

Klasifikasi Gangguan Jaringan *Low Bandwidth* Bank Tabungan Negara (BTN) dengan Perbandingan Metode *K-Nearest Neighbor* dan Metode *Naive Bayes* (Studi Kasus : PT. Ipwan Global Telecomm)

Sashmita Salsabeela dan Latifah

STMIK Jakarta STI&K

Jl. BRI No.17, Radio Dalam Kebayoran Baru Jakarta Selatan 12140

E-mail: sashmitasalsabeela@gmail.com, latifahbahrudinsuryobroto@gmail.com

Abstrak

Jaringan internet sudah tidak asing lagi pada era komputerisasi saat ini. Jaringan internet dapat dibedakan menurut besaran transfernya yaitu tinggi dan rendah. Besaran rendah atau disebut juga *low bandwidth* sering digunakan pada perusahaan perbankan untuk perangkat dengan *virtual private network* (VPN) seperti ATM, *Openbooth*, dan lain sebagainya. Begitupun dengan BANK TABUNGAN NEGARA (BTN) yang memasang jaringan internet *low bandwidth* dengan bantuan orang ketiga yaitu PT IPWAN GLOBAL TELECOMM. Pada saat ini PT IPWAN GLOBAL TELECOMM telah memasang jaringan *internet low bandwidth* milik BANK TABUNGAN NEGARA (BTN) sebanyak lebih dari 350 titik yang menyebabkan banyaknya terjadi kasus gangguan pada jaringan. Belum adanya klasifikasi terhadap gangguan jaringan *low bandwidth* menjadi masalah dalam penelitian ini. Oleh karena itu peneliti membuat penelitian klasifikasi jaringan internet dengan membandingkan metode Naive Bayes dan K-Nearest neighbor untuk menemukan metode dengan *accuracy* terbaik. Hasil dari penelitian ini menunjukan *Naive bayes* mendapatkan *accuracy* sebesar 53 % sedangkan *K-Nearest Neighbor* mendapatkan *accuracy* sebesar 52%. Hal ini menunjukan *Naive Bayes* merupakan metode terbaik dalam penelitian ini. Kemudian, hasil dari penemuan metode terbaik tersebut menghasilkan sebuah *prototype* berupa sistem klasifikasi gangguan jaringan low bandwidth dengan menggunakan bahasa pemrograman *Hypertext Preprocessor* (PHP) dan database Mysql.

Kata kunci : *Data Mining*, Klasifikasi, *K-Nearest Neighbor*, *Naive Bayes*, Gangguan jaringan *low bandwidth*

Pendahuluan

Jaringan internet adalah hal yang biasa digunakan di era komputerisasi seperti saat ini dan telah digunakan di berbagai aspek kehidupan termasuk lembaga pemerintahan maupun swasta. Salah satunya adalah perbankan yang membutuhkan jaringan internet yang dapat di klasifikasikan berdasarkan *bandwidth* yang tinggi maupun rendah. *Bandwidth* adalah besaran transfer yang dilakukan dalam satu waktu saat mengakses jaringan internet. Penggunaan *bandwidth* besar mencakup penggunaan jaringan koneksi LAN untuk komputer dan juga wifi yang umum digunakan sehari-hari, sedangkan *bandwidth* kecil mencakup ATM, *Payment point*, dan sebagainya yang merupakan bagian dari *Virtual Private Network* (VPN).

Dalam mengelola jaringan internet berklasifikasi

low bandwidth atau *bandwidth* rendah, bank membutuhkan orang ketiga yang merupakan perusahaan penyedia layanan jaringan internet untuk memenuhi banyaknya kebutuhan pemasangan di lokasi yang tersebar diseluruh Indonesia. Salah satu dari pihak ketiga yang ditunjuk adalah PT IPWAN GLOBAL TELECOMM yang bergerak di bidang jaringan dan dipercayakan untuk mengelola jaringan internet low bandwidth BANK TABUNGAN NEGARA (BTN). Total lokasi pemasangan jaringan low bandwidth BANK TABUNGAN NEGARA (BTN) yang dipercayakan pada PT IPWAN GLOBAL TELECOMM mencapai lebih dari 350 titik lokasi diseluruh Indonesia dengan gangguan yang tercatat selama periode tahun 2017 sampai dengan tahun 2020 sebanyak lebih dari 600 kasus gangguan jaringan. Banyaknya lokasi pemasangan yang tidak terlepas dari gangguan

jaringan menimbulkan kendala dalam proses *maintenance*, sehingga dibutuhkan analisis data guna meningkatkan performa kecepatan dalam perbaikan dengan mengetahui skala gangguan jaringan.

Analisa data tersebut dapat dilakukan dengan menggunakan algoritma data mining, dengan mengubah algoritma dari kumpulan data menjadi sebuah informasi yang sebelumnya tidak diketahui. Banyak metode yang dapat dilakukan untuk melakukan termasuk dalam analisis klasifikasi data dalam algoritma *data mining* seperti *Support Vector Machine*, *Naïve Bayes*, *Constant*, *KNN*, *Tree*, *Random Forest*, *C45* dan lain sebagainya. Berdasarkan penelitian yang dilakukan oleh Iqbal Chun Gui, dkk[5]. perbandingan LOF, *Naive Bayes*, dan ABM. untuk mendeteksi potensi pelanggan jaringan 4G di China menemukan bahwa *Naïve Bayes Classifier* adalah metode yang paling baik digunakan dengan akurasi 67.35% dibandingkan dengan LOF dan ABM yaitu 38% dan 60% . Amani Ghazzawi[3] dalam penelitiannya menghasilkan ID3 dan *Naïve Bayes Classifier* adalah metode terbaik dibandingkan dengan 2 perbandingannya yaitu *KNN* dan *Random Trees*. Penelitian yang dilakukan Rahman dkk[13] menyatakan bahwa pada penelitian klasifikasi air bersih Data Mining *KNN* dan *Naive Bayes* memiliki akurasi yang tinggi masing-masing yaitu 82,42% dan 70,32% . Hasil pengujian lain menunjukkan nilai akurasi terbaik yaitu metode *Tree* , *Naïve Bayes* dengan nilai CA sebesar 0,972 yang dilakukan oleh Irvi Oktanisa[4].

Berdasarkan penelitian terdahulu dan kebutuhan yang diinginkan untuk mengklasifikasikan data agar memudahkan dalam pengambilan keputusan dan meningkatkan performa jaringan berdasarkan hasil klasifikasi data yang terbaik maka akan dibuat analisis data mining untuk klasifikasi gangguan jaringan low bandwidth BANK TABUNGAN NEGARA (BTN) dengan studi kasus PT IPWAN GLOBAL TELECOMM menggunakan perbandingan metode *Naïve Bayes Classifier* dan metode *K-Nearest Neighbor*.

Data gangguan jaringan *low bandwidth* pada pemasangan jaringan BANK TABUNGAN NEGARA (BTN) yang dilakukan oleh PT IPWAN GLOBAL TELECOMM selama 3 tahun terakhir berjumlah lebih dari 600 data. Kemudian data tersebut akan melalui proses untuk menghilangkan noise dan pemilihan atribut yang sesuai dengan tujuan penelitian. Menurut Wilem Musu, dkk[17], data mining sangat dipengaruhi oleh komposisi data latih dan data testing. Data testing terbaik didapat oleh komposisi dengan perbandingan data latih dan testing sebesar 70 banding 30 dan 80 banding 20 dengan accuracy lebih dari 99% . Oleh karena itu pada penelitian ini data training dan data testing akan dibagi menjadi 80 banding 20 yaitu berupa 80% data latih dan 20% data testing. Data yang telah melalui proses tahapan-tahapan tersebut yang kemudian akan melalui proses data mining dengan algoritma *Naive Bayes* dan *K-Nearest neighbor*

yang kemudian hasil accuracy kedua metode tersebut akan dilihat hasil terbaik dalam perbandingannya. Menurut penelitian terdahulu kedua metode tersebut memang metode yang paling mudah digunakan karna sifatnya yang bebas terhadap atribut sehingga memudahkan peneliti dalam menentukan nilai kelas yang diinginkan[12]. Hasil akhir dari penelitian adalah membuat sebuah prototype berupa sistem aplikasi dengan metode terbaik yang telah melewati perbandingan dalam penelitian ini.

Tinjauan Pustaka

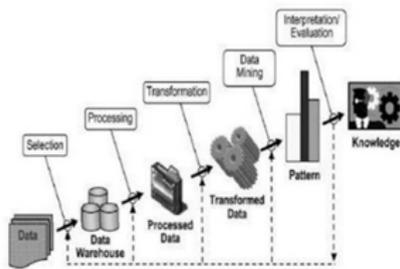
Data mining adalah proses untuk mengumpulkan data dengan menggunakan algoritma matematika, dan statistika serta *mechine learning* dan kecerdasan buatan guna pencarian pola atau informasi dan pengetahuan yang dapat bermanfaat dan memiliki nilai informasi lebih dari berbagai data yang sebelumnya belum diketahui secara manual[6]. Data mining merupakan proses ekstraksi data dalam sebuah pola untuk memanen informasi dan pengetahuan yang dapat dilakukan dengan otomatis dan cepat bahkan tanpa bantuan oleh programmer.

Menurut Yuli Mardi[7]. Fungsi data mining secara umum adalah sebagai berikut:

1. *Description* (Diskripsi) Fungsi dari diskripsi adalah untuk menggambarkan sebuah pola yang memiliki kecenderungan dalam sebuah data. Contoh adalah keterangan dan fakta tentang siapa yang paling profesional dalam pemilihan presiden. Deskripsi dari kemungkinan tersebut memungkinkan memberikan penjelasan untuk suatu pola.
2. *Estimation* (Estimasi) Estimasi memiliki fungsi yang mirip dengan klasifikasi, namun memiliki variable dengan nilai numerik. Proses nya dibangun dengan menggunakan data dengan nilai dari variable yang akan prediksi. Sebagai contoh adalah estimasi nilai mahasiswa pada semester ini.
3. *Prediction* (Prediksi) Prediksi memiliki kegunaan yang hampir mirip dengan estimasi dan klasifikasi. Namun, metode ini memiliki hasil dan nilai yang menunjukkan masa depan. Salah satu contoh adalah prediksi harga 3 bulan mendatang dan presentase kenaikan kecelakaan yang akan terjadi tahun depan.
4. *Classification* (Klasifikasi) Klasifikasi adalah data mining yan memiliki target variable berdasarkan kategori. Sebagai contoh klasifikasi dapat di golongkan menjadi 3 kategori yaitu tinggi, sedang, dan rendah. Salah satu contoh kasus yaitu menentukan terjadinya kecurangan atau tidak dalam transaksi kartu kredit.

5. *Clustering* (Pengklusteran) Fungsi dari pengklusteran adalah mengelompokkan record dan mengamati untuk membentuk sebuah kelas objek yang memiliki kesamaan. Berbeda dengan klasifikasi, clustering tidak memiliki variable. Algoritma kluster melakukan pembagian pada seluruh data menjadi kelompok yang mirip dan kelompok tersebut memiliki nilai maksimal dan juga minimal. Contohnya adalah kelompok-kelompok konsumen dalam target pemasaran sebuah perusahaan.
6. *Association* (Asosiasi) Asosiasi bisa juga disebut keranjang bisnis yang tugasnya untuk menemukan atribut yang muncul pada waktu tertentu. Contoh kasusnya adalah meneliti jumlah pelanggan yang memberikan respons positif pada pelayanan yang diberikan.

Tahapan data mining menurut Fayyad,U[9] adalah sebagai berikut:



Gambar 1: Tahapan Data Mining[8]

1. *Data Selection* Merupakan tahapan untuk memilih atau seleksi dari sekumpulan data. Data yang telah di proses melalui seleksi akan digunakan dalam proses selanjutnya dan akan disimpan terpisah dari data sebelumnya.
2. *Pre-Processing/Cleaning* Proses cleaning merupakan proses penghilangan *noise*, duplikasi, dan data yang tidak konsisten. Proses ini juga bertujuan untuk memperbaiki kesalahan-kesalahan yang terdapat pada data seperti pengetikan dan sebagainya.
3. *Transformation* Tahapan ini merupakan proses untuk mentransformasi data yang telah dipilih sebelumnya, sehingga data telah sesuai dengan kebutuhan penelitian sehingga dapat digunakan dalam proses data mining.
4. *Data Mining* Tahapan ini adalah tahapan yang menggunakan sebuah algoritma pada data terpilih untuk menemukan informasi dan pola dalam sebuah data.
5. *Interpretation/Evaluation* Tahapan ini merupakan tahapan akhir yang merupakan pemeriksaan apakah informasi dan pola yang

ditemukan dalam proses sebelumnya bertentangan dengan hipotesa atau fakta yang ada sebelumnya.

Metode K-Nearest Neighbor

Metode yang juga dikenal dengan KNN ini merupakan metode yang didalam prosesnya menggunakan hasil dari *query instance* yang telah diklasifikasikan berdasarkan data terbanyak dari kelas label pada KNN. Tujuan dari algoritma KNN adalah untuk mengklasifikasikan objek berdasarkan data training dan atribut. KNN bekerja pada *query instance* pada jarak terpendek ke data training. Salah satu cara menghitung jarak dekat atau jauh dengan menggunakan metode *Euclidean Distance*[14].

Euclidean Distance memang biasanya sering digunakan untuk menghitung jarak dan menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak dua objek. Berikut ini merupakan persamaan *Euclidean Distance*.

$$d(x_i k, x_j k) = \left(\sum_{k=1}^m (x_i k - x_j k)^2 \right) \quad (1)$$

Keterangan:

$x_i k$ = Nilai x pada data training

$x_j k$ = Nilai x pada data testing

m = btsd jumlah banyaknya data

$d(x_i k, x_j k)$ = Jarak antara data testing dan data training

Metode Naïve Bayes

Naïve Bayes adalah metode untuk klasifikasi data dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu yang berisikan tentang peluang masa depan berdasarkan pengalaman dimasa sebelumnya. Teorema tersebut kemudian dikombinasikan dengan *Naïve Bayes* dimana diasumsikan kondisi antar atribut menjadi saling bebas. Klasifikasi dengan menggunakan Metode *Naïve Bayes* mengemukakan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak memiliki hubungan dengan ciri kelas lainnya[12].

Persamaan teorema Bayes adalah:

$$P(H|X) = \frac{P(H|X).P(H)}{P(X)} \quad (2)$$

Keterangan :

X : kelas data yang belum diketahui

H : Hipotesis dari data X merupakan suatu kelas yang spesifik

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

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

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

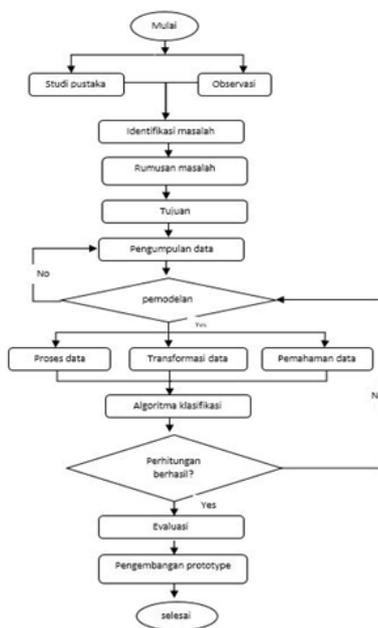
$P(X)$: Probabilitas X

Software RapidMiner Software

RapidMiner merupakan perangkat lunak dengan sifat terbuka atau *open source* yang dapat menjadi sebuah solusi untuk melakukan analisis data mining. *RapidMiner* memiliki fitur yang lengkap seperti banyaknya algoritma yang tersedia di dalamnya dan bentuk grafis yang canggih dan dapat memudahkan pengguna dalam pengoperasian proses *data mining*.

Metodologi Penelitian

Berikut merupakan tahapan metode penelitian yang akan dilakukan dalam penelitian ini dan dapat dilihat pada Gambar 2.



Gambar 2: Alur metodologi penelitian

Pada gambar 2 dapat dijelaskan alur metodologi penelitian diawali dengan melakukan studi pustaka terhadap penelitian-penelitian terdahulu dan observasi permasalahan yang terjadi. Kemudian dari masalah tersebut tersebut dibuatlah identifikasi, rumusan masalah, serta tujuan penelitian. Proses selanjutnya adalah mengumpulkan data. Jika data telah siap untuk proses selanjutnya maka proses dilanjutkan dengan melakukan pemodelan dengan tahapan memproses data dengan memilih atribut yang sesuai dan menghilangkan kesalahan data, melakukan transformasi data dengan merubah data menjadi kategori numerik, lalu memahami data dari setiap atribut yang ada. Namun, jika data yang ada masih kurang untuk masuk dalam tahap pemodelan, proses akan kembali pada pengumpulan data. Tahapan selanjutnya setelah tahap pemodelan, yaitu proses algoritma klasifikasi dengan menggunakan KNN dan Naive Bayes. Jika pengklasifikasian berhasil maka proses akan diteruskan pada

tahap evaluasi. Namun, jika proses klasifikasi gagal, proses harus kembali pada tahap pemodelan. Hal tersebut dikerenakan pemodelan berpengaruh penting pada berhasil atau tidaknya sebuah perhitungan data mining.

Pembahasan dan Hasil Penelitian

Pengumpulan data

Data sekunder yang bersumber dari aplikasi gangguan jaringan PT IPWAN GLOBAL TELECOMM pada lokasi pemasangan jaringan BANK TABUNGAN NEGARA (BTN) diambil pada kurung waktu dari tahun 2018 sampai dengan tahun 2020 yang memiliki data yang berjumlah 536 record. Kemudian data melewati proses pemodelan dengan tahapan data *cleansing* yaitu proses menghilangkan *null* atau *noisy*, sehingga data tereduksi menjadi sebesar 487 data. Data pada penelitian ini akan diproses dengan menggunakan 80% data latih dan 20% data uji, sehingga jumlah untuk data training sebesar 390 data dan data testing sebesar 97 data.

Pra- Pemrosesan Data

Data yang diperoleh sebesar 536 data yang dimulai dari tahun 2018 sampai dengan tahun 2020 memiliki *data noisy* atau *null* yang artinya data tersebut tidak memiliki value, maka data melalui proses *data cleansing* untuk menghilangkan data yang tidak memiliki *value* agar dapat diproses dengan baik dalam penelitian. Untuk *data noisy* atau *null* terdapat pada tabel dibawah ini:

Tabel 1: Proses Data Cleansing

No	Lokasi	Cabang	Jenis pemasangan	Nomer	Jenis penanganan
1	ATM CYBER PARK BEKASI	BEKASI	ATM	81197911350	perbaikan
2	ATM SPBU PENGASI NAN	BEKASI	ATM	85814160723	perbaikan
3	KPOS BOGOR	BOGOR	Payment Point kPOS	81188678889	perbaikan
...					
...					
487	KPOS CIAWI	BOGOR	Payment Point kPOS	85814190876	perbaikan

Pemahaman Data

Pemahaman data merupakan proses analisa atribut-atribut yang akan digunakan pada penelitian. Atribut yang terdiri dari 10 atribut pada *datasheet* sebelumnya dan direduksi sesuai dengan kebutuhan penelitian menjadi 4 atribut yang kemudian diklasifikasikan lagi berdasarkan kategori. Setelah melakukan seleksi atribut dan pemahaman

data, hasilnya kemudian dilanjutkan kembali dengan proses *feature selection* untuk memilih lagi atribut yang telah diseleksi jika masih ada yang kurang relevan dan harus dihilangkan. Pada penelitian ini proses akan menggunakan *feature selection* dengan 1 atribut calss label yaitu atribut “kategori penanganan” yang berfungsi mengoptimalkan kinerja suatu model dengan menggunakan *software RapidMiner*. Setelah menjalankan proses *feature selection*, maka atribut yang dianggap tidak signifikan akan dihilangkan.

Transformasi Data

Tahap transformasi adalah data dalam bentuk tabel yang sudah dikelompokan dirubah menjadi format numerik. Berikut proses transformasi data menjadi data numerik Klasifikasi cabang pada tabel 2.

Tabel 2: Transformasi Atribut Klasifikasi Cabang

Klasifikasi Cabang	Transformasi
JABODETABEK	1
LUAR JABODETABEK	2

Pada atribut Jenis Pemasangan yang sudah dikelompokan dan dirubah dengan format numerik pada tabel 3.

Tabel 3: Transformasi Atribut Jenis pemasangan

Jenis Pemasangan	Transformasi
ATM	1
Open Booth	2
Kpos	3
Payment Point	4
Back up link	5

Pada atribut Kartu yang sudah dikelompokan dan dirubah dengan format numerik pada tabel 4.

Tabel 4: Transformasi Atribut Kartun

Jenis Pemasangan	Transformasi
Indosat	1
Telkomsel	2
Smart Fren	3

Setelah data melewati proses transformasi menjadi nilai numerik, penelitian akan diteruskan dengan proses perhitungan dengan metode *K-Nearest Neighbor* dan *Naive Bayes*. Fungsi dari proses transformasi adalah agar memudahkan dalam proses perhitungan data mining sehingga perlunya membuat grup berdasarkan kriteria pada data tersebut.

Data Latih dan Data Testing

Setelah data di proses dalam tahap transformasi, data yang berjumlah 487 record yang didalam data tersebut terdapat 210 data gangguan berat, 149 data gangguan sedang, dan 128 data gangguan ringan digunakan untuk data latih dan data testing. Data latih adalah data yang digunakan untuk di proses data mining setelah melewati tahap-tahap sebelumnya. Sedangkan testing data merupakan bagian yang diambil dari data training yang berfungsi untuk menguji rule klasifikasi. Pada penelitian ini data menggunakan perbandingan 80:20.

Pemodelan Algoritma K-Nearest Neighbor

Hasil proses data mining dapat dilihat pada tab result. Pada tab performance vector dapat melihat hasil accuracy yaitu sebesar 51.55% pada proses KNN. Hasil dari proses KNN dapat dilihat pada gambar 3.

accuracy: 51.55%				
	true BERAT	true SEDANG	true RINGAN	class precision
pred BERAT	43	19	19	53.00%
pred SEDANG	7	7	2	43.75%
pred RINGAN	0	0	0	0.00%
class recall	86.00%	26.92%	0.00%	

Gambar 3: Hasil Proses perhitungan K-NN

Pemodelan Algoritma Naive Bayes

pada tab result. Pada tab *performance vector* dapat melihat hasil accuracy metode *Naive Bayes* sebesar 52,58% dapat dilihat pada gambar 4 .

accuracy: 51.55%				
	true BERAT	true SEDANG	true RINGAN	class precision
pred BERAT	43	19	19	53.00%
pred SEDANG	7	7	2	43.75%
pred RINGAN	0	0	0	0.00%
class recall	86.00%	26.92%	0.00%	

Gambar 4: Hasil Accuracy Pada Metode Naive Bayes

Pengujian Model

Pengujian *Confusion Matrix* untuk dataset yang telah melalui proses metode KNN yaitu berupa 80% data latih dan 20% data testing yang dapat dilihat pada tabel 5

Tabel 5: Confusion Matrix Pengujian Data Dengan KNN

Prediksi	Class			class precision
	BERAT	SEDANG	RINGAN	
BERAT	43	18	14	57%
SEDANG	9	6	5	30%
RINGAN	0	1	1	50%
class recall	83%	24%	5%	

Untuk perhitungan pada Tabel 5 dapat dihitung sebagai berikut :

1. Untuk nilai presisi (menunjukkan kualitas) dari kelas Berat, dapat dilihat pada persamaan 1.

$$PrecisionBerat = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FP_i} = \frac{43}{43 + 18 + 14} = 57\%$$

2. Untuk nilai presisi (menunjukkan kualitas) dari kelas Ringan, dapat dilihat pada persamaan 2.

$$Precision.Sedang = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FP_i} = \frac{6}{9 + 6 + 5} = 30\%$$

3. Untuk nilai presisi (menunjukkan kualitas) dari kelas Ringan, dapat dilihat pada persamaan 3.

$$PrecisionRingan = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FP_i} = \frac{0}{0 + 1 + 1} = 50\%$$

4. Untuk nilai recall (menunjukkan kualitas) dari kelas Berat, dapat dilihat pada persamaan 4.

$$RecallBerat = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FN_i} = \frac{43}{43 + 9 + 1} = 83\%$$

5. Untuk nilai recall (menunjukkan kualitas) dari kelas Sedang, dapat dilihat pada persamaan 5

$$RecallSedang = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FN_i} = \frac{6}{6 + 18 + 1} = 24\%$$

6. Untuk nilai recall (menunjukkan kualitas) dari kelas Ringan, dapat dilihat pada persamaan 6.

$$RecallRingan = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FN_i} = \frac{1}{14 + 5 + 1} = 50\%$$

7. Untuk nilai accuracy yang dapat secara keseluruhan yang seperti pada persamaan 7.

$$accuracy = \frac{\sum_{i=1}^l TP_i + TN_i}{\sum_{i=1}^l TP_i + TN_i + FP_i + FN_i} = \frac{43 + 6 + 1}{43 + 18 + 14 + 9 + 6 + 5 + 0 + 1 + 1} = 52\%$$

Kemudian tahap selanjutnya hasil akan dibandingkan antara metode Naïve Bayes perhitungan secara manual dengan perhitungan menggunakan software RapidMiner terhadap hasil perbedaan antara tingkat accuracy, precision, dan recall. Untuk tingkat hasil perbandingan terdapat pada Tbel 6.

Tabel 6: Hasil perbandingan Manual dan Rapid Miner KNN

Class precision	Manual	57%	30%	50%
	RapidMiner	53,09%	43,75%	0%
Class recall	Manual	83%	24%	5%
	RapidMiner	86%	26,92%	0%
Accuracy	Manual	52%		
	RapidMiner	51,55%		

Tahap selanjutnya adalah pengujian Confusion Matrix untuk dataset yang telah melalui proses metode Naïve Bayes yaitu berupa 80% data latih dan 20% data testing yang dapat dilihat pada tabel 7

Tabel 7: Confusion Matrix Pengujian Data Dengan Naïve Bayes

Prediksi	Class			class precision
	BERAT	SEDANG	RINGAN	
BERAT	42	17	17	55%
SEDANG	8	9	4	43%
RINGAN	0	0	0	0%
class recall	84%	35%	0%	

Untuk perhitungan pada tabel 7 dapat dihitung sebagai berikut :

8. Untuk nilai presisi (menunjukkan kualitas) dari kelas Berat, dapat dilihat pada persamaan 8.rumus

$$PrecisionBerat = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FP_i} = \frac{42}{42 + 17 + 17} = 55\%$$

9. Untuk nilai presisi (menunjukkan kualitas) dari kelas Ringan, dapat dilihat pada persamaan 9.

$$PrecisionSedang = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FP_i} = \frac{9}{8 + 9 + 4} = 43\%$$

10. Untuk nilai presisi (menunjukkan kualitas) dari kelas Ringan, dapat dilihat pada persamaan 10.

$$PrecisionRingan = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FP_i} = \frac{0}{0 + 0 + 0} = 0\%$$

11. Untuk nilai recall (menunjukkan kualitas) dari kelas Berat, dapat dilihat pada persamaan 11.

$$RecallBerat = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FN_i} = \frac{42}{42 + 8 + 0} = 84\%$$

12. Untuk nilai recall (menunjukkan kualitas) dari kelas Sedang, dapat dilihat pada persamaan 12.

$$RecallSedang = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FN_i} = \frac{9}{17 + 9 + 0} = 35\%$$

13. Untuk nilai recall (menunjukkan kualitas) dari kelas Ringan, dapat dilihat pada persamaan 13.

$$RecallRingan = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FN_i} = \frac{0}{17 + 4 + 0} = 0\%$$

14. Untuk nilai recall (menunjukkan kualitas) dari kelas Sedang, dapat dilihat pada persamaan 12.

$$Accuracy = \frac{\sum_{i=1}^l TP_i + TN_i}{\sum_{i=1}^l TP_i + TN_i + FP_i + FN_i} = \frac{42 + 9 + 0}{42 + 17 + 17 + 8 + 9 + 4 + 0 + 0 + 0} = 53\%$$

Tahap selanjutnya hasil akan dibandingkan antara metode *Naive Bayes* perhitungan secara manual dengan perhitungan menggunakan *software RapidMiner* terhadap hasil perbedaan antara tingkat *accuracy*, *precision*, dan *recall*. Untuk tingkat hasil perbandingan terdapat pada tabel 8.

Tabel 8: Confusion Matrix Pengujian Data Dengan KNN

Prediksi	Class			class precision
	BERAT	SEDANG	RINGAN	
BERAT	43	18	14	57%
SEDANG	9	6	5	30%
RINGAN	0	1	1	50%
class recall	83%	24%	5%	

Hasil dan Evaluasi

Pada hasil perbandingan metode antara KNN dan *Naive Bayes* terdapat hasil dari tingkat *accuracy*, *precision*, dan *recall* tersebut. Pada penelitian ini data yang digunakan adalah 80:20 dari 536 keseluruhan data. Pada penelitian KNN menggunakan k-3 dimana dalam proses datanya membandingkan dari k-1 hingga k-10 dan mendapat hasil yang lebih baik pada k-3. Untuk tingkat hasil perbandingan antara masing-masing metode dapat dilihat pada tabel 9.

Penutup

Berdasarkan pembahasan yang telah dijelaskan sebelumnya, maka dalam penelitian gangguan jaringan pada pemasangan jaringan low bandwidth menggunakan metode K-Nearest Neighbor dan metode *Naive Bayes* dapat disimpulkan bahwa KNN menghasilkan *accuracy* sebesar 52% dengan *precision* 46% dan *recall* 37%. Sedangkan metode *Naive Bayes* menghasilkan *accuracy* sebesar 53% dengan *precision* 33% dan *recall* 40% yang artinya metode *Naive Bayes* lebih baik pada penelitian ini dalam tingkatan *accuracy*. Sehingga pada penelitian ini digunakan metode *Naive Bayes* pada sistem untuk klasifikasi gangguan jaringan low bandwidth dan diharapkan dengan adanya klasifikasi gangguan jaringan low bandwidth dapat mempermudah PT IPWAN GLOBAL TELECOMM dalam melihat sekala prioritas gangguan jaringan guna meningkatkan performa pelayanan dimasa mendatang.

Daftar Pustaka

[1] Fitri Handayani dan Feddy Setio Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110", Jurnal Teknik Elektro, Vol. 7 No. 1, 2015.

[2] Edy Widodo dan Pathyatus Sa,'adah, "Klasifikasi Data Penjualan Alat Tulis Kantor (ATK) Terlaris untuk Optimasi Strategi Pemasaran di

- Toko Citramedia Menggunakan Metode Naive Bayes”, Jurnal Teknik Informatika Fakultas Teknik, Volume 10 Nomor 2, 2019.
- [3] Amani Ghazzawi, Basma Alharbi, “Analysis of Customer Complaints Data using Data Mining Techniques”, *Procedia Computer Science*, Volume 163, pp: 62-69, 2019.
- [4] Irvi Oktanisa , Ahmad Aff Supianto, “Perbandingan Teknik Klasifikasi dalam Data Mining untuk Bank Direct Marketing ”, *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, Vol. 5, No. 5, 2018.
- [5] Chun Gui , Qiang Lin, “Comparison of three data mining algorithms for potential 4G customers prediction”, *Jurnal mathematics and computer science*. (Vol. 6, No. 1), 2016.
- [6] S. Widaningsih, “Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, KNN dan SVM”, *Jurnal Tekno Insentif*, 13(1), pp. 16–25, 2019.
- [7] Yuli Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5”, *Jurnal Edik Informatika*, V2.i2(213-219), 2016.
- [8] A.N. Khormarudin, “Teknik Data Mining: Algoritma K-Means Clustering”, *Ilmu Komputer*, pp. 1–12, 2016.
- [9] U. Fayyad, “Advances in Knowledge Discovery and Data Mining”, MIT Press, 1996.
- [10] Senna Hendrian, “Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan”, *Jurnal Prodi Studi Informatika*, v11i3.2777, 2018.
- [11] Bustami, “Penerapan Algoritma Naive Bayes Untuk Nasabah Asuransi”, *Jurnal Informatika*, 8(1), pp. 884–898, 2014.
- [12] L. M. Chaira, “Pemilihan Jenis Asuransi Berdasarkan Demografi Calon Pemegang Polis Dengan Metode Naïve Bayes Classifier”, *Limits: Journal of Mathematics and Its Applications*, 13(2), p. 11, 2016.
- [13] M.A. Rahman, N. Hidayat dan A. Afif Supianto, “Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)”, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (JPTIIK)*, 2(12), pp. 925–928.
- [14] W. Ode, N. Kadir dan B.S. Pramono, “Penerapan Data Mining Dengan Metode Knearest Neighbor (Knn) Untuk Mengelompokan Minat Konsumen Asuransi (PT. JASARA-HARJA PUTERA)”, 5(1), pp. 97–104, 2019.
- [15] Saiful Ulya, M. Arief Soeleman, Fikri Budiman, “Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Priorita Bantuan Pembangunan Desa”, *Techno.COM*, Vol.20,No.1:83-96, 2021.
- [16] Indrayati, Devi Sugianti, M. Adib Al Karomi, “Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Penyakit Diabetes Melitus”, *Prosiding SNATIF*, 2017.
- [17] Wilem Musu, Abdul Ibrahim dan Heriadi, “Pengaruh Komposisi Data Training dan Testing Terhadap Akurasi Algoritma C4.5”, *Program Studi Manajemen Informatika, Universitas Dipa Makassar*, Vol 10 No 1 (2021): SISITI, 2021.