

# Analisis Layanan Pelanggan PT PLN Berdasarkan Media Sosial Twitter Dengan Menggunakan Metode Naïve Bayes Classifier

Tri Prasetyo<sup>1</sup>, Hadi Zakaria<sup>2\*</sup>, Pandu Wiliantoro<sup>3\*</sup>

<sup>1,2,3</sup>Teknik, Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: <sup>1</sup>[dosen02669@unpam.ac.id](mailto:dosen02669@unpam.ac.id), <sup>2\*</sup>[dosen00274@unpam.ac.id](mailto:dosen00274@unpam.ac.id), <sup>3\*</sup>[dosen02668@unpam.ac.id](mailto:dosen02668@unpam.ac.id)

**Abstrak**– Layanan untuk daya dan energi pada saat ini masih dikelola oleh perusahaan dari Perusahaan Listrik Negara (PLN) yang menjadi salah satu perusahaan dalam lingkup Badan Usaha Milik Negara (BUMN). Perusahaan PT. PLN ini memiliki tujuan yaitu melayani untuk kepentingan umum serta menjaga mutu dan kualitas dari daya dan energi itu sendiri. Dengan berorientasi pada kepuasan pelanggan oleh sebab itu PT. PLN akan selalu menjaga kualitas agar keuntungan yang didapatkan menjadi semakin besar pada setiap waktunya. Dilakukan pembagian pada opini publik yaitu opini negative yang berisi keluhan masyarakat, opini positive yang berisi dukung masyarakat, kemudian opini netral yang berisi laporan atau pengaduan dari masyarakat. Kata kunci yang digunakan untuk menjaring respon dan opini masyarakat terhadap layanan PT PLN dalam proses web scraping tersebut adalah menggunakan kata kunci yaitu pln tangerang kabupaten”, “pln tangerang kota”, “pln tangerang selatan”, “pln banten. Berdasarkan penelitian yang telah dilakukan dengan analisis sentiment melalui penggunaan twitter berdasarkan tweet yang telah dilakukan dengan menggunakan metode naïve bayes classifier dengan hasil akurasi lebih dari 80%.

**Kata Kunci:** Sentimen, Layanan, PT PLN, Twitter, Naïve Bayes Classifier

**Abstract**– Services for power and energy are currently still managed by a company from the State Electricity Company (PLN), which is one of the companies within the scope of State-Owned Enterprises (BUMN). Company PT. This PLN has the aim of serving the public interest and maintaining the quality and quality of power and energy itself. By being oriented to customer satisfaction, PT. PLN will always maintain quality so that the profits get bigger every time. Public opinion is divided into negative opinions containing public complaints, positive opinions containing community support, then neutral opinions containing reports or complaints from the public. The keywords used to capture public responses and opinions on PT PLN services in the web scraping process are using keywords, namely pln tangerang district", "pln tangerang city", "pln tangerang south", "pln banten. Based on research that has been done with sentiment analysis through the use of twitter based on tweets that have been carried out using the nave Bayes classifier method with more than 80% accuracy results.

**Keywords:** Sentiment, Service, PT PLN, Twitter, Naïve Bayes Classifier

## 1. PENDAHULUAN

Teknologi berupa internet saat ini telah menjadi bagian dari masyarakat pada kehidupan sehari-hari dimana setiap aktifitas data diketahui memanfaatkan jaringan internet. Jumlah pengguna internet setiap tahun mengalami kenaikan. Melalui jaringan internet lahirlah sebuah media sosial yang pada fungsi awalnya adalah untuk membantu melakukan interaksi tanpa adanya kendala dalam segi jarak dan waktu. Berbagai dampak positive dan negative pada dasarnya banyak ditemukan pada penggunaan media sosial dimana penggunaan media sosial menjadi jauh dari kata bermanfaat. Salah satu media sosial yang sudah lama berkembang dan telah banyak digunakan oleh masyarakat adalah twitter. Banyak masyarakat yang telah menggunakan media sosial seperti twitter tidak terkecuali pelanggan dari PT PLN. Berbagai keluhan, laporan serta belum dapat menerima kebijakan yang dikeluarkan PT PLN dapat dicurahkan pada media sosial seperti twitter.

Twitter merupakan media sosial yang banyak digunakan oleh sebagian kalangan yang memiliki smartphone. Dengan twitter masyarakat dapat melakukan opini terkait berbagai hal yang menyangkut dengan kehidupan sosial seperti sebuah layanan entah dari perusahaan yang dikelola oleh swasta ataupun perusahaan yang dikelola oleh negara.

Salah satu media sosial biasa digunakan oleh PT. PLN adalah twitter. Melalui akun twitter @pln\_123 segala macam opini masyarakat yang ingin menyuarakan seperti pelayanan, kualitas, mutu serta hal yang ingin dilaporkan dapat ditampung dengan wadah melalui akun tersebut. Akun @pln\_123 merupakan akun resmi yang dikelola dibawah naungan PT. PLN.

Dengan cakupan wilayah PT PLN yang tersebar diseluruh Indonesia analisis opini penting dilakukan karena dengan menganalisis opini masyarakat pada salah satu wilayah hal ini dapat menjadi contoh bagi wilayah lain yang belum maksimal dalam melayani masyarakat sehingga dapat diminimalisir kesalahan dan kekurangan berdasarkan wilayah yang sedang di analisis saat ini.

Wilayah yang daerahnya sangat luas dan banyak pelanggan dari PT PLN adalah pada wilayah provinsi Banten. Dengan total 5,3 juta pelanggan serta konsumsi listrik 54,65 Giga Watt/Hour (Gwh) Banyaknya pelanggan yang berada pada provinsi Banten tentu harus dilayani melalui kualitas yang memadai sehingga kualitas pelayanan dari PT PLN tidak berkurang dari waktu ke waktu.

Analisis opini berupa analisis layanan dilakukan agar PT PLN dapat mengetahui bagaimana kualitas pelayanan serta mutu dimana nantinya hasil analisis tersebut dapat dijadikan landasan dalam menentukan kebijakan lanjutan atau mengubah kebijakan yang telah ada agar pelayanan terhadap masyarakat dapat terus meningkat menjadi lebih baik lagi

Opini Sebagai masyarakat yang juga menjadi pelanggan dari PT PLN sangat penting untuk menilai kualitas layanan dari PT. PLN. Oleh sebab itu guna melakukan analisis layanan diperlukan suatu metode yang mampu mengklasifikasikan pelayanan untuk masyarakat atau pelanggan oleh sebab itu digunakan algoritma Naïve Bayes. Dengan menggunakan algoritma naïve bayes maka opini yang berisi layanan negative, layanan positive serta netral dapat diklasifikasikan.

Analisis layanan dapat mendukung dalam hal mengetahui opini di pelanggan sehingga PT PLN dapat menjawab permasalahan serta memberikan solusi yang diharapkan pelanggan. Analisis layanan juga diterapkan bertujuan untuk mengetahui pandangan pelanggan tentang kualitas layanan yang sedang dijalankan.

Melalui analisis layanan ini akan menghasilkan informasi berupa kualitas layanan dari PT PLN kepada pelanggan di daerah provinsi Banten berupa opini negative, opini positive dan opini netral. Dengan hasil tersebut dapat dijadikan informasi yang berguna bagi perusahaan untuk dapat agar memperbaiki kualitas layanan yang telah diterapkan. Berdasarkan latar belakang yang telah dijelaskan, dilakukan penelitian agar bertujuan untuk mengetahui layanan bagi masyarakat yang menjadi pelanggan PT PLN khususnya daerah provinsi Banten dengan menggunakan metode algoritma Naïve Bayes. Oleh sebab itu penelitian ini berjudul “ANALISIS LAYANAN PELANGGAN PT. PLN BERDASARKAN MEDIA SOSIAL TWITTER DENGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER (Studi Kasus : PT PLN Banten)”.

## 2. METODOLOGI PENELITIAN

### 2.1 Waktu dan Lokasi Penelitian

Penelitian ini akan dilaksanakan dalam waktu kurang lebih 6 (enam) bulan sejak ditanda tangannya kontrak perjanjian penelitian.

Tabel 1. Waktu Penelitian

No	Kegiatan	Waktu Penelitian						
		Nov 2021	Des 2021	Jan 2022	Feb 2022	Mar 2022	Apr 2022	Mei 2022
1	Studi Literatur	█						
2	Pengumpulan Data		█					
3	Pelabelan Data					█		
4	Pembuatan Model					█		
5	Proses Klasifikasi						█	
6	Penulisan							█

## 2.2 Responden Penelitian

Responden penelitian ini adalah masyarakat yang berhubungan secara langsung terhadap kualitas pelayanan PT PLN di provinsi Banten melalui tweets pada media sosial twitter.

## 2.3 Metode Pengumpulan Data

Pengumpulan data respon dan opini masyarakat Indonesia terhadap kualitas layanan PT. PLN provinsi Banten dilakukan dengan menggunakan teknik web scraping pada media sosial twitter dalam bentuk tweets. Kegiatan web scraping tersebut dilakukan dengan menggunakan API twitter dengan menggunakan package 'rtweet' (Basu R, 2017). 'rtweet' merupakan package yang dirancang untuk dapat mengumpulkan dan mengatur data twitter menggunakan API twitter dari aplikasi R (Pramana S, 2018).

Kata kunci yang digunakan untuk menjangkau respon dan opini masyarakat terhadap layanan PT PLN dalam proses web scraping tersebut adalah menggunakan kata kunci yaitu "pln tangerang kabupaten", "pln tangerang kota", "pln Tangerang selatan", "pln banten". Kata kunci yang digunakan dinilai dapat menjangkau semua

opini masyarakat khususnya wilayah didalam provinsi Banten terhadap kualitas layanan PT PLN di media sosial twitter. Data yang diambil merupakan data dari tanggal 28 April 2022 sampai tanggal 05 Mei 2022 karena adanya keterbatasan pengumpulan data.

## 2.4 Metode Persiapan

Selanjutnya, dilakukan kegiatan persiapan data/preprocessing untuk mempersiapkan data agar siap untuk dianalisis. Kegiatan preprocessing tersebut meliputi tahapan pembersihan tweets terhadap unsur-unsur yang tidak dibutuhkan dalam analisis, yaitu delete duplicate atau menghapus tweets respon dan opini masyarakat yang sama persis, delete URL atau menghapus link yang terdapat pada tweets, menghapus mentions dan hastags, menghapus emoji, menghapus punctuation, melakukan normalisasi kata, menghapus kata yang tidak penting (stopwords removal), dan mengubah format tulisan menjadi huruf kecil. Tahapan normalisasi kata berguna untuk mentransformasi kata singkatan, typo, dan kata berlebih menjadi sebuah kata yang formal . (Salsabila NA, 2018).

Penghapusan kata yang tidak penting (stopwords) berguna untuk mengurangi waktu sistem dalam merunning data. Kamus stopwords yang digunakan pada penelitian ini menggunakan kamus yang dibuat (FZ., 2003). Tahapan preprocessing merupakan tahapan yang paling penting dalam penelitian yang menggunakan data hasil text mining, karena pada tahapan ini sangat menentukan hasil analisis yang akan didapatkan.

Analisis sentimen atau sering disebut dengan istilah opinion mining merupakan gabungan dari beberapa teknik, meliputi natural language processing (NLP), information retrieval (IR), dan data mining (DM) yang mengolah atau menganalisis opini, sentimen, dan emosi yang diekspresikan dalam bentuk teks pada suatu entitas. Pada analisis sentimen, model klasifikasi bisa digunakan untuk menentukan sentimen ke dalam dua kelas atau lebih. Task sentiment analysis bisa dilakukan menggunakan pendekatan teknik machine learning ataupun pendekatan berbasis lexicon (Ravi, 2015). Dataset yang digunakan dalam penelitian ini adalah public timeline tweet bahasa Indonesia yang merupakan hasil pencarian berdasarkan tweet user yang mengandung kata layanan layanan PLN Banten. Dataset didapatkan dengan cara scraping memanfaatkan API dari twitter kemudian disimpan dalam bentuk database. Data yang diambil merupakan data dari tanggal 28 April 2022 sampai tanggal 05 Mei 2022.

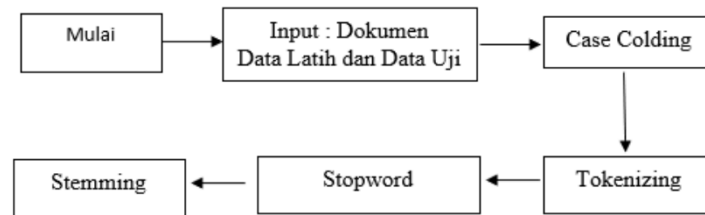
Selanjutnya, adalah tahapan case folding yaitu mengubah semua huruf dalam data menjadi lowercase (huruf kecil) semua, hal ini dikarenakan analisis sentimen tidak melihat besar kecilnya huruf, sebagai contoh kata 'Hitung' dan 'hitung' dianggap sama. Kemudian tokenizing, yaitu tahap dimana setiap kalimat yang ada didalam data tes dipisah menjadi setiap kata. Tahapan selanjutnya adalah stopword atau mengubah kata yang tidak penting, dalam hal ini kata-kata yang tidak mengandung sentimen dengan contoh kata hubung. Terakhir adalah mengubah

semua kata yang ada di dalam data tes ke dalam kata dasar (root word), tahapan ini disebut juga sebagai proses stemming. Setelah semua proses preprocessing selesai, maka data tes siap untuk

diproses dalam analisis sentimen

## 2.5 Preprocessing

Preprocessing merupakan teknik data mining yang melibatkan perubahan data mentah menjadi sebuah format yang terstruktur dan dimengerti. Data mentah seringkali tidak lengkap, tidak konsisten dan mungkin mengandung banyak kesalahan. Teknik preprocessing terbukti dapat menyelesaikan masalah tersebut. Ada beberapa tahapan dalam teknik preprocessing yaitu tokenization, case folding, filtering dan stemming (Kiprono, 2016).



**Gambar 1. Proses Preprocessing**

Case folding merupakan tahapan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil (lowercase). Tahapan kedua, yaitu tokenizing merupakan tahapan pemotongan string input berdasarkan tiap kata yang menyusunnya. Tokenizing juga digunakan untuk membuang beberapa karakter yang dianggap sebagai tanda baca. Selanjutnya, stopwords adalah tahap mengambil kata penting dari hasil tokenizing dengan algoritma stopwords removal (membuang kata yang kurang penting). Stopword merupakan kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag of words. Tahap terakhir, stemming adalah proses normalisasi dalam sistem information retrieval yang digunakan untuk mencari kata dasar (root word) yang terdapat dalam suatu dokumen atau term dengan mengacu pada aturan-aturan tertentu.

## 2.6 Naïve Bayes Classifier

Naïve Bayes Classifier merupakan sebuah metode klasifikasi dengan probabilitas sederhana yang mengaplikasikan teorema bayes dengan tidak ketergantungan (independent) yang tinggi. Algoritma naïve bayes digunakan sebagai penggolong beberapa masalah dunia nyata seperti analisis sentimen, deteksi spam email, pengelompokan otomatis email, pengurutan email berdasarkan prioritas dan kategorisasi dokumen. Model klasifikasi naïve bayes menghitung probabilitas posterior suatu kelas, berdasarkan pada distribusi kata-kata dalam dokumen (Ravi, 2015). Klasifikasi naïve bayes dibangun oleh data pelatihan untuk memperkirakan probabilitas dari setiap kategori yang terdapat pada ciri dokumen yang diuji. Sistem akan dilatih dengan menggunakan data baru (data latih dan data uji) dan selanjutnya diberi tugas untuk menebak nilai fungsi target dari data tersebut (Destuardi dan Surya, 2009). Secara umum, proses klasifikasi dengan menggunakan naïve bayes dapat dilihat dari persamaan :

$$P(cj/wi) = \frac{P(wi/cj) \times P(cj)}{P(wi)}$$

$P(cj/wi)$ : peluang kategori j ketika terdapat kemunculan kata i

$P(wi/cj)$ : peluang sebuah kata i masuk kedalam kategori j

$P(cj)$  : peluang kemunculan sebuah kategori j

$P(cj/wi)$ : peluang kemunculan sebuah kata

Peluang kemunculan sebuah kata bisa dihilangkan pada perhitungan klasifikasi karena peluang kemunculan kata tidak akan berpengaruh pada perbandingan hasil klasifikasi setiap kategori. Proses klasifikasi dapat disederhanakan sebagai berikut :

$$P(cj/wi) = P(wi/cj) \times P(cj)$$

Untuk menghitung prior atau peluang kemunculan suatu kategori pada semua dokumen dapat dilakukan dengan menggunakan persamaan :

$$P(c) = \frac{Nc}{N}$$

$Nc$  : banyak kategori  $c$  pada dokumen latih

$N$  : banyak keseluruhan dokumen yang digunakan

**Tabel 1. Confusion Matrix**

<i>Class</i>	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Negative</i>	<i>False Positive (FP)</i>	<i>False Negative (TN)</i>

### 2.7 Confusion Matrix

Confusion matrix adalah matrix 2x2 yang merepresentasikan hasil klasifikasi biner pada suatu dataset. Terdapat beberapa rumus umum yang dapat digunakan untuk menghitung performa klasifikasi. Hasil dari nilai accuracy, precision dan recall bisa ditampilkan dalam presentase.

#### 1. Accuracy

Accuracy adalah jumlah proporsi prediksi yang benar. Adapun rumus perhitungan akurasi dapat dilihat dari persamaan dibawah ini (Manning, 2009).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

#### 2. Precision

Precision adalah proporsi jumlah dokumen teks yang relevan terkendali diantara semua dokumen teks yang terpilih oleh sistem. Rumus precision dapat dilihat pada persamaan dibawah ini (Manning, 2009).

$$Precision = \frac{TP}{TP + FP}$$

#### 3. Recall

Recall adalah proporsi jumlah dokumen teks yang relevan terkendali diantara semua dokumen teks relevan yang ada pada koleksi. Rumus recall (Manning, 2009) dapat dinyatakan dengan dibawah ini :

$$Recall = \frac{TP}{TP + FN}$$

### 2.8 Rapidminer Studio

Rapid miner studio adalah aplikasi opensource yang digunakan untuk melakukan proses data mining. Rapid miner menyediakan prosedur data mining dan machine learning, dalam machine learning mempunyai bagian yaitu, ETL (extraction, transformation, learning) Rapid miner mempunyai tools tipe nilai seperti (Imelda A. Muis & Muhammad Affandes, 2015).

1. Nominal  
Nominal adalah dimana tipe nilai yang berdasarkan nilai secara pertagori
2. Numeric  
Numeric adalah mengambil secara umum
3. Integer  
Integer adalah tipe nilai yang digunakan untuk bilangan, seperti bilangan bulat
4. Real  
Real adalah dimana tipe nilai yang digunakan secara nyata
5. Text  
Text adalah dimana tipe nilai yang digunakan untuk teks bebas tanpa terstruktur
6. Binominal  
Binominal adalah dimana tipe nilai yang digunakan terdiri dari dua nilai
7. Polynominal  
Polynominal adalah dimana nominal lebih dari dua nilai
8. Date-time

Date-time adalah digunakannya untuk tanggal dan waktu

### 3. ANALISA DAN PEMBAHASAN

Dalam penelitian ini, Penulis menggunakan Rapidminer Studio sebagai wadah untuk pembuatan analisis sentiment berdasarkan data tweet yang didapatkan dari twitter . Rapidminer studio memiliki tampilan yang mudah sekali untuk digunakan karena sederhana sehingga dapat dengan cepat dapat dipahami cara kerjanya. Rapidminer Studio mempunyai kelebihan yaitu dapat menampilkan kesalahan yang terjadi dan keharusan user untuk melengkapi data atau tools yang dapat melengkapi proses analisis sentiment

#### 3.1 Data Latih

Pada pelatihan ini menjelaskan proses membentuk model prediksi dari data yang telah diketahui kelasnya. Data yang digunakan dalam penelitian ini terdiri dari 2 jenis yaitu data latih dan data uji. Data latih yang digunakan di ambil dari kumpulan data tweet yang telah di labeli sesuai sentimennya secara manual, terdapat 2 jenis label yang dibagi menjadi sentiment positive dan negative. Pada penelitian ini peneliti menggunakan data latih sebanyak 41 data dan menghasilkan akurasi 88.24%. Berikut tampilan data latih pada program rapidminer studio

Row No.	y	prediction(y)	Text
1	negative	negative	@pln_123 terjadi pemadaman listrik di Kp. Cengklong RT. 024 RW. 012 Kosambi - Tangerang
2	negative	negative	@pln_123 daerah Sepatan Kab. Tangerang mati lampu ada gangguan apa ya min?
3	positive	positive	PLN Optimalisasi Infrastruktur Kelistrikan di Bali - Berita Kota Tangerang Selatan - Bidik Tangsel <a href="https://t.co/Sty6FtrVOU">https://t.co/Sty6FtrVOU</a>
4	negative	negative	@pln_123 @baybajang Min respon dm dong, kab.tangerang mati listrik juga nih
5	negative	negative	@pln_123 @jandream7 Iya nih kapan selesai penanganya? . Daerah tanah tinggi tangerang mati total sejak hari ini 02
6	negative	positive	RT @JP6String: @pln_123 min mohon info apa penyebab Taman Royal 1 Tangerang mati lampu? Terus perbaikannya
7	negative	negative	@pln_123 gimana sih tangerang mati lampu nih hari raya gelap-gelapan masa!!!!
8	negative	negative	@pln_123 ini kenapa ya di daerah jurumudi, benda, tangerang mati lampu? Padahal lagi hari raya loh ini ?? baru hujan

**Gambar 1. Data Latih**

Arti dari y dan prediction(y) adalah kelas dari kalimat negative dan positive. Kemudian Text adalah isi dari tweet yang telah diambil dari twitter. Data yang diinputkan berupa data dari format excel.

#### 3.2 Data Uji

Pada data uji ini menjelaskan proses puncak untuk mengetahui tingkat ke akuratan pada analisis. Pada proses ini peneliti akan melakukan penginputan data dan memperoleh hasil berupa kelas berupa label positive dan negative. Hasil ini didapatkan berdasarkan data latih yang sebelumnya telah dimasukkan kedalam program rapidminer studio. Berikut tampilan data uji pada program rapidminer studio.

Row No.	prediction(y)	Text
1	negative	@pln_123 terjadi pemadaman listrik di Kp. Cengklong RT. 024 RW. 012 Kosambi - Tangerang
2	negative	@pln_123 daerah Sepatan Kab. Tangerang mati lampu ada gangguan apa ya min?
3	positive	PLN Optimalisasi Infrastruktur Kelistrikan di Bali - Berita Kota Tangerang Selatan - Bidik Tangsel <a href="https://t.co/Sty6FtrVOU">https://t.co/Sty6FtrVOU</a>
4	negative	@pln_123 @baybajang Min respon dm dong, kab.tangerang mati listrik juga nih
5	negative	@pln_123 @jandream7 Iya nih kapan selesai penanganya? . Daerah tanah tinggi tangerang mati total sejak hari ini 02 mei 2022 pukul 16
6	positive	RT @JP6String: @pln_123 min mohon info apa penyebab Taman Royal 1 Tangerang mati lampu? Terus perbaikannya berapa lama? Ter
7	negative	@pln_123 gimana sih tangerang mati lampu nih hari raya gelap-gelapan masa!!!!
8	negative	@pln_123 ini kenapa ya di daerah jurumudi, benda, tangerang mati lampu? Padahal lagi hari raya loh ini ?? baru hujan sebentar udah me

**Gambar 2. Data Uji**

Hasil dari data uji dengan hasil yang didapatkan oleh peneliti dengan cara manual untuk mendapatkan ketepatan prediksi. Hasil dari uji manual yang sudah dilakukan peneliti diinputkan kedalam program rapidminer studio. Disini peneliti menggunakan 17 data. Oleh karena itu pengujian



yang dilakukan peneliti dapat disimpulkan bahwa hasil pengujian menunjukkan hasil yang sesuai dengan yang diharapkan.

### 3.3 Confusion Matrix

Confusion Matrix dilakukan untuk menguji hasil klasifikasi dengan mengukur nilai kebenaran dari sistem. Jika dataset memiliki 2 kelas/label, yaitu positive dan negative. Dalam penelitian ini entri dari dua baris dan kolom confusion matrix dirujuk sebagai true and false positive dan true and false negative. Seperti pada tabel dibawah ini.

**Tabel 2. Confusion Matrix**

	Predicted	
	Positive	Negative
Positive	True Positive	True Negative
Negative	False Positive	False Negative

True Positive (TP) merupakan jumlah record positive yang diklasifikasikan sebagai positive, false positive (FP) adalah jumlah record negative yang diklasifikasikan sebagai positive, false negative (FN) adalah jumlah record positive yang diklasifikasikan sebagai negative, true negative (TN) adalah jumlah record negative yang diklasifikasikan sebagai negative. Dan klasifikasi text, biasanya pengukuran akurasinya menggunakan pengukuran Accuracy

## 4. IMPLEMENTASI

### 4.1 Accuracy

Accuracy merupakan presentase dari data yang telah diproses kedalam sebuah program salah satunya didalam program rapid miner studio. Accuracy yang dapat dihasilkan dari proses analisa dan perhitungan dapat ditunjukkan pada persamaan dibawah ini

$$Accuracy = \frac{TP+TN}{(TP+FP+FN+TN)}$$

Dari 17 data testing yang sudah diproses ketepatan prediksinya, didapatkan hasil TP : 0, TN: 10, FP: 5, FN: 2.

accuracy: 88.24%

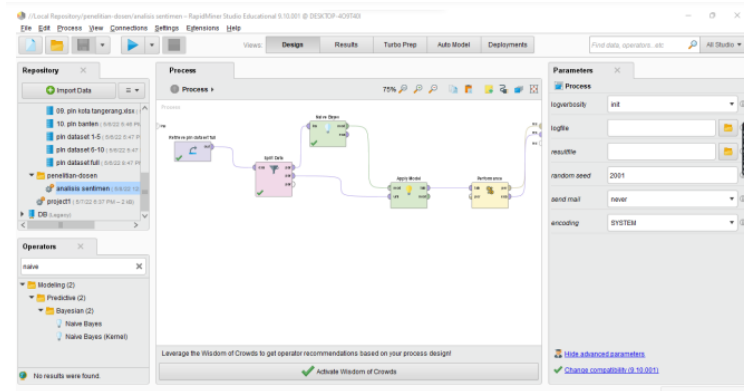
	true negative	true positive	class precision
pred. negative	10	0	100.00%
pred. positive	2	5	71.43%
class recall	83.33%	100.00%	

**Gambar 3. Accuracy**

Didapatkan hasil dari accuracy antara uji manual sebesar 88.24% dengan demikian memiliki angka presentase yang cukup tinggi diklasifikasikan dengan tepat oleh sistem. Hasil yang diperoleh dari pengujian yang telah dilakukan bahwa perhitungan accuracy sama dengan hasil yang dilakukan secara manual sehingga didapatkan sesuai yang diharapkan yaitu nilai yang cukup tinggi karena data testing yang diinputkan memiliki kosakata yang ada pada data training, sehingga sistem dapat mempelajari dari data training dan dapat memproses data testing dengan sangat baik.

### 4.2 Analisis Sentimen

Antarmuka rapidminer studio ini berfungsi untuk pengguna secara umum untuk melakukan analisis pada data seperti analisis sentiment serta untuk melihat hasil dari proses analisa dengan menginputkan sebuah data tweet dari twitter dan memiliki kelas berupa positive atau negative. Antarmuka yang sediakan ini menggunakan inputan berupa text yang nantinya dapat menggunakan beberapa operator atau algoritma yang sudah disediakan rapidminer studio untuk dapat mengolah sebuah data dan menghasilkan informasi. Data yang diinputkan bisa berupa data training dan juga data testing menghasilkan output berupa tabel atau presentase.



**Gambar 4. Antarmuka Analisis Sentimen**

**4.3 Tabel Analisis Sentimen**

Tabel analisis sentiment pada rapidminer studio ini berfungsi menampilkan hasil data yang telah diolah dengan menggunakan operator naïve bayes sehingga pengguna secara umum tidak perlu lagi kesulitan untuk menghitung secara manual data yang telah di inputkan kedalam program.

No.	y	prediction(y)	Text	From-User	confidence(n...	confidence(positiv...	Created-At	From-User-Id	To-I
	negative	negative	@pln_123 ter...	Rulie	1	0	May 4, 2022 1...	424811032	pln...
	negative	negative	@pln_123 da...	cd17j7	1	0	May 3, 2022 7...	2568136323	pln...
	positive	positive	PLN Optmail...	Info Tangsel	0.000	1.000	May 3, 2022 6...	411473395	?
	negative	negative	@pln_123 @...	Raa	1	0	May 2, 2022 7...	1354436726...	pln...
	negative	negative	@pln_123 @...	EBEN ESER ...	1	0	May 2, 2022 5...	1271815657	pln...
	negative	positive	RT @JP6Sbt...	Andika Prad...	0	1	May 2, 2022 5...	78028096	?
	negative	negative	@pln_123 gi...	del	1	0	May 2, 2022 5...	1065312190...	pln...
	negative	negative	@pln_123 in...	S a n t o s a o	1	0	May 2, 2022 5...	589809004	pln...
	negative	negative	@pln_123 m...	JP	1	0	May 2, 2022 5...	1147506044...	pln...
	negative	negative	@pln_123	son of father	1	0	May 2, 2022 5...	1206070295...	pln...
	negative	positive	Viral Beli Pul...	Don Lomax	0.000	1.000	May 2, 2022 1...	174193809	?
	negative	negative	@pln_123 M...	djoker	1	0	Apr 30, 2022 ...	1331316147...	pln...
	negative	negative	@pln_123 m...	ridwan	1	0	May 2, 2022 6...	1409352681...	pln...
	positive	positive	Ciptakan Ras...	Polda Banten	0.000	1.000	May 3, 2022 1...	1402525522...	?

**Gambar 5. Tabel Analisis Sentimen**

**4.4 Statistik Analisis Sentimen**

Statistik analisis sentiment pada rapidminer studio ini berfungsi menampilkan hasil berupa statistic secara lengkap sesuai dari label yang telah ditentukan user dengan menggunakan data yang telah diolah memanfaatkan operator naïve bayes sehingga pengguna secara umum tidak perlu lagi membuat presentase secara manual pada program lain.

Name	Type	Missing	Statistics	Filter (16 / 16 attributes)
y	Binominal	0	negative: positive: 12: 5	negative (12), positive (5)
prediction(y)	Binominal	0	negative: positive: 10: 7	negative (10), positive (7)
confidence(negative)	Real	0	0	0.566
confidence(positive)	Real	0	1	0.412
Created-At	Date time	0	Earliest date: Apr 26, 2022 6:55 PM Latest date: May 4, 2022 11:30 AM	5d 16h 34m 32s
From-User	Polynomial	0	Least: (0)	Most: Andika Pradipta (1)
From-User-Id	Real	0	Min: 78028096 Max: 1497862700617322500	Average: 685735027836022140

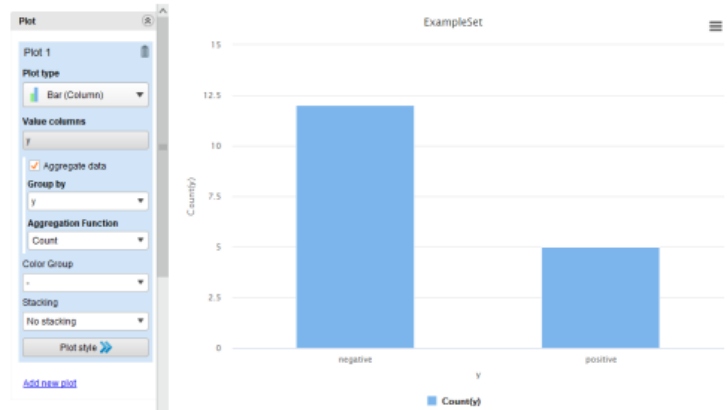
**Gambar 6. Statistik Analisis Sentimen**

**4.5 Visualisasi Analisis Sentimen**

Tampilan visual dari analisis sentiment pada rapidminer studio ini berfungsi menampilkan hasil berupa grafik secara lengkap sesuai dari label yang telah ditentukan user. Pilihan visualisasi dapat diatur sesuai keinginan dan kebutuhan dari user sehingga tampilan informasi berdasarkan data



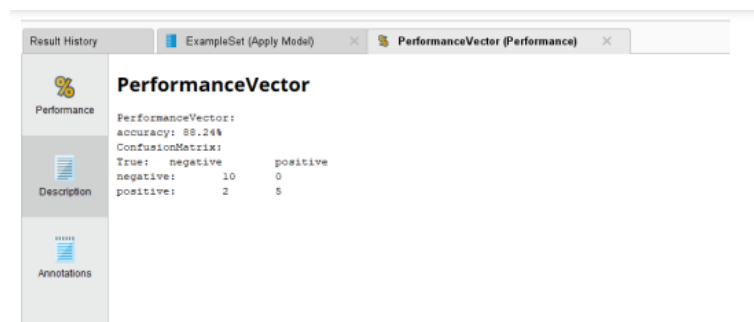
yang telah diolah dapat lebih mudah dipahami sehingga user dapat memudahkan dalam menyimpulkan namun tetap memanfaatkan operator naïve bayes.



**Gambar 7. Visualisasi Analisis Sentimen**

#### 4.6 Performance Vector

Tampilan performance vector dari analisis sentiment pada rapidminer studio ini berfungsi menampilkan hasil berupa accuracy secara akurat sesuai dengan inputan data yang telah diinputkan user.



**Gambar 8. Performance Vector**

## 5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dengan analisis sentiment melalui penggunaan twitter berdasarkan tweet yang telah dilakukan dengan menggunakan metode naïve bayes classifier dengan hasil akurasi lebih dari 80%. Hasil analisa untuk layanan PLN khususnya wilayah Banten dan sekitarnya tersebut fokusnya pada pelayanan yang telah dilakukan dimana respon atau pendapat negative masih lebih banyak dari respon positive.

## REFERENCES

- Basu R, K. A. (2017). Harnessing Twitter Data for Analyzing Public Reactions to Transportation Policies. [https://www.researchgate.net/publication/321997978\\_Harnessing\\_Twitter\\_Data\\_for\\_Analyzing\\_Public\\_Reactions\\_to\\_Transportation\\_Policies\\_Evidences\\_from\\_the\\_Odd-Even\\_Policy\\_in\\_Delhi\\_India](https://www.researchgate.net/publication/321997978_Harnessing_Twitter_Data_for_Analyzing_Public_Reactions_to_Transportation_Policies_Evidences_from_the_Odd-Even_Policy_in_Delhi_India).
- Destuardi dan Surya, S. (2009). *Klasifikasi Emosi Untuk Teks Bahasa Indonesia Menggunakan Metode Naive Bayes*. Surabaya: Institut Teknologi Sepuluh November.
- FZ., T. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*.
- Imelda A. Muis & Muhammad Affandes, M. (2015). *Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet Sains*. Riau: Sains, Teknologi dan Industri UIN Sultan Syarif Kasim.



- Kiprono, K. W. (2016). Comparative Twitter Sentiment Analysis Based on Linear and Probabilistic Models. *International Journal on Data Science and Technology*, 41-45.
- Manning, C. R. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Pramana S, Y. B. (2018). *Data Mining dengan R Konsep Setara Implementasi*. Bogor: IN MEDIA.
- Ravi, V. a. (2015). A Survey on Opinion Mining and Sentiment Analysis: Task, Approaches and Applications. *Elsevier Knowledge-Based Systems*, 14-46.
- Salsabila NA, W. Y. (2018). Colloquial Indonesian Lexicon. *International Conference on Asian Language Processing*, 236-9.
- Telkom, U. (2021). *14.04.1010\_bab1.pdf*. Bandung: openlibrary.telkomuniversity.ac.id.