	Kapasitas (Liter/detik)			Kapasitas Produksi (M3)			Kapasitas Terpasang (M3)		Klasifikasi
		Sumber	Terpasang	Produksi	Per-Jam	Per-Hari	Per-Bulan	Per-Hari	Per-Bulan	
1	Saribu Jawa	55	27	24,5	88,2	1499,4	46841,4	232,8	72316,8	?
2	Saribu Asih	50	35	28	100,8	1713,6	53121,6	3024	93744	?
3	Mual Code	10	7,5	6	21,6	54	1674	648	20088	?
4	Pancur Napitu	15	8,8	8,8	31,68	380,16	11784,96	760,32	23569,92	?
5	Bah Timuran	70	7,5	7,5	27	491,4	15233,4	648	20,088	?
6	Sampan Taon	175	10	10	36	799,2	24775,2	864	26784	?
7	Rambung	40	10	10	36	608,4	18860,4	864	26784	?
8	Sinaksak	47	20	20	72	1368	42408	1728	53568	?
9	Bah Pamujian	130	10	10	36	637,56	19764,36	864	26784	?
10	Tambun Sari	18	7,5	7,5	27	330,75	10253,25	648	20088	?
11	Serbelawan	6	10	6,5	23,4	135,72	4207,32	691,2	26784	?
12	Marihat HT 2	15	9	7	25,2	428,4	13280,4	777,6	24105,6	?
13	Bah Gunung	28	12,5	8	28,8	489,6	15177	1080	33480	?
14	Bah Sibagari	45	18	15	54	918	28458	1555,2	48211,2	?
15	Sumur Bor	15	7,5	7,5	27	364	11288	501,2	30215	?
16	Bah Ram	14,52	8,3	8,3	29,88	188,24	15277	575,52	22145	?
17	Sirongit	9	7	5	18	306	9486	804	18200	?
18	Sinaman	10	5	2	7,2	122,4	3794,4	432	12396	?
19	Bah Bukare	6	4	1,5	5,4	91,8	2845,8	345,6	10785	?

Hasil belum diketahui klasifikasinya karena variabel tersebut merupakan hasil prediksi dari data yang dihitung dari setiap variabel. Untuk menghitung nilai distribusi gaussian, menggunakan persamaan (3). Berikut adalah proses menghitung nilai distribusi gaussian :

P(Sumber = 55 | Layak) = $\frac{1}{\sqrt{2\pi * 17,08150}} * e^{-\frac{(55-29,4444444)^2}{(2*291,77778)}}$

P(Sumber = 55 | Layak) = $\frac{1}{42,8061262963} * 2,718281^{-1,1191503766}$

P(Sumber = 55 | Layak) = $\frac{1}{42,8061262963} * 0,3265571272$

P(Sumber = 55 | Layak) = 0,0076287475

Penyelesaian diatas menggunakan satu sampel Data *Testing* yaitu alternatif Saribu Jawa dengan variabel Sumber. Untuk menghitung nilai distribusi gaussian dengan hasil Tidak Layak, langkah penyelesaian sama untuk menghitung nilai distribusi gaussian pada setiap variabel.

3.1.3 Probabilitas Akhir Setiap Kelas

Menghitung probabilitas akhir untuk setiap kelas yaitu memasukkan semua data nilai distribusi gaussian yang ada ke dalam satu kelas yang sama, seperti berikut:

P(X|Layak)=0,00762874745799175*0,0247251023136586*0,023459954863804*0,00651328940057145*0,000383832956781163*0,000012274513749396 6*0,00141337790054551*0,0000091734767902799

P(X|Layak)= 0,000000000000000000000000176059

Nilai diatas adalah probabilitas akhir untuk kelas dengan hasil Layak. Untuk mencari probabilitas kelas dengan hasil Tidak Layak, langkah-langkah nya sama seperti menghitung probabilitas akhir kelas dengan hasil Layak.

3.1.4 Probabilitas Akhir

Probabilitas akhir didapat melalui perhitungan nilai probabilitas akhir kelas. Perhitungan probabilitas akhir adalah sebagai berikut.

P(Layak|X) = 0,34615 * 0,000000000000000000000000176059

P(Layak|X) = 0,000000000000000000000000060943

Setelah mendapatkan probabilitas akhir, langkah terakhir dilakukan normalisasi, dengan cara sebagai berikut:

P(Layak)= P(Layak|X) / (P(X|Layak)+ P(X|TidakLayak))

P(Layak)=0,000000000000000000000000060943/(0,000000000000000000000000176059 + 0,000000000000000000000000000000))

P(Layak)= 2,8888888889

Tabel 5. Hasil Klasifikasi

No	Unit	Lokasi	Probabilitas Akhir		Keterangan
			Kelas Layak	Kelas Tidak Layak	
1	Totap Majawa	Saribu Jawa	2,88888888888889	0,00000000000000	Layak
2	Tanah Jawa	Saribu Asih	88,3003455896004	0,00000000000000	Layak
3	Huta Bayu	mual code	0,00000096686658	88,3003445462191	Tidak Layak
4	Raja Maligas	pancur napitu	0,00196581379776	81,8223504896037	Tidak Layak
5	Bangun	bah timuran	0,15219089523414	61,1912817133176	Tidak Layak
6	Kerasaan	sampan taon	0,00000487939654	717,077723609511	Tidak Layak
7	Perdagangan	Rambung	51,8115856270783	0,00000369779141	Layak

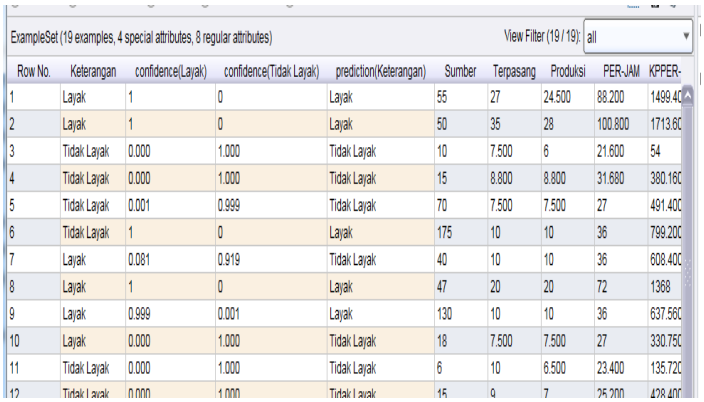
ISSN: 2686-0260
September 2019
Hal:750-761

8	Sinaksak	Sinaksak	72,589382133935	0,00000000000000	Layak
9	Karang Sari	bah pamujian	1434157066,04203	0,00723746126326	Layak
10	Karang Sari	Tambun Sari	53,5772655658045	0,33633965267819	Layak
11	Serbelawan	Serbelawan	0,00028013969988	53,5771288703454	Tidak Layak
12	Tiga Dolok	marihat HT 2	0,00051099716737	109,788683009909	Tidak Layak
13	Tiga Dolok	bah gunung	0,87444575109991	59,9584497854806	Tidak Layak
14	Raya Bayu	bah sibagari	64,8021045623453	0,00000000000000	Layak
15	Seribu Dolok	sumur bor	0,00028347091791	64,8018044278679	Tidak Layak
16	Sindar Raya	bah ram	0,00078104187024	61,2035091642921	Tidak Layak
17	Panombean Pane	sirongit	0,00003650031919	62,7007722753373	Tidak Layak
18	Bangun Pane	sinaman	81,8036945077325	0,02278986001042	Layak
19	Nagori Sinasih	bah bukare	0,03133600048842	81,8016242419948	Tidak Layak

3.2. Hasil

Berdasarkan hasil klasifikasi yang terdapat pada tabel 5, dapat dilihat bahwa dalam penentuan lokasi layak maupun tidak layak dapat dilihat dari nilai probabilitas akhir dari kelas layak dan tidak layak. Jika nilai probabilitas akhir layak lebih besar daripada kelas tidak layak maka hasil yang diperoleh berada pada kelas layak, dan jika nilai probabilitas akhir kelas tidak layak lebih besar dari kelas layak maka diperoleh hasil tidak layak.

Nilai probabilitas di atas akan diuji dengan data sebanyak 19 data dan diselesaikan dengan menggunakan *tools rapidminer* sehingga diperoleh hasil klasifikasi seperti pada gambar 2 berikut :


ExampleSet (19 examples, 4 special attributes, 8 regular attributes) View Filter (19 / 19): all

Row No.	Keterangan	confidence(Layak)	confidence(Tidak Layak)	prediction(Keterangan)	Sumber	Terpasang	Produksi	PER-JAM	KPPER-
1	Layak	1	0	Layak	55	27	24.500	88.200	1499.40
2	Layak	1	0	Layak	50	35	28	100.800	1713.60
3	Tidak Layak	0.000	1.000	Tidak Layak	10	7.500	6	21.600	54
4	Tidak Layak	0.000	1.000	Tidak Layak	15	8.800	8.800	31.680	380.160
5	Tidak Layak	0.001	0.999	Tidak Layak	70	7.500	7.500	27	491.400
6	Tidak Layak	1	0	Layak	175	10	10	36	799.200
7	Layak	0.081	0.919	Tidak Layak	40	10	10	36	608.400
8	Layak	1	0	Layak	47	20	20	72	1368
9	Layak	0.999	0.001	Layak	130	10	10	36	637.560
10	Layak	0.000	1.000	Tidak Layak	18	7.500	7.500	27	330.750
11	Tidak Layak	0.000	1.000	Tidak Layak	6	10	6.500	23.400	135.720
12	Tidak Layak	0.000	1.000	Tidak Layak	15	9	7	25.200	428.400

Gambar 2. Hasil Perhitungan Dengan Menggunakan *Tools RapidMiner*

Pada gambar 2 dapat dilihat bahwa hasil dari perhitungan manual adalah sama dengan hasil perhitungan menggunakan *tools rapidminer.*

3.3. 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*. 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 *Classfier* ditunjukan pada gambar berikut:

Table View Plot View

accuracy: 78.95%

	true Layak	true Tidak Layak	class precision
pred. Layak	5	1	83.33%
pred. Tidak Layak	3	10	76.92%
class recall	62.50%	90.91%	

Gambar 3. Nilai *Accuracy Performance*

Keterangan :

a. Jumlah prediksi Layak dan kenyataannya benar Layak adalah 5 *record*.
b. Jumlah prediksi Tidak Layak dan kenyataannya benar Layak adalah 3 *record*.
c. Jumlah prediksi Layak dan kenyataannya benar Tidak Layak adalah 1 *record*.
d. Jumlah prediksi Layak dan kenyataannya benar Tidak Layak adalah 10 *record*.

Pada gambar Nilai *Accuracy Performance* dijelaskan bahwa prediksi Layak memiliki nilai 5 dengan *class precision* 83,33%, prediksi Tidak Layak memiliki nilai 10 dengan *class precision* 76,92 %. Sehingga total *Accuracy* yang diperoleh sebesar 78,95%.

4. KESIMPULAN

Berdasarkan pembahasan sebelumnya dapat disimpulkan bahwa:

a. Penerapan *Datamining* dengan menggunakan algoritma naive bayes pada klasifikasi lokasi pembangunan sumber 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 19 Lokasi dengan menggunakan dua kelas. Dari hasil perhitunngan Algoritma *Naive Bayes* diperoleh klasifikasi dengan kelas Layak sebanyak 8 lokasi dan kelas Tidak Layak sebanyak 11 lokasi.
b. Pengujian data pada *Rapidminer 5.3* menggunakan naive bayes dapat menampilkan dua kelas dari hasil klasifikasi dengan tingkat akurasi sebesar 78,95%.

DAFTAR PUSTAKA

[1] M. G. Sadewo, A. P. Windarto, and D. Hartama, "Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means," *InfoTekJar (Jurnal Nasional Informatika dan Teknologi Jaringan)*, vol. 2, no. 1, pp. 60–67, 2017.

[2] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, "Memanfaatkan Algoritma K-Means dalam Menentukan Pegawai yang Layak Mengikuti Asessment Center untuk Clustering Program SDP," *CESS (Journal of Computer Engineering System and Science)*, vol. 3, no. 1, pp. 87–93, 2018.

[3] M. G. Sadewo, A. P. Windarto, and A. Wanto, "Penerapan Algoritma Clustering dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi dengan K-Means," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 2, no. 1, pp. 311–319, 2018.

[4] R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi Rapidminer dengan Metode K-Means (Study Kasus : Imunisasi Campak pada Balita Berdasarkan Provinsi)," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018.

[5] S. Sudirman, A. P. Windarto, and A. Wanto, "Data Mining Tools | RapidMiner : K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops In Indonesia," *IOP Conference Series: Materials Science and Engineering*, vol. 420, no. 12089, pp. 1–8, 2018.

[6] A. P. Windarto, L. S. Dewi, and D. Hartama, "Implementation of Artificial Intelligence in Predicting the Value of Indonesian Oil and Gas Exports With BP Algorithm," *International Journal of Recent Trends in Engineering & Research (IJRTER)*, vol. 3, no. 10, pp. 1–12, 2017.

[7] I. S. Sulistyorini, M. Edwin, and A. S. Arung, "ANALISIS KUALITAS AIR PADA SUMBER MATA AIR DI KECAMATAN KARANGAN DAN KALIORANG KABUPATEN KUTAI TIMUR," *Jurnal Hutan Tropis*, vol. 4, no. 1, pp. 64–76, 2016.

[8] D. L. Fithri, "MODEL DATA MINING DALAM PENENTUAN KELAYAKAN PEMILIHAN TEMPAT TINGGAL MENGGUNAKAN METODE NAIVE BAYES," *Jurnal SIMETRIS*, vol. 7, no. 2, pp. 725–730, 2016.

[9] H. Brilian Argario, N. Hidayat, and R. Kartika Dewi, "Implementasi Metode Naive Bayes Untuk Diagnosis Penyakit Kambing (Studi Kasus : UPTD . Pembibitan Ternak dan Hijauan Makanan Ternak Kec. Singosari Malang)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 8, pp. 2719–2723, 2018.

[10] W. Muslehatin, M. Ibnu, and Mustakim, "Penerapan Naïve Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau," *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)*, pp. 18–19, 2017.

[11] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec Journal*, vol. 2, no. 3, pp. 207–217, 2015.

[12] S. F. Rodiyansyah, "NAÏVE BAYES CLASSIFICATION UNTUK PENENTUAN KELAYAKAN DONOR DARAH," pp. 156–159, 2014.

[13] Y. I. Kurniawan, "PERBANDINGAN ALGORITMA NAIVE BAYES DAN C.45 DALAM KLASIFIKASI DATA MINING," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 5, no. 4, pp. 455–464, 2018.

[14] T. Arifin, "METODE DATA MINING UNTUK KLASIFIKASI DATA SEL NUKLEUS DAN SEL RADANG BERDASARKAN ANALISA TEKSTUR," *Informatika*, vol. II, no. 2, pp. 425–433, 2015.

[15] M. Guntur, J. Santony, and Yuhandri, "Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko," *Jurnal RESTI(Rekayasa Sistem dan Teknologi Informasi)*, vol. 2, no. 1, pp. 354–360, 2018.