

# Analisis Laporan Gangguan Pelanggan Menggunakan Kombinasi Regresi Linear Berganda dan Klasifikasi *Decision Tree* Untuk Penentuan Komposisi Tim Yantek (Studi Kasus PLN UP3 Makassar Selatan)

Yanuardhi Arief Budiyono<sup>(1)</sup>, Agus Budi Raharjo<sup>(2)</sup>, Diana Purwitasari<sup>(3)</sup>

Institut Teknologi Sepuluh Nopember  
Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota SBY, Jawa Timur, Indonesia

Email: <sup>1</sup>abe.yanuardhi@pln.co.id, <sup>2</sup>agus.budi@its.ac.id, <sup>3</sup>diana@if.its.ac.id

---

## Tersedia Online di

<http://www.jurnal.unublitar.ac.id/index.php/briliant>

---

## Sejarah Artikel

Diterima pada 30 Juni 2022  
Disetujui pada 25 Juli 2022  
Dipublikasikan pada 22 November 2022  
Hal. 1076-1090

---

## Kata Kunci:

Laporan Gangguan; Machine Learning; Regresi; *Decision Tree*

---

## DOI:

<http://dx.doi.org/10.28926/briliant.v7i4.1037>

**Abstrak:** PT PLN (Persero) terus berupaya meningkatkan layanan kepada pelanggan dalam hal durasi penanganan laporan gangguan pelanggan melalui implementasi *command center*. Dengan total pelanggan 549.649, *command center* PLN UP3 Makassar Selatan bertujuan untuk memusatkan pemantauan laporan gangguan pelanggan dan pengaturan petugas pelayanan teknik di lapangan untuk mempercepat durasi penanganan laporan gangguan pelanggan. Dalam rentang waktu tiga tahun sejak Januari 2019 hingga Desember 2021, PLN UP3 Makassar Selatan memiliki laporan gangguan pelanggan sebanyak 225.309 laporan. Penelitian ini bertujuan untuk memprediksi pemenuhan durasi penanganan laporan gangguan sesuai penetapan target kinerja UP3 Makassar Selatan. Analisis prediksi menggunakan metode Regresi Linear pada variabel jumlah laporan dengan target durasi penanganan gangguan pada data laporan gangguan pelanggan yang bersumber dari Aplikasi Keluhan dan Pelayanan Terpadu (APKT). Selanjutnya klasifikasi menggunakan metode *Decision Tree* untuk mengetahui

durasi penanganan gangguan memenuhi target kinerja atau tidak. Hasil penelitian menunjukkan bahwa metode yang diajukan memberikan kinerja yang baik berdasarkan evaluasi *Coefficient of Determination* ( $R^2$ ), *Root Mean Square Error* (RMSE), dan *ROC Curve*. Model yang dibangun diharapkan dapat diimplementasikan kedalam sistem pengaturan petugas teknik sehingga memberikan komposisi jumlah tim untuk menurunkan durasi penanganan laporan gangguan.

## PENDAHULUAN

PLN sebagai perusahaan yang menyediakan jasa pelayanan dan berorientasi pada kepuasan pelanggan, melakukan transformasi layanan digital yang sebelumnya hanya melalui *Call center 123*, saat ini dikembangkan aplikasi *PLN Mobile* dan *Command Center* yang menjadi salah satu wujud kepedulian PLN terhadap peningkatan pelayanan laporan gangguan pelanggan. Dengan total jumlah pelanggan sebanyak 549.649 pada bulan Desember 2021, UP3 Makassar Selatan dalam rentang waktu bulan Januari tahun 2019 hingga bulan Desember 2021, terdapat laporan gangguan pelanggan sebanyak 225.309. Terkumpulnya informasi laporan gangguan pelanggan di *database* dengan jumlah yang besar dan informasi yang beraneka ragam, maka laporan gangguan pelanggan menjadi salah satu cara untuk belajar tentang akar masalah yang harus ditangani dengan serius. Hal ini

menjadi sangat diperlukan untuk mewujudkan persepsi aktif perusahaan atas permintaan pelanggan dan mengendalikan masalah dengan benar (Shuangshuang, et al., 2020).

Data laporan gangguan dari pelanggan dapat digunakan sebagai ukuran yang akurat untuk menghitung seberapa sukses suatu layanan, terutama dengan transformasi teknologi informasi dalam hal ini *Machine Learning* (Ghazzawi & Alharbi, 2019). *Machine learning* yang berfokus pada pengolahan data besar dan mempelajari pola serta karakter data sangat membantu analisis data dengan banyak metode dapat digunakan untuk pengolahan data (Irfan, et al., 2019). Penggunaan *machine learning* sebelumnya telah dilakukan untuk mengukur kepuasan pelanggan menggunakan tiga algoritma prediksi, yaitu *Tree*, *Support Vector Machine*, dan *Ensemble* dari data layanan, data pemadaman listrik, dan indeks keandalan (Cavalcante Siebert, et al., 2019). Hasil dari model kombinasi *Liner Regression*, *Ensemble (boost)*, *Ensemble (bag)*, *Regression Tree*, *Support Vector Machine (SVM)* dan *GPR* memiliki nilai yang sangat baik dengan perbedaan sebesar 1,36% terhadap hasil survey pelanggan.

Dalam penelitian yang lain, prediksi dilakukan untuk mendapatkan informasi *churn* pelanggan. Metode yang digunakan adalah pengelompokan *K-Means* yang selanjutnya hasil dari pengelompokan dilakukan klasifikasi menggunakan metode *Decision Tree* (Mirjana, et al., 2021). Selanjutnya metode *chi-squared automatic interaction detector* (CHAID) digunakan untuk mengidentifikasi faktor *churn* yang paling berpengaruh pada kelompok dengan tingkat *churn* tertinggi. Hasilnya mengungkapkan bahwa *churn* pada pelanggan telekomunikasi terjadi terutama yang terkait pada variabel *tenure*, dengan kemungkinan terbesar pada pelanggan yang baru bergabung dengan layanan perusahaan.

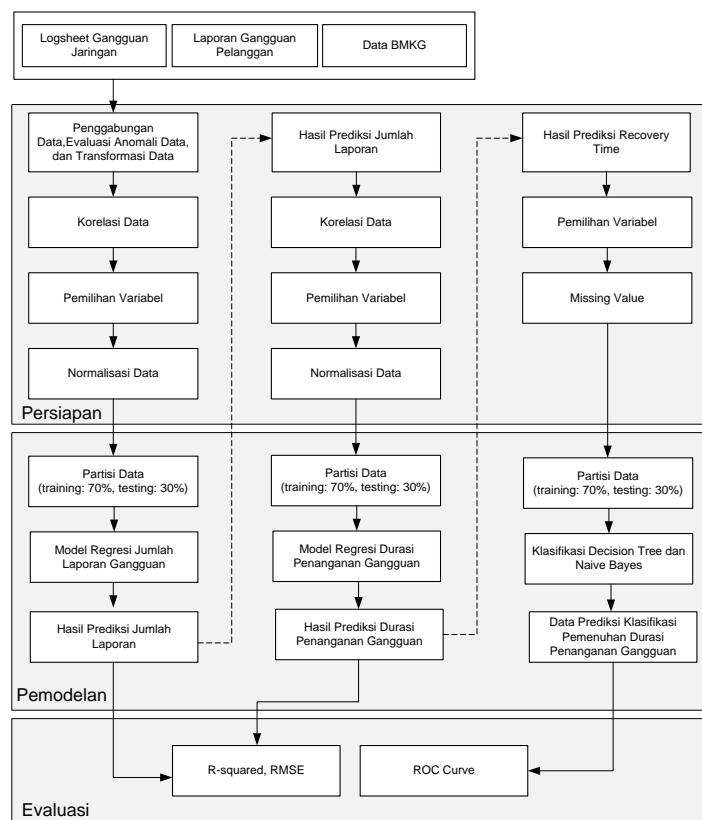
Penelitian berbasis deret waktu untuk memprediksi pengaduan pelanggan diajukan dengan menerapkan algoritma berbasis ARIMA, Regresi Linear berganda, dan jaringan saraf *backpropagation* (BP) (Xin, et al., 2019). Penelitian tersebut mempelajari dan membandingkan tiga algoritma prediksi keluhan pelanggan. Prediksi dilakukan berdasarkan identifikasi tren jangka panjang, variasi musiman, variasi siklus, dan variasi yang tidak teratur. Variabel target yang diambil adalah pengaduan pelanggan tenaga listrik jibei. Hasil pemodelan menampilkan bahwa dengan nilai variasi kesalahan data 10%, 20%, 30%, dan 40% maka algoritma Jaringan Saraf BP memberikan akurasi terbaik dalam memprediksi keluhan pelanggan. Penelitian berbasis deret waktu yang lain untuk memprediksi keluhan pelanggan menggunakan algoritma berbasis ARIMA, *Multiple Linear Regression* (MLP), dan *Backpropagation Neural Network* (BPNN) (Shuangshuang, et al., 2020). Nilai prediksi dari ketiga model tersebut dibandingkan satu sama lain dan hasil dari prediksi BPNN merupakan yang terbaik. Selanjutnya pengoptimalan algoritma genetika digunakan untuk memperbaiki luaran model. Hasilnya menunjukkan bahwa model yang dioptimalkan memiliki kesalahan relatif kurang dari 40%.

Saat ini PLN belum memiliki sistem yang dapat memprediksi laporan gangguan pelanggan. Di sisi lain, berbagai penelitian di atas menunjukkan bahwa *Machine Learning* dapat diterapkan untuk memperkirakan jumlah keluhan pelanggan listrik dan peningkatan layanan. Oleh karena itu, penelitian ini

mengajukan pengembangan teknik pada *Machine Learning* pada studi kasus prediksi laporan gangguan pelanggan di PLN, khususnya PLN UP3 Makassar Selatan. Pengembangan metode yang diajukan meliputi teknik pengolahan data pada fase pra-proses dan kombinasi metode regresi dan klasifikasi. Prediksi durasi penanganan gangguan dihitung menggunakan metode Regresi Linear dengan data masukkan jumlah laporan gangguan harian yang diambil dari data besar aplikasi APKT PLN UP3 Makassar Selatan selama rentang bulan Januari tahun 2019 hingga bulan Desember tahun 2021. Hasil regresi yaitu jumlah laporan gangguan dan durasi penanganan gangguan yang akan menjadi variabel target proses selanjutnya dengan melakukan prediksi klasifikasi terhadap pemenuhan target kinerja durasi penanganan gangguan dengan metode klasifikasi *Decision tree*.

## METODE

Diagram alir metode penelitian dapat dilihat pada *Gambar 1*. Bab ini menjelaskan metode yang digunakan untuk memprediksi jumlah laporan gangguan, durasi penanganan gangguan, dan klasifikasi target kinerja penanganan gangguan. Dalam penelitian dibagi menjadi tiga tahapan yaitu persiapan, pemodelan, dan evaluasi.



Gambar 1. Diagram Alir Tahapan Penelitian.

#### A. Persiapan

Penelitian menggunakan data laporan gangguan pelanggan, data BMKG, dan logsheet gangguan jaringan. Data laporan gangguan pelanggan diambil dalam jangka waktu tiga tahun dengan total 225.309 baris dan 41 kolom yang ditransformasikan menjadi data agregasi laporan harian gangguan pelanggan. Selanjutnya penggabungan tiga data, yaitu data agregasi laporan harian gangguan pelanggan, data BMKG, dan data logsheet gangguan jaringan. Analisis data mentah dilakukan untuk memilih variabel tiap data laporan yang digunakan pada penelitian. Terbentuk struktur dataset dan variabel yang digunakan pada penelitian ini dengan jumlah 730 baris dan 12 variabel ditunjukkan pada Tabel 1. Proses manipulasi dilakukan terhadap data kosong (*missing value*) dengan mengganti data menggunakan nilai rata-rata mengacu pada Persamaan (1).

$$\bar{x} = \frac{\sum x}{n} = \frac{x_1+x_2+\dots+x_n}{n} \quad (1)$$

Setelah dipastikan semua dataset terisi, proses setelahnya mengukur nilai kekuatan hubungan antara variabel target prediksi dan variabel prediktor yang dapat diketahui melalui uji korelasi linear. Koefisien korelasi merupakan nilai yang digunakan untuk memahami nilai hubungan antara variabel (Safitri, W, 2014). Berikut adalah interpretasi nilai koefisien korelasi (Yeager, 2022).

1.  $0,00 < |r| < 0,199$  menunjukkan hubungan kedekatan sangat lemah
2.  $0,20 < |r| < 0,399$  menunjukkan hubungan kedekatan lemah
3.  $0,40 < |r| < 0,599$  menunjukkan hubungan kedekatan sedang (moderate)
4.  $0,60 < |r| < 0,799$  menunjukkan hubungan kedekatan kuat
5.  $0,80 < |r| < 1,000$  menunjukkan hubungan kedekatan sangat kuat

Pada penelitian juga dilakukan normalisasi data pada variabel prediktor dan target menggunakan pendekatan *Min-Max Normalization* dengan rentang 0,01 hingga 1. Pada tahap klasifikasi, data masukkan yang diperoleh pada target klasifikasi bersifat *imbalance* sehingga digunakan teknik *smote* untuk menyeimbangkan data klasifikasi. *Smote* adalah metode *over-sampling* terhadap data minoritas diperbanyak dengan menggunakan data uraian dari penggandaan data pada kelas minoritas (Sutoyo & Fadlurrahman, 2020). Persamaan (2) digunakan pada teknologi *smote*:

$$X_{new} = X_i + (\hat{X}_i - X_i)x\delta \quad (2)$$

- $X_i$  : Vektor dari fitur kelas minoritas.  
 $\hat{X}_i$  : *k-nearest neighbors* untuk  $X_i$ .  
 $\delta$  : Angka acak 0 sampai 1.

Tabel 1. Struktur Dataset Penelitian.

No	Nama Field	Satuan	Tipe	Keterangan	Variabel
1	Bulan		Month/Year	Bulan Berjalan	Fitur
2	Kali Padam	kali	Numerical	Jumlah Gangguan Jaringan	Fitur
3	Level Cuaca	mm	Numerical	Parameter Curah Hujan	Fitur
4	Laporan MTG	kali	Numerical	Jumlah Laporan Gangguan Pelanggan Posko Mattoangin	Target1
5	Durasi MTG	hours	Numerical	Durasi Penanganan Gangguan	Target2
6	Klasifikasi Durasi MTG	Y/N	Character	Klasifikasi Durasi Penanganan Gangguan	Target3
7	Laporan PNK	kali	Numerical	Jumlah Laporan Gangguan Pelanggan Posko Panakkukang	Target1
8	Durasi PNK	hours	Numerical	Durasi Penanganan Gangguan	Target2
9	Klasifikasi Durasi PNK	Y/N	Character	Klasifikasi Durasi Penanganan Gangguan	Target3
10	Laporan MSGM	kali	Numerical	Jumlah Laporan Gangguan Pelanggan Posko Sungguminasa	Target1
11	Durasi SGM	hours	Numerical	Durasi Penanganan Gangguan	Target2
12	Klasifikasi Durasi SGM	Y/N	Character	Klasifikasi Durasi Penanganan Gangguan	Target3

## B. Pemodelan

Pada tahap pemodelan, peneliti membentuk tiga alur proses, yang pertama proses prediksi jumlah laporan gangguan pelanggan menggunakan metode regresi linear berganda. Analisis regresi merupakan analisis statistik yang mempelajari hubungan antara dua atau lebih data sehingga salah satu data dapat diprediksi (Pangesti, 2016). Secara umum, model linear berganda yang digunakan untuk memprediksi dapat dilihat pada Persamaan (3):

$$Y = \alpha + b_1X_1 + b_2X_2 + b_3X_3 + b_nX_n \quad (3)$$

Keterangan :

- Y : Variabel Respon (*dependent*)
- $\alpha$  : Konstanta
- $b_1, b_2, b_3, b_n$  : Koefisien Regresi
- $X_1, X_2, X_3, X_n$  : Variabel Prediktor (*independent*)

Tahap ini dilakukan untuk mendapatkan ramalan jumlah laporan gangguan pelanggan dimasa yang akan datang. Nilai hasil prediksi digunakan sebagai variabel prediktor pada proses selanjutnya, yaitu prediksi durasi penanganan gangguan. Proses prediksi durasi penanganan gangguan menggunakan metode yang sama dengan proses pertama yaitu regresi linear berganda. Dengan mendapatkan nilai ramalan durasi penanganan gangguan pelanggan, maka dapat diprediksi apakah kondisi tersebut memenuhi target kinerja durasi penanganan laporan gangguan atau tidak memenuhi. Alur proses terakhir, dengan variabel bulan, hari, dan cuaca maka kita dapat memprediksi klasifikasi seperti apa durasi penanganan gangguan dapat

terpenuhi dan sebaliknya. Metode yang digunakan adalah *Decision Tree* dibandingkan dengan metode klasifikasi *Naïve Bayes*.

Menurut (Rokach & Maimon, 2005), *Decision Tree* merupakan metode klasifikasi yang membuat partisi-partisi dan terdiri dari node yang terbentuk oleh akar pohon dengan setiap node terdapat satu masukan. Node yang memiliki keluaran disebut tes node, jika tidak memiliki keluaran maka disebut *leaf* atau node keputusan (Rokach & Maimon, 2005). Salah satu algoritma yang digunakan untuk membuat model *Decision Tree* adalah C4.5 yang mampu menangani data numerik dan diskrit (Narulita et al., 2021). Konsep entropi digunakan untuk menentukan pada variabel mana sebuah keputusan akan terbagi (Oktafia & Pardede, 2008). Persamaan (4) digunakan untuk menghitung entropi.

$$Entropy (S) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 \quad (4)$$

$p_1$  : Jumlah populasi atau sampel terpenuhi target

$p_2$  : Jumlah populasi atau sampel tidak terpenuhi target

Salah satu metode klasifikasi yang sering digunakan untuk memprediksi kelas target yang bersifat nominal adalah *Decision Tree* dan *Naïve Bayes* (Rosandy, 2016). *Naïve Bayes* adalah metode klasifikasi paling sederhana dengan menghitung segala kemungkinan yang terjadi dimasa depan dengan asumsi kebebasan pengaruh yang kuat antar variabel yang disebut *Naïve* (Lalwani et al., 2022).

### C. Evaluasi

Untuk menjalankan semua metode di atas, langkah pertama yang dilakukan adalah membagi data latih sebesar 70% dan data uji sebesar 30% sebagai data masukkan setiap model yang dijalankan. Pengukuran kinerja dan akurasi prediksi model yang dibentuk menggunakan indikator *Root Mean Squared Error* (RMSE), *R-squared* ( $R^2$ ), dan *ROC Curve*.

## HASIL DAN PEMBAHASAN

Berdasarkan hasil analisis penggabungan data logsheet gangguan, laporan gangguan pelanggan, dan data BMKG, maka peneliti dapat menentukan variabel data yang dibutuhkan untuk model prediksi. Selanjutnya metode analisis pra pemrosesan yang dilakukan dan model seperti apa yang dibuat. Terakhir, evaluasi implementasi hasil model yang telah dibuat.

### A. Prediksi Jumlah Laporan Gangguan

Pengujian diawali dengan uji korelasi antar variabel data gabungan untuk mengetahui kekuatan hubungan antar variabel. Uji korelasi dilakukan untuk mendapatkan variabel yang dipilih menjadi prediktor dalam model prediksi menggunakan teknik korelasi *pearson* seperti pada Tabel 2. Hasil korelasi prediktor dengan korelasi tertinggi terhadap target adalah level cuaca dan kali padam gangguan. Selanjutnya normalisasi data dengan metode *Min-Max Normalization*, sehingga variabel data yang dipilih berada pada rentang skala yang sama dengan hasil pada

Tabel 3.

Tabel 3. Korelasi Variabel Target Jumlah Laporan.

VARIABEL FITUR	SATUAN	LAPORAN MTG	LAPORAN PNK	LAPORAN SGM	KRITERIA KORELASI MTG	KRITERIA KORELASI PNK	KRITERIA KORELASI SGM
TAHUN	Year	0,251	0,385	0,463	Lemah	Lemah	Sedang
BULAN	Month	0,043	0,111	0,172	Sangat Lemah	Sangat Lemah	Sangat Lemah
TANGGAL	Date	-0,131	-0,075	-0,061	Sangat Lemah	Sangat Lemah	Sangat Lemah
NUM HARI	Day	-0,007	-0,072	-0,045	Sangat Lemah	Sangat Lemah	Sangat Lemah
KALI PADAM	kali	0,587	0,703	0,716	Sedang	Kuat	Sedang
LEVEL CUACA	mm	0,633	0,656	0,705	Kuat	Kuat	Kuat

Tabel 4. Normalisasi Data Prediksi Jumlah Laporan Gangguan.

Variabel	Data Mentah			Data Normalisasi		
	Min	Mean	Max	Min	Mean	Max
Bulan	1	6,51	12	0,01	0,51	1
Kali Padam	1	13,40	43	0,01	0,32	1
Level Cuaca	0	13,37	150	0,01	0,10	1
Laporan MTG	14	53,73	627	0,01	0,07	1
Laporan PNK	15	61,52	293	0,01	0,18	1
Laporan SGM	28	83,46	518	0,01	0,12	1

Data masukkan model dilakukan partisi dengan membagi jumlah data menjadi 70% sebagai data latih model dan 30% sebagai data uji model. Pemilihan data latih dan data uji dilakukan menggunakan metode *linear sampling* terhadap data populasi. Pada proses pembelajaran data, variabel target adalah jumlah laporan masing-masing posko, yaitu Jumlah Laporan Mattoangin, Jumlah Laporan Panakkukang, dan Jumlah Laporan Sungguminasa. Untuk variabel prediktor yang digunakan adalah variabel Bulan, Kali Padam, dan Level Cuaca. Hasil model pembelajaran prediksi jumlah laporan setiap posko adalah sebagai berikut:

Tabel 2. Statistik Model Pembelajaran Posko Mattoangin.

Statistik Regresi Linear Berganda Posko Mattoangin				
Variable	Coeff.	Std. Err.	t-value	P> t
Bulan	0.0009	0.0031	0.2886	0.773
Kali Padam	0.1173	0.008	146.798	0.0
Level Cuaca	0.0743	0.0069	107.396	0.0
Intercept	0.0252	0.0029	87.156	0.0

Tabel 5. Statistik Model Pembelajaran Posko Panakkukang.

Statistik Regresi Linear Berganda Posko Panakkukang				
Variable	Coeff.	Std. Err.	t-value	P> t
Bulan	0.0179	0.0086	20.775	0.0383
Kali Padam	0.3247	0.0221	146.834	0.0
Level Cuaca	0.1607	0.0192	83.905	4.44E-16
Intercept	0.0431	0.008	53.923	1.07E-7

Tabel 6. Statistik Model Pembelajaran Posko Sungguminasa.

Statistik Regresi Linear Berganda Posko Sungguminasa				
Variable	Coeff.	Std. Err.	t-value	P> t
Bulan	0.031	0.0064	48.307	1.81E-6
Kali Padam	0.2866	0.0165	174.105	0.0
Level Cuaca	0.1654	0.0143	115.954	0.0
Intercept	-0.0032	0.006	-0.5377	0.5911

Tabel 7. R-squared Metode Pembelajaran dan Koefisien Prediksi Jumlah Laporan.

POSKO	R <sup>2</sup>	$\alpha$	b1	b2	b3	Keterangan
Mattoangin	0,598	0,0252	0,00091	0,1172	0,0743	Kemampuan variabel prediktor dalam menjelaskan variansi variabel target adalah sebesar 59,8%, dan masih ada 40,2% variansi variabel target tidak mampu dijelaskan variabel prediktor
Panakkukang	0,559	0,0431	0,0179	0,3247	0,1607	Kemampuan variabel prediktor dalam menjelaskan variansi variabel target adalah sebesar 55,9%, dan masih ada 45,1% variansi variabel target tidak mampu dijelaskan variabel prediktor
Sungguminasa	0,672	-0,0032	0,031	0,2866	0,1654	Kemampuan variabel prediktor dalam menjelaskan variansi variabel target adalah sebesar 67,2%, dan masih ada 32,8% variansi variabel target tidak mampu dijelaskan variabel prediktor

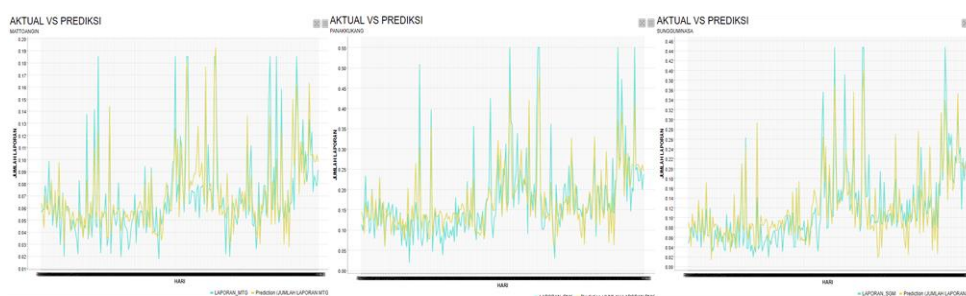
Data luaran model pembelajaran dilakukan pengujian menggunakan 30% data populasi untuk mendapatkan nilai prediksi Jumlah Laporan Gangguan Pelanggan. Luaran hasil uji prediksi diukur menggunakan parameter r-squared (R<sup>2</sup>) dan Root Mean Squared Error (RMSE) sesuai

berikut.



Tabel 8. Statistik Hasil Uji Regresi Jumlah Laporan.

Statistik Hasil Uji Regresi Jumlah Laporan		
Posko	R <sup>2</sup>	RMSE
Mattoangin	0,697	0,019
Panakkukang	0,672	0,056
Sungguminasa	0,679	0,046



Gambar 2. Line Plot Data Aktual dan Prediksi Jumlah Laporan.

Nilai terbaik *r-squared* luaran model pada posko Mattoangin yang menunjukkan angka di atas 69,7%, disimpulkan bahwa variabel prediktor dapat menjelaskan 69,7% variansi variabel targetnya. 30,3% variansi target tidak dapat dijelaskan oleh variabel prediktor. Sedangkan nilai RMSE seluruh posko di atas 90% sehingga disimpulkan bahwa luaran model memiliki akurasi yang sangat baik.

#### B. Prediksi Durasi Penanganan Gangguan

Salah satu variabel yang digunakan sebagai data masukkan model prediksi durasi penanganan gangguan adalah data luaran model prediksi jumlah laporan posko. Nilai korelasi antara variabel prediktor dan variabel target menggunakan teknik korelasi pearson ditunjukkan pada Tabel 9.

Tabel 9. Korelasi Variabel Target Durasi.

VARIABEL FITUR	SATUAN	DURASI MTG	DURASI PNK	DURASI SGM	KRITERIA KORELASI MTG	KRITERIA KORELASI PNK	KRITERIA KORELASI SGM
BULAN	Month	-0,009	0,029	0,081	Sangat Lemah	Sangat Lemah	Sangat Lemah
KALI PADAM	kali	0,579	0,643	0,604	Sedang	Kuat	Kuat
LEVEL CUACA	mm	0,783	0,774	0,774	Kuat	Kuat	Kuat
PREDIKSI JUMLAH LAPORAN MTG	kali	0,732	-	-	Kuat	-	-
PREDIKSI JUMLAH LAPORAN PNK	kali	-	0,753	-	-	Kuat	-

<b>PREDIKSI JUMLAH LAPORAN SGM</b>	<i>kali</i>	-	-	0,738	-	-	Kuat
--	-------------	---	---	-------	---	---	------

Seperti model sebelumnya, dilakukan normalisasi dataset variabel prediktor dan variabel respon dengan metode *Min-Max Normalization* sehingga menghasilkan data sebagai berikut.

Tabel 10. Normalisasi Data Prediksi Durasi Penanganan Gangguan.

Variabel	Data Mentah			Data Normalisasi		
	Min	Mean	Max	Min	Mean	Max
<b>Bulan</b>	1	6,55	12	1	6,51	12
<b>Kali Padam</b>	0,01	0,31	1	0,01	0,31	1,00
<b>Level Cuaca</b>	0,00	15,71	150	0,00	15,71	150
<b>Prediksi Laporan MTG</b>	0,03	0,07	0,19	0,01	0,27	1
<b>Prediksi Laporan PNK</b>	0,06	0,17	0,48	0,01	0,28	1
<b>Prediksi Laporan SGM</b>	0,01	0,12	0,04	0,01	0,19	1
<b>Durasi MTG</b>	0,70	1,24	3,45	0,01	0,21	1
<b>Durasi PNK</b>	0,90	1,42	4,59	0,01	0,15	1
<b>Durasi SGM</b>	1,10	1,61	3,88	0,01	0,19	1

Selanjutnya melakukan proses pembelajaran model prediksi durasi penanganan gangguan pelanggan menggunakan partisi data masukkan sebesar 70% sebagai data latih. Pemilihan data latih model prediksi menggunakan metode *linear sampling*. Variabel prediktor pada model ini adalah Level Cuaca, dan Prediksi Jumlah Laporan dengan variabel target adalah Durasi Penanganan

Tabel 11. Statistik Model Pembelajaran Durasi Posko Sungguminasa.

Statistik Regresi Durasi Sungguminasa				
Variable	Coeff.	Std. Err.	t-value	P> t
<b>LEVEL_CUACA</b>	0,0032	0,0006	55.508	1,25E-07
<b>Prediction (JUMLAH LAPORAN SGM)</b>	0,2015	0,0944	21.344	0,0344
<b>Intercept</b>	0,0809	0,0209	38.651	0,0002

Tabel 12. Statistik Model Pembelajaran Durasi Posko Panakkukang.

Statistik Regresi Durasi Panakkukang				
Variable	Coeff.	Std. Err.	t-value	P> t
<b>LEVEL_CUACA</b>	0,0018	0,0005	39.806	0,0001
<b>Prediction (JUMLAH LAPORAN PNK)</b>	0,2862	0,0769	37.235	0,0003
<b>Intercept</b>	0,0438	0,0168	26.056	0,0101

Laporan Gangguan masing-masing posko. Hasil model pembelajaran prediksi durasi penanganan gangguan adalah sebagai berikut.

Tabel 13. Statistik Model Pembelajaran Durasi Posko Mattoangin.

Statistik Regresi Durasi Mattoangin				
Variable	Coeff.	Std. Err.	t-value	P> t
LEVEL_CUACA	0.0041	0.0006	66.081	6.28E-10
Prediction (JUMLAH LAPORAN MTG)	0.067	0.1023	0.6544	0.5138
Intercept	0.1232	0.0211	5.848	2.97E-8

Tabel 14. Nilai R-squared dan Koefisien Variabel Prediksi Durasi.

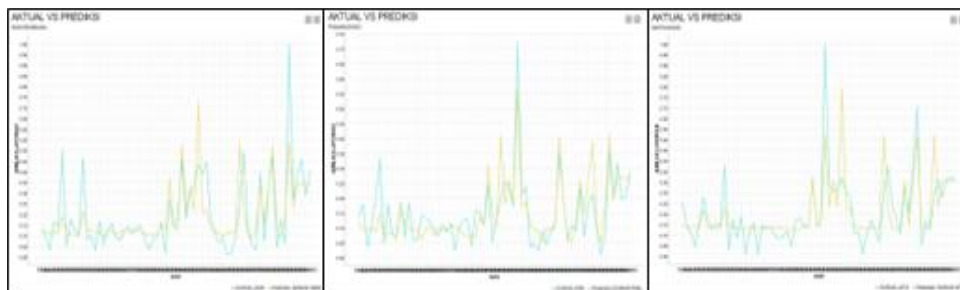
POSKO	R <sup>2</sup>	$\alpha$	b1	b2	Keterangan
Mattoangin	0,6427	0,0809	0,0032	0,2015	Kemampuan variabel prediktor dalam menjelaskan variansi variabel target adalah sebesar 64,3%, dan masih ada 35,7% variansi variabel target tidak mampu dijelaskan variabel prediktor
Panakkukang	0,6214	0,0438	0,0018	0,2862	Kemampuan variabel prediktor dalam menjelaskan variansi variabel target adalah sebesar 62,1%, dan masih ada 37,9% variansi variabel target tidak mampu dijelaskan variabel prediktor
Sungguminasa	0,6527	-0,1232	0,0041	0,067	Kemampuan variabel prediktor dalam menjelaskan variansi variabel target adalah sebesar 65,3%, dan masih ada 34,7% variansi variabel target tidak mampu dijelaskan variabel prediktor

Data luaran model pembelajaran dilakukan pengujian menggunakan 30% data populasi untuk memperoleh nilai prediksi Durasi Penanganan Laporan Gangguan Pelanggan. Nilai akurasi dari luaran hasil uji prediksi diukur

menggunakan parameter *r-squared* ( $R^2$ ) dan *Root Mean Squared Error* (RMSE)

Tabel 15. Statistik Hasil Uji Regresi Durasi Penanganan.

Statistik Hasil Uji Regresi Durasi Penanganan Gangguan		
POSKO	$R^2$	RMSE
Mattoangin	0,519	0,112
Panakkukang	0,613	0,068
Sungguminasa	0,547	0,117



Gambar 3. *Line Plot* Data Aktual dan Prediksi Jumlah Laporan.

sesuai .

Nilai *r-squared* luaran model terbaik di posko Panakkukang dengan nilai 61,3% sehingga disimpulkan bahwa variabel prediktor menjelaskan variansi variabel target sebesar 61,3% dan 39,7% variansi target tidak dapat dijelaskan oleh variabel prediktor. Nilai RMSE terbaik di posko Panakkukang dengan nilai 0,068 diikuti posko Mattoangin dengan nilai 0,112 dan posko Sungguminasa sebesar 0,117. Disimpulkan bahwa kesalahan hasil uji mendekati nol dan akurasi hasil prediksi memiliki nilai sangat baik dengan akurasi di atas 90%.

### C. Klasifikasi Target Kinerja Durasi Penanganan Gangguan

Uji klasifikasi dilakukan untuk mengetahui variabel yang mempengaruhi durasi penanganan gangguan pelanggan memenuhi target atau tidak memenuhi target. Luaran model prediksi durasi penanganan gangguan menjadi salah satu variabel masukkan untuk model klasifikasi. Proses normalisasi dilakukan untuk variabel masukkan model klasifikasi dan teknik *smote* digunakan untuk menyeimbangkan data klasifikasi yang *unbalance*. Dengan membagi 70% sebagai data latih dan 30% sebagai data uji, dua metode klasifikasi yaitu *Decision Tree* dan *Naïve Bayes* digunakan agar dapat dibandingkan hasil pemodelannya. Hasil luaran model ditunjukkan pada Tabel 4 berikut ini.

Tabel 16. Statistik Hasil Uji Klasifikasi.

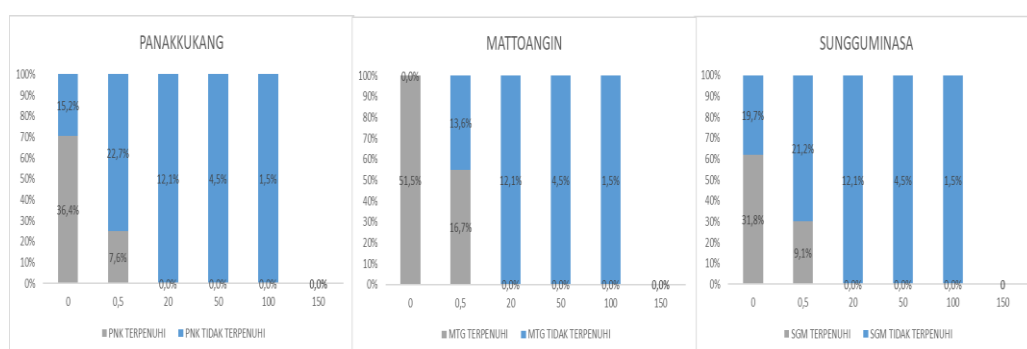
VARIABEL	DECISION TREE			NAÏVE BAYES		
	ACCURACY	ROC CURVE	COHEN'S KAPPA (k)	ACCURACY	ROC CURVE	COHEN'S KAPPA (k)
POSKO PANAKKUKANG	0,80	0,92	0,59	0,60	0,82	0,29

<b>POSKO SUNGGUMINASA</b>	0,90	0,87	0,73	0,85	0,94	0,58
<b>POSKO MATTOANGIN</b>	0,90	0,99	0,79	0,80	0,97	0,55

Hasil model, metode klasifikasi *Decision Tree* memberikan nilai akurasi yang lebih baik pada pengujian seluruh posko dibandingkan metode klasifikasi *Naïve Bayes*.

#### D. Pembahasan Hasil Model

Regresi linier berganda dilakukan untuk mendapatkan nilai prediksi jumlah laporan gangguan pelanggan dan durasi penanganan gangguan. Keunggulan menggunakan regresi linier berganda yaitu mampu melakukan generasilisasi dan pemecahan pola data tertentu, mampu melakukan akuisisi informasi meskipun tidak memiliki kepastian, dan mampu menjalankan analisis perhitungan secara



Gambar 4. Sebaran Klasifikasi Durasi Penanganan Berdasarkan Level Cuaca.

paralel sehingga proses menjadi lebih cepat. Pemilihan variabel prediktor menggunakan uji korelasi dengan pengaruh kategori sedang dan kuat memberikan nilai *r-squared* mencapai 69,7% atau kategori kuat untuk prediksi jumlah laporan, dan nilai *r-squared* 61,3% atau kategori sedang untuk prediksi durasi penanganan laporan gangguan.

Selanjutnya, klasifikasi dilakukan untuk mengetahui apakah prediksi dimasa depan durasi penanganan laporan gangguan memenuhi target kinerja atau tidak pada satu hari dengan kali padam gangguan dan kondisi cuaca tertentu serta jumlah laporan gangguan pelanggan yang telah diprediksikan. Menggunakan dua metode klasifikasi yang dibandingkan untuk mengukur hasil prediksi yaitu *Decision Tree* dan *Naïve Bayes*, hasil metode *Decision Tree* memberikan akurasi yang lebih baik dengan nilai akurasi mencapai 90% dibandingkan 85% pada metode *Naïve Bayes*.

Dengan akurasi hasil prediksi tersebut maka kemungkinan terjadinya target terpenuhi durasi penanganan laporan gangguan pelanggan sebesar 51% atau setara dengan 560 hari dan tidak terpenuhi target durasi penanganan laporan gangguan sebesar 49% atau setara dengan 538 hari. Semakin tinggi level cuaca, maka durasi penanganan gangguan didominasi klasifikasi target tidak terpenuhi seperti yang digambarkan pada Gambar 4.

## KESIMPULAN

Pembentukan model prediksi laporan gangguan pelanggan dan durasi penanganan laporan gangguan melalui kombinasi Regresi Linear Berganda dan Klasifikasi *Decision Tree* terbukti dapat dijalankan dan memberikan hasil akurasi prediksi jumlah laporan sebesar 69,7% dan prediksi durasi penanganan gangguan sebesar 61,73%. Hal ini tentu dengan melaksanakan tahapan pra pemrosesan data melalui uji korelasi untuk pemilihan variabel masukkan model, missing value, normalisasi *Min-Max*, dan pembagian data latih dan data uji yang optimal. Luaran model prediksi jumlah laporan sebagai variabel masukkan model prediksi durasi penanganan gangguan yang selanjutnya dilakukan klasifikasi untuk mengetahui klasifikasi variabel-variabel yang memenuhi atau tidak memenuhi durasi penanganan gangguan. Metode klasifikasi *Decision Tree* merupakan metode klasifikasi yang lebih unggul jika dibandingkan dengan metode klasifikasi *Naïve Bayes* melalui kurva ROC memiliki tingkat akurasi yang lebih baik.

## SARAN

Untuk penelitian selanjutnya peningkatan akurasi hasil regresi dan klasifikasi dapat dilakukan menggunakan metode yang lainnya. Untuk metode prediksi regresi dapat menggunakan metode Regresi Poisson, LSTM, atau ARIMA. Sedangkan metode klasifikasi dapat menggunakan metode klasifikasi Random Forest, Support Vector Machine, atau Gradient Boosted.

## DAFTAR RUJUKAN

- Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. *Computing*, 104(2), 271–294. <https://doi.org/10.1007/s00607-021-00908-y>
- Narulita, S., Oktaga, A. T., & Susanti, I. (2021). Pengujian Akurasi Model Prediksi Menggunakan Metode Data Mining Classification Decision Tree Algoritma C4.5. *Jurnal Media Aplikom*, 13(2), 15–29.
- Oktafia, D., & Pardede, D. D. L. C. (2008). *Perbandingan Kinerja Algoritma Decision Tree Dan Naive Bayes Dalam Memprediksi Kebangkrutan*. 2008, 2008.
- Pangesti, S. (2016). *SATS4312: Modul 1 Regresi Linear Sederhana*. 52.
- Rokach, L., & Maimon, O. (2005). Decision Trees. *Cycle*, 1897(Figure 1), 44–45. <https://doi.org/10.1007/0-387-25465-X>
- Rosandy, T. (2016). Perbandingan Metode Naive Bayes Classifier Dengan Metode Decision Tree (C4.5) Untuk Menganalisa Kelancaran Pembiayaan (Study Kasus: KSPPS / BMT AL-FADHILA). *Jurnal Teknologi Informasi Magister Darmajaya*, 2(01), 52–62.
- Safitri, W, R. (2014). Analisis Korelasi Pearson Dalam Menentukan Hubungan Antara Kejadian Demam Berdarah Dengue Dengan Kepadatan Penduduk Di Kota Surabaya Pada Tahun 2012 - 2014. *Jurnal Kesehatan Masyarakat*, 1(3), 1–9.
- Sutoyo, E., & Fadlurrahman, M. A. (2020). Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network. *Jurnal Edukasi Dan*

- Anditya, A. et al., 2020. Developing an LSTM-based Classification Model of IndiHome Customer Feedbacks. *IEEE*.
- Bahrawi, A., 2016. Prediksi Keputusan Menggunakan Klasifikasi. *Jurnal Kominfo*, Volume 20, pp. 1-10.
- Francisco V, O. & Rosaria, S., 2021. Machine learning for marketing on the KNIME Hub The development of a Machine learning for marketing on the KNIME Hub. *Elsevier*, Volume 137, pp. 393-410.
- Géron, A., 2019. *Hands-on Machine Learning with Scikit-Learn, Keras&TensorFlow*. s.l.:O'Reilly.
- Ghazzawi, A. & Alharbi, B., 2019. Analysis of Customer Complaints Data using Data Mining Techniques. *Elsevier*, Volume 163, pp. 62-69.
- Irfan, U. et al., 2019. A Churn Prediction Model using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE*, Volume 7, pp. 60134 - 60149.
- Jiahan, D. et al., 2020. Power User Sensitivity Analysis and Power Outage Complaint Prediction. *IOP*.
- Ning, L., Hua, L., Jie, L. & Guangquan, Z., 2011. A Customer Churn Prediction Model in Telecom Industry Using Boosting. *IEEE*.
- Putra, J. W. G., 2020. *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning*. 1.4 ed. Tokyo: s.n.
- S Ummugulthum, N. & S, B., 2014. Customer Relationship Management Classification Using Data Mining Techniques. *IEEE*.
- Shuangshuang, G., Linlin, T., Xiaoyan, G. & Zheng, H., 2020. Power Customer Complaint Prediction Model Based on Time Series Analysis. *IIETA*, 34(4), pp. 471-477.
- Suzanne, E., 2012. ROC CurvesVWhat are They and How are They Used?. *Point of Care*, 11(1), pp. 16-21.
- Xin, X. et al., 2019. A Comparative Study of Customer Complaint Prediction Model of Time Series, Multiple Linear Regression and BP Neural Network. *IOP*.