

FORECASTING THROUGHPUT CAPACITY ON 5GHZ WIRELESS RADIO NETWORK USING LINEAR REGRESSION METHOD

PERAMALAN KAPASITAS THROUGHPUT PADA JARINGAN RADIO WIRELESS 5GHZ DENGAN METODE REGRESI LINIER

Muhammad Rizky Pratama¹, Armansyah²
^{1,2} Universitas Islam Negeri Sumatera Utara
prizky428@gmail.com¹, armansyah@uinsu.ac.id²

Abstract - In an internet network, internet service providers must ensure the quality of the devices used is in good condition. One of the devices used is a 5GHz wireless radio where the stability of the fluctuating throughput capacity affects the quality of the device as a network transmission. This research uses machine learning with a simple linear regression model to predict the final throughput capacity based on the initial throughput and optimize it in the future. The regression model is used because it has a good accuracy level in forecasting. The data used is time series throughput data from the Ubnt Litebeam 5AC Gen.2 5GHz wireless radio device for the period of January-July 2024, using data split technique, data from January-June is used as training data and data from July as test data. The forecasting results obtained indicate the model's error rate measured using Mean Square Error (MSE) of 184,682 and Root Mean Squared Error (RMSE) of 13,589. From the forecasting results obtained, it can be used as a reference to evaluate the feasibility of wireless radio devices used by internet service provider for the future.

Keywords – *Forecasting, Throughput, Radio Wireless, Simple Linear Regression.*

Abstrak – Pada sebuah jaringan internet, penyedia layanan internet harus memastikan kualitas perangkat yang digunakan dalam keadaan baik. Salah satu perangkat yang digunakan yaitu radio *wireless* 5ghz dimana kestabilan kapasitas *throughput* yang naik turun tidak menentu mempengaruhi kualitas perangkat sebagai transmisi jaringan. Penelitian ini menggunakan *machine learning* dengan model regresi linier sederhana untuk meramalkan kapasitas *throughput* akhir berdasarkan *throughput* awal dan mengoptimalkannya di masa mendatang. Model regresi digunakan karena memiliki tingkat akurasi yang baik dalam melakukan peramalan. Data yang digunakan merupakan data *time series throughput* perangkat radio *wireless* Ubnt Litebeam 5AC Gen.2 5Ghz periode Januari-Juli 2024, menggunakan teknik split data, data periode Januari-Juni digunakan sebagai data latih dan data periode Juli sebagai data uji. Hasil peramalan yang didapat mengindikasikan tingkat kesalahan model diukur menggunakan Mean Square Error (MSE) sebesar 184,682 dan Root Mean Squared Error (RMSE) sebesar 13,589. Dari hasil peramalan yang didapat, bisa dijadikan acuan untuk mengevaluasi layak tidaknya perangkat radio wireless yang digunakan perusahaan penyedia layanan internet untuk kedepannya.

Kata Kunci – Peramalan, *Throughput*, Radio *Wireless*, Regresi Linier Sederhana.

I. PENDAHULUAN

Perusahaan penyedia layanan internet pada saat ini umumnya menggunakan dua jenis alat transmisi jaringan, yaitu *fiber optic* dan radio *wireless*. Radio *wireless* dengan frekuensi 5 Ghz biasa digunakan penyedia layanan internet untuk mentransmisikan jaringan dari *access point* ke *client* yang jaraknya lebih dari 1km. Salah satu tantangan yang sering dihadapi banyak penyedia layanan internet yaitu kestabilan jaringan yang digunakan. Kestabilan jaringan yang dimaksud dapat dilihat berdasarkan jumlah *throughput* jaringan yang sampai pada *client*. *Throughput* merupakan ukuran efektivitas transfer data yang dinyatakan dalam satuan bit per detik (BPS). Penghitungan *throughput* dilakukan dengan membagi jumlah total paket data yang berhasil diterima di tujuan selama periode tertentu dengan durasi waktu pengamatan. Konsep *throughput* juga dapat diterapkan untuk mengukur kinerja *transfer file* dalam berbagai kondisi jaringan [1]. *Throughput* disebut sebagai indikator utama dari kualitas jaringan, yang kapasitasnya dipengaruhi oleh frekuensi yang digunakan, interferensi antar frekuensi, dan *bandwith* yang disediakan. Untuk mencapai kapasitas *throughput* yang dikirimkan maksimal, kekuatan sinyal (*signal strength*) antar perangkat yang terhubung juga harus maksimal. Kualitas *signal strength* itu pula dipengaruhi oleh spesifikasi perangkat yang digunakan, penempatan perangkat serta *pointing* antar perangkat radio *wireless* yang terkoneksi [2]. Kapasitas *throughput* pada jaringan radio *wireless* ditentukan oleh tingkat *throughput* yang diperoleh dari kemacetan jaringan, sehingga perlu dipastikan bahwa kemacetan tersebut berfungsi secara maksimal agar keseluruhan *throughput* bisa terkirim dengan maksimal pula [3]. Dikarenakan kapasitas *throughput* yang naik turun tidak menentu, maka perlu dilakukannya peramalan kapasitas *throughput* akhir yang terkirim pada *client* agar hal-hal yang mempengaruhi kapasitas *throughput* dapat dioptimalkan kedepannya.

Peramalan atau *forecasting* adalah teknik untuk memproyeksikan kebutuhan masa depan, mencakup aspek kuantitas, kualitas, waktu, dan lokasi, guna memenuhi permintaan barang atau jasa. Di antara berbagai metode peramalan yang ada, Regresi Linier merupakan salah satu pendekatan yang umum digunakan untuk mengembangkan model prediktif. Regresi linier sederhana merupakan teknik analisis yang mengkaji hubungan antara satu variabel independen dan satu variabel dependen yang memiliki keterkaitan linier. Metode ini termasuk dalam kategori analisis *time series* kuantitatif, di mana waktu menjadi landasan utama dalam melakukan prediksi dan peramalan. Dalam konteks ini, model peramalan yang diaplikasikan adalah model sebab-akibat, yang mempertimbangkan pengaruh variabel-variabel tertentu terhadap variabel yang sedang diramalkan [4]. Pemilihan metode regresi linier sederhana didasarkan pada kebutuhan peramalan jangka menengah, yang umumnya mencakup periode bulanan dengan rentang waktu kurang dari satu tahun. Penelitian ini menggunakan pendekatan kuantitatif yang datanya diambil dari data historis berupa *time series* bulanan pada kapasitas *throughput client* pada jaringan radio *wireless* 5 Ghz [5]. Dengan dataset yang akan digunakan mengacu pada parameter *throughput rx* dari perangkat *access point* Ubiquiti Litebeam 5AC Gen 2 5 Ghz. Lokasi penelitian dilakukan di salah satu perusahaan IT Infrastruktur, yaitu PT. Zathco. Penelitian ini menggunakan dataset *throughput* awal dan *throughput* akhir tertinggi periode Januari - Juli 2024 untuk memprediksi *throughput* akhir *client* radio *wireless* 5 Ghz periode Agustus 2024. Tujuan dilakukan peramalan *throughput* ini nantinya dari hasil peramalan yang dilakukan, dapat dijadikan acuan oleh perusahaan penyedia layanan internet terkait untuk meningkatkan kualitas dan performa perangkat radio *wireless* yang digunakan, serta mengoptimalkan variabel yang mempengaruhi kapasitas *throughput* yang dikirimkan ke *client*.

Penelitian mengenai *throughput* sebelumnya dilakukan oleh [2] melakukan analisis kinerja penerapan standard protokol keamanan IEEE 802.11 pada layanan Wi-Fi yang dapat mempengaruhi kapasitas *throughput* yang dihasilkan *access point* frekuensi 2,4 Ghz dan 5 Ghz. Penelitian lainnya yang dilakukan oleh [6] yang melakukan analisis perbandingan kinerja jaringan pada proses tethering untuk membandingkan kapasitas *throughput* yang dihasilkan oleh jaringan WLAN 2,4 Ghz dan 5 Ghz di pedesaan dan di perkotaan. Sedangkan untuk penelitian regresi linier sederhana menggunakan data historis, dilakukan oleh [7] yang menggunakan model regresi untuk memprediksi persediaan obat jenis tablet yang memperoleh akurasi sebesar 98,505%.

II. SIGNIFIKANSI STUDI

A. *Machine Learning* dan *Supervised Learning*

Machine learning, sebagai subbidang kecerdasan buatan (*artificial intelligence*), merupakan pendekatan pemrograman yang bertujuan menciptakan komputer dengan kecerdasan menyerupai manusia. Sistem ini dirancang untuk meningkatkan pemahaman secara otomatis melalui pengalaman. Fokus utama *machine learning* adalah pengembangan sistem yang dapat belajar mandiri dan membuat keputusan tanpa memerlukan pemrograman berulang oleh manusia, yang artinya sistem dapat beradaptasi terhadap perubahan yang terjadi [8]. *Supervised learning* merupakan salah satu teknik pembelajaran dalam *machine learning* yang memanfaatkan dataset berlabel (*labeled data*) sebagai data *training* untuk melatih mesin. Dengan pendekatan ini, mesin dapat mengidentifikasi label *input* berdasarkan fitur-fitur yang ada, yang kemudian digunakan untuk melakukan prediksi, peramalan, atau klasifikasi [9]. Dalam konteks tertentu, *supervised learning* diaplikasikan pada masalah regresi, khususnya ketika variabel *output* berupa nilai riil, seperti pada penelitian kali ini yang nantinya *output* dari model regresi yang diramalkan berupa nilai kapasitas *throughput* akhir. Penulis membuat dan mengembangkan model regresi nantinya menggunakan bahasa pemrograman Python, dan menggunakan Jupyter Notebook sebagai teks editor yang dirancang khusus untuk memudahkan *programmer* dalam membuat program dengan efisiensi waktu, kemudahan pengembangan, dan kompatibilitas dengan sistem. Selain itu, Python juga merupakan bahasa pemrograman paling populer yang banyak digunakan oleh *programmer* untuk *Machine Learning* dan *Artificial Intelligence* [10]. Popularitas python ini salah satunya dikarenakan tersedianya banyak *library* (modul-modul siap untuk digunakan) yang mendukung dalam bidang *Machine Learning*. *Library* yang penulis gunakan dalam membangun model regresi ini yaitu Pandas, NumPy, Scikit-learn, Matplotlib, dan Seaborn.

B. *Regresi Linier*

Regresi linier merupakan metode yang menghubungkan satu atau lebih variabel independen (X) dengan variabel respon (Y). Metode ini efektif untuk meramalkan nilai suatu variabel berdasarkan variabel lainnya dan memberikan prediksi yang lebih akurat [11]. Sebagai alat analisis, regresi linier memudahkan pemahaman hubungan antara variabel-variabel tersebut, sehingga menghasilkan model peramalan yang lebih tepat.[12]. Persamaan dari regresi linier sederhana dapat dilihat sebagai berikut:

$$Y = a + bX \quad (1)$$

di mana Y adalah variabel terikat (respon), X adalah variabel bebas (prediktor), a adalah *intercept* (titik potong dengan sumbu Y), dan b adalah koefisien regresi yang menggambarkan tingkat perubahan Y untuk setiap perubahan satu unit X . Dalam penelitian ini penulis menggunakan metode regresi linier sederhana, menggunakan data *throughput* awal dan

throughput akhir periode Januari-Juli 2024 untuk meramalkan throughput akhir periode Agustus 2024. Kedua variabel ini dipilih untuk membangun model regresi daripada variabel lainnya dikarenakan memiliki korelasi yang tinggi, dimana kapasitas *throughput* akhir yang dikirim bergantung dengan berapa kapasitas *throughput* awalnya, dan nilai *throughput* akhir sendiri pastinya tidak akan melebihi dari nilai *throughput* awalnya. Dan dikarenakan hanya dua variabel ini yang memiliki korelasi yang kuat pula, data yang tersedia cocok menggunakan metode regresi linier sederhana dan tidak mendukung untuk menggunakan regresi linier berganda.

C. Peramalan

Peramalan adalah suatu proses analitis yang bertujuan untuk memproyeksikan kejadian atau hasil di masa mendatang dengan memanfaatkan data historis. Proses ini melibatkan penggunaan model matematis untuk menghitung probabilitas kejadian di masa depan berdasarkan pola-pola yang teridentifikasi dari data masa lalu [13]. Pada penelitian ini, pendekatan peramalan *throughput* menggunakan metode kuantitatif dengan dataset berupa *time series*. Dataset *time series* terdiri dari serangkaian observasi yang dicatat secara berurutan dalam interval waktu tertentu, yang bisa bervariasi dari skala harian, mingguan, bulanan, hingga tahunan. Untuk itu dalam penelitian ini penulis menggunakan data *time series* periode Januari-Juli 2024 dari perangkat radio wireless 5 Ghz untuk dijadikan model regresi yang nantinya akan diramalkan *throughput* akhir periode Agustus 2024 dari perangkat tersebut.

D. Radio Wireless

Radio *wireless* merupakan alat jaringan yang digunakan untuk menyalurkan data antara dua jaringan dari jarak yang jauh dan biasanya sulit dijangkau dengan media jaringan kabel [14]. Radio *wireless* atau *Wireless Local Area Network* (WLAN) dalam dunia penyedia layanan internet (ISP) sampai saat ini masih banyak digunakan dikarenakan penggunaan radio *wireless* lebih efisien dalam harga daripada penggunaan kabel fiber, disamping instalasinya yang juga lebih mudah . Radio *Wireless* memiliki beberapa frekuensi dalam penangkapan sinyalnya, untuk penelitian ini penulis mengambil sampel data penggunaan radio *wireless* untuk jarak jauh (diatas 1km) digunakan radio *wireless* dengan frekuensi 5 Ghz [15].

E. Throughput

Throughput merupakan *bandwidth* yang sesungguhnya, yang terukur pada suatu ukuran waktu tertentu dalam mentransmisikan berkas pada sebuah jaringan, baik itu jaringan dengan kabel maupun tanpa kabel (*wireless*). [2]. Berbeda dengan *bandwidth* walaupun satuannya sama *bit per second* (bps), tetapi *throughput* lebih menggambarkan *bandwidth* sebuah jaringan yang sebenarnya pada suatu waktu, pada kondisi dan jaringan tertentu selama interval waktu tertentu dibagi oleh durasi interval waktu tersebut pula [16]. Untuk lebih jelasnya nilai *throughput* dapat dihitung menggunakan rumus sebagai berikut:

$$\text{Throughput} = \frac{\text{jumlah data yang dikirim/bandwidth (bps)}}{\text{waktu pengiriman data (s)}} \quad (2)$$

III. HASIL DAN PEMBAHASAN

A. Presentasi dan Representasi Data

Pertama-tama sebelum membuat model regresi linier sederhana dari sebuah peramalan, terlebih dahulu kita mempresentasi dan merepresentasikan data awal atau data mentah yang akan kita lakukan penelitian. Data mentah terdiri dari 213 *record* data dimulai dari Januari sampai dengan Juli 2024, yang terdiri dari data nama perangkat radio *wireless access point* dan *client* yang digunakan, jarak antar perangkat (satuan km), tanggal dan waktu, kekuatan signal (satuan dBm (decibel-milliwatt)), frekuensi yang digunakan dalam koneksi antar perangkat, *throughput* awal dan *throughput* akhir (dalam satuan Mbps). Beberapa data mentah dapat dilihat pada tabel berikut:

TABEL I
DATA MENTAH

Access Point	Client	Jarak (km)	Tanggal	Jam	Signal Strength (dBm)	Frek.	Throughput Awal (Mbps)	Throughput Akhir (Mbps)
Ubnt Litebeam 5AC Gen 2 5GHZ	Ubnt Litebeam 5AC Gen 2 5GHZ	2,85	Senin, 01/01/2024	16:35:50	-65	5145	236,14	98,58
Ubnt Litebeam 5AC Gen 2 5GHZ	Ubnt Litebeam 5AC Gen 2 5GHZ	2,85	Selasa, 02/01/2024	15:45:07	-68	5145	221,03	93,73
Ubnt Litebeam 5AC Gen 2 5GHZ	Ubnt Litebeam 5AC Gen 2 5GHZ	2,85	Rabu, 03/01/2024	17:06:25	-71	5145	193,73	72,31
...
Ubnt Litebeam 5AC Gen 2 5GHZ	Ubnt Litebeam 5AC Gen 2 5GHZ	2,85	Senin, 29/07/2024	14:37:36	-65	5495	187,36	92,61
Ubnt Litebeam 5AC Gen 2 5GHZ	Ubnt Litebeam 5AC Gen 2 5GHZ	2,85	Selasa, 30/07/2024	16:19:21	-68	5495	155,94	68,96
Ubnt Litebeam 5AC Gen 2 5GHZ	Ubnt Litebeam 5AC Gen 2 5GHZ	2,85	Rabu, 31/07/2024	18:14:38	-67	5495	104,01	79,21

Setelah mengetahui data mentah yang akan digunakan, selanjutnya yaitu merepresentasikan data apa saja yang nantinya akan dilakukan penelitian dan dibuat menjadi model peramalan regresi linier sederhana. Variabel yang akan digunakan dalam membangun model regresi linier sederhana ini yaitu *Throughput* Awal dan *Throughput* Akhir dari *client*. Pemilihan kedua variabel ini dikarenakan kapasitas *throughput* akhir yang nilainya tidak akan jauh berbeda dan tidak akan melebihi nilai dari variabel *throughput* awal. Alasan ini menunjukkan korelasi yang kuat antara kedua variabel ini dan dijadikan landasan yang kuat mengapa kedua variabel ini digunakan sebagai acuan untuk membuat model regresi dalam penelitian ini.

B. Perhitungan Manual Menggunakan Regresi Linier Sederhana

Setelah mengetahui *field* data apa saja yang akan digunakan untuk membuat model regresi, tahap selanjutnya yaitu melakukan perhitungan manual menggunakan metode yang sudah ditetapkan yaitu regresi linier sederhana. Data yang akan digunakan sebagai data latih ialah data *Throughput* Awal dan *Throughput* Akhir pada perangkat periode Januari-Juni 2024. Sedangkan yang akan dijadikan sebagai data uji untuk meramalkan *Throughput* Akhir pada Agustus 2024 ialah data *Throughput* Awal periode Juli 2024.

1. Identifikasi Variabel dan Menghitung XY dan X²

Variabel yang dijadikan sebagai koefisien (faktor penyebab (X)) pada penelitian ini adalah *Throughput* Awal, sedangkan variabel yang akan dijadikan sebagai *intersect* (akibat nya (Y)) adalah *Throughput* Akhir. Berdasarkan data pada Tabel I, data yang akan dilatih dan dilakukan perhitungan manual yaitu data *Throughput* Awal dan *Throughput* Akhir periode Januari-Juni 2024 dengan total data berjumlah 182, maka diperoleh hasil perhitungan XY dan X² seperti pada tabel berikut:

TABEL II
MENGHITUNG XY DAN X²

Tanggal	Throughput Awal (X)	Throughput Akhir (Y)	XY	X ²
Senin, 01/01/2024	236,14	98,58	23.278,6812	55.762,10
Selasa, 02/01/2024	221,03	93,73	20.717,1419	48.854,26
Rabu, 03/01/2024	193,73	72,31	14.008,6163	37.531,31
...
Jum'at, 28/06/2024	153,74	89,99	13.835,0626	23.635,99
Sabtu, 29/06/2024	220,98	199,41	44.065,6218	48.832,16
Minggu, 30/06/2024	226,30	211,58	47.880,554	51.211,69
Total	Σ = 31.889,86	Σ = 18.167,4	Σ = 3.476.358,755	Σ = 6.155.397,59

2. Mencari Nilai Intersect (a)

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$a = \frac{(18.167,44)(6.155.397,59) - (31.889,86)(3.476.358,755)}{182(6.155.397,59) - (31.889,86)^2}$$

$$a = \frac{111.827.816.392,4696 - 110.860.594.006,7243}{1.120.282.361,38 - 1.016.963.170,8196}$$

$$a = \frac{967.222.385,7453}{103.319.190,5604}$$

a = 9,3615

3. Mencari Nilai Koefisien (b)

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{182(3.476.358,755) - (31.889,86)(18.167,44)}{182(6.155.397,59) - (31.889,86)^2}$$

$$b = \frac{632.697.293,41 - 579.357.118,1584}{1.120.282.361,38 - 1.016.963.170,8196}$$

$$b = \frac{53.340.175,2516}{103.319.190,5604}$$

$$b = 0,5162$$

4. Menentukan Persamaan Regresi

Setelah nilai *intersect* dan koefisiennya didapat, maka diperoleh persamaan regresi menggunakan rumus persamaan regresi linier dengan hasilnya sebagai berikut:

$$Y = a + bX$$

$$Y = 9,3615 + 0,5162X \tag{5}$$

5. Menghitung Peramalan

Setelah mendapatkan persamaan regresi seperti pada persamaan (5), kita gunakan persamaan tersebut untuk menghitung peramalan *Throughput* Akhir pada periode Agustus 2024, yaitu menggunakan data periode Juli 2024 sebagai data uji. Hasil peramalan dapat dilihat dalam tabel sebagai berikut:

TABEL III
HASIL PERAMALAN THROUGHPUT AKHIR PERIODE AGUSTUS 2024

Tanggal	Throughput Akhir (Hasil Peramalan(Y))	Tanggal	Throughput Akhir (Hasil Peramalan(Y))
Kamis, 01/08/2024	93,05	Sabtu, 17/08/2024	100,48
Jum'at, 02/08/2024	70,45	Minggu, 18/08/2024	140,98
Sabtu, 03/08/2024	137,06	Senin, 19/08/2024	78,35
Minggu, 04/08/2024	125,57	Selasa, 20/08/2024	112,69
Senin, 05/08/2024	96,93	Rabu, 21/08/2024	102,16
Selasa, 06/08/2024	126,11	Kamis, 22/08/2024	68,81
Rabu, 07/08/2024	115,69	Jum'at, 23/08/2024	63,18
Kamis, 08/08/2024	106,61	Sabtu, 24/08/2024	117,12
Jum'at, 09/08/2024	109,01	Minggu, 25/08/2024	109,38
Sabtu, 10/08/2024	97,81	Senin, 26/08/2024	97,48
Minggu, 11/08/2024	139,79	Selasa, 27/08/2024	72,38
Senin, 12/08/2024	73,51	Rabu, 28/08/2024	130,00
Selasa, 13/08/2024	135,07	Kamis, 29/08/2024	106,08
Rabu, 14/08/2024	118,34	Jum'at, 30/08/2024	89,86
Kamis, 15/08/2024	102,82	Sabtu, 31/08/2024	118,28
Jum'at, 16/08/2024	62,80		

C. Membangun Model Regresi Dengan Python

Langkah pertama sebelum membuat model regresi dari sebuah peramalan adalah *preprocessing* data, yang mencakup pembersihan data dari *outlier (Handling Outlier)*, dan *matrix correlation* dari data yang telah melalui *handling outlier*. Setelah melalui *preprocessing* data, barulah data digunakan untuk membuat model regresi dan dilakukan peramalan menggunakan model yang telah dibuat. Untuk langkah terakhir yang tak kalah penting yaitu evaluasi dari model regresi yang dibangun menggunakan MSE dan RMSE.

1. Preprocessing Data

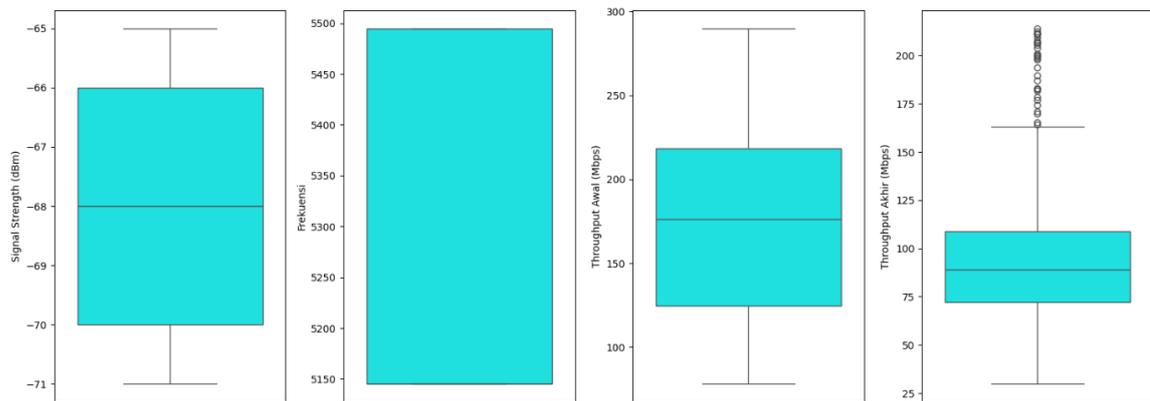
Data yang digunakan merupakan data yang sama seperti pada Tabel I, setelah mengetahui data mentah yang akan digunakan untuk membuat model, kita dapat mengetahui info mengenai data yang akan kita gunakan, mulai dari info tipe datanya, *null values*, *unique values*, *duplicated values* dan *unique sample* dari data mentah yang akan digunakan seperti pada gambar berikut:

```
[5]:
```

	column	datatype	null_values	unique_values	duplicated_values	unique_sample
0	Access Point	object	0	1	0	[Ubnt Litebeam 5AC Gen 2 5GHZ]
1	Client	object	0	1	0	[Ubnt Litebeam 5AC Gen 2 5GHZ]
2	Jarak	object	0	1	0	[2,85 km]
3	Tanggal	object	0	213	0	[Senin, 01/01/2024, Selasa, 02/01/2024, Rabu, ...]
4	Jam	object	0	213	0	[16:35:50, 15:45:07, 17:06:25, 18:49:09, 17:32...
5	Signal Strength (dBm)	int64	0	7	0	[-65, -68, -71, -67, -69]
6	Frekuensi	int64	0	2	0	[5145, 5495]
7	Throughput Awal (Mbps)	float64	0	210	0	[236.14, 221.03, 193.73, 167.58, 147.66]
8	Throughput Akhir (Mbps)	float64	0	210	0	[98.58, 93.73, 72.31, 60.59, 84.34]

Gambar 1. Info Data Mentah

Pada Gambar 1 dapat dilihat *features* yang termasuk *numeric* atau angka ada 4 yaitu Signal Strength (dBm), Frekuensi, *Throughput* Awal (Mbps) dan *Throughput* Akhir (Mbps). Dari keempat *numeric features* tersebut, kita coba tampilkan outlier yang ada pada data seperti pada gambar berikut:



Gambar 2. Outlier Data

Pada Gambar 2 terlihat bahwa *outlier* pada data banyak terdapat pada *features Throughput Akhir*. Untuk mengidentifikasi jumlah *outlier* pada *Throughput Akhir*, penulis menggunakan metode IQR untuk mengetahui nilai batas bawah (*lower bound*) dan batas atas (*upper bound*) pada *outlier*. Selanjutnya diketahui *value* berapa saja yang termasuk dalam *outlier* dan mengetahui jumlah *outlier* pada data. Proses *handling outlier* dapat dilihat pada gambar berikut:

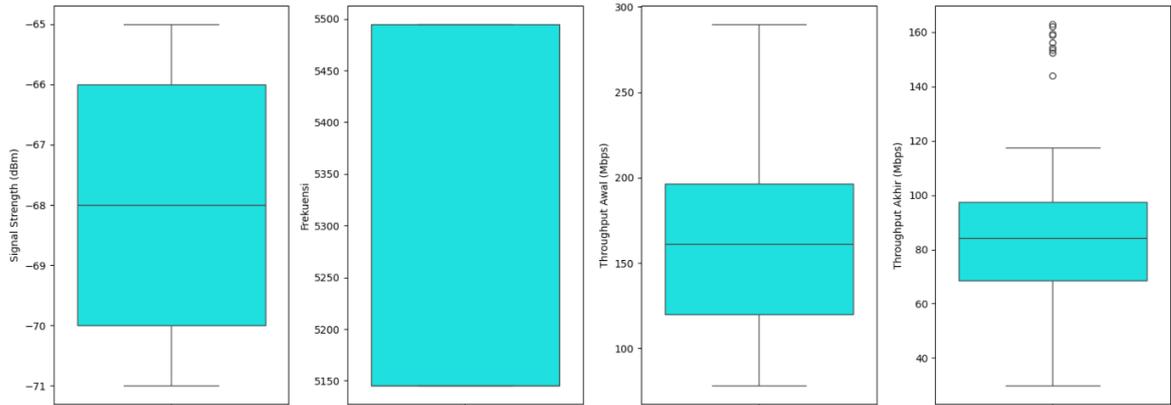
```
[8]:
```

	Q1	Q3	IQR	Lower Bound	Upper Bound	Outliers
Signal Strength (dBm)	-70.00	-66.00	4.00	-76.000	-60.000	[]
Frekuensi	5145.00	5495.00	350.00	4620.000	6020.000	[]
Throughput Awal (Mbps)	124.51	218.58	94.07	-16.595	359.685	[]
Throughput Akhir (Mbps)	72.08	108.71	36.63	17.135	163.655	[200.18, 205.44, 204.07, 178.42, 182.81, 211.2...

Gambar 3. Metode IQR

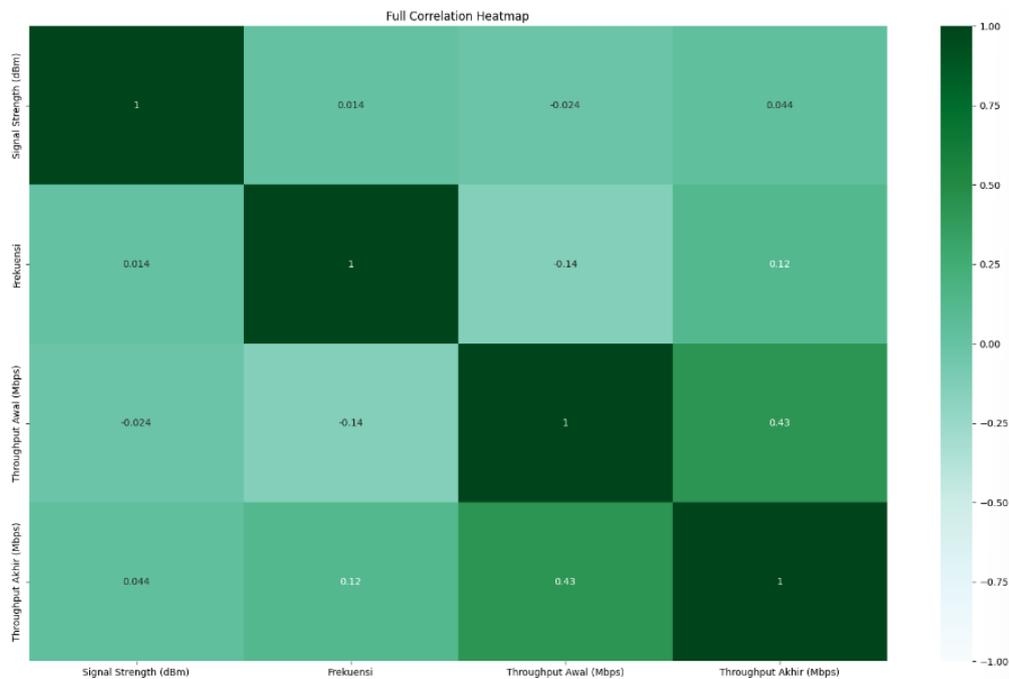
```
[9]: df.loc[(df['Throughput Akhir (Mbps)'] < 17.135) | (df['Throughput Akhir (Mbps)'] > 163.655)].shape[0]
[9]: 33
```

Gambar 4. Jumlah Data Outlier



Gambar 5. Handling Outlier Setelah Hapus Outlier Data

Pada Gambar 4 terlihat data outlier ada 33 data dan nantinya data yang tersisa setelah dibersihkan ada 180 record data. Penghapusan outlier ini berdampak pada nilai sebaran data dalam *matrix correlation* yang dapat dilihat pada gambar berikut:



Gambar 6. Matrix Correlation Features

Pada Gambar 6 terlihat sebaran data pada *numeric features* yang mendekati 1 selain *Throughput Akhir* ialah *Throughput Awal*, dimana kedua *features* ini nantinya akan diidentifikasi sebagai nilai X dan Y untuk membangun model regresi.

2. Model Regresi

Membangun model regresi menggunakan *features Throughput Awal* sebagai nilai X dan *Throughput Akhir* sebagai nilai Y dengan data seperti pada gambar berikut:


```
[79]: model.coef_
[80]: model.intercept_
[79]: array([[0.23550211]])
[80]: array([[44.6692681]])
```

(a) (b)

Gambar 9. (a) Nilai Koefisien (b) Nilai Intersect

Setelah diketahui nilai koefisien dan intersect dari data yang diuji, maka model regresi yang didapatkan dapat dilihat pada persamaan berikut:

$$Y = a + bX$$

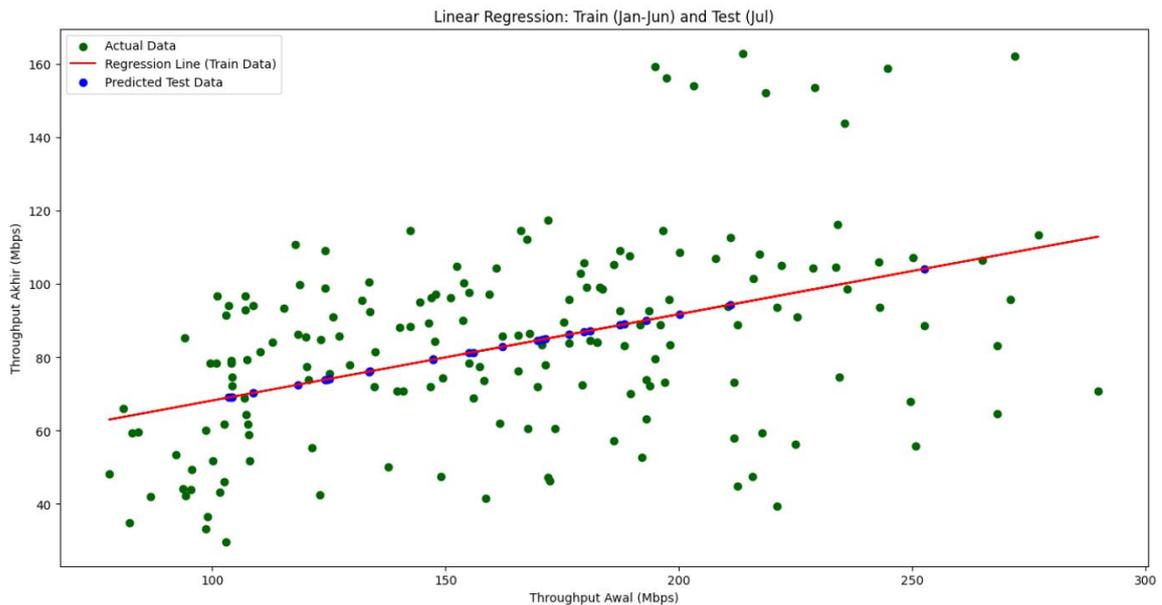
$$Y = 44,6692681 + 0,23550211X \tag{6}$$

Dengan didapatnya model regresi seperti pada persamaan 6, maka kita dapat melakukan peramalan terhadap nilai Y, yaitu *Throughput* Akhir periode Agustus 2024 berdasarkan data uji pada periode Juli 2024. Didapatkan hasil peramalan *Throughput* Akhir *client* periode 2024 seperti pada gambar berikut:

```
[20]: array([ 82.85122501,  72.53858766,  79.3775689 ,  74.13529196,
            84.62220087,  89.03786541,  90.13059519,  85.02255445,
            104.17594096,  73.93511517,  94.38847332,  87.30927993,
            69.04844641,  86.24010035,  81.16502991,  76.14412495,
            91.80972523,  87.00548221,  69.22271797,  70.28012244,
            76.16767516,  84.8718331 ,  88.79294321,  81.39346695,
            69.16384244])
```

Gambar 10. Hasil Peramalan *Throughput* Akhir *Client* Periode Agustus 2024

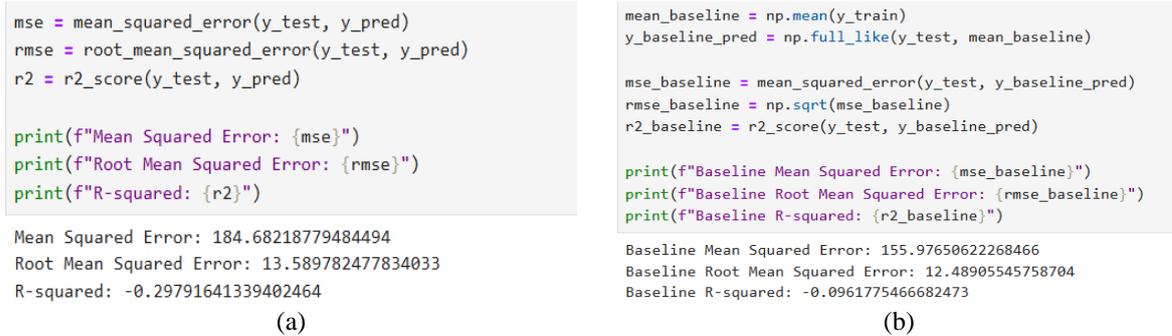
Dengan hasil peramalan yang didapat, berikut ditampilkan model regresi menggunakan *plot scatter* seperti pada gambar berikut:



Gambar 11. Plot Scatter Model Regresi Peramalan *Throughput* Akhir

3. Evaluasi Model Regresi

Setelah didapat model regresi dan telah dilakukannya peramalan, perhitungan terakhir yaitu menghitung error pada model regresi dengan mengambil jumlah total nilai absolut kesalahan peramalan individual dan membaginya dengan total periode data yang diramalkan menggunakan metode MSE dan RMSE. Setelah mendapatkan nilai MSE dan RMSE dari model, dapat juga dibandingkan dengan model baseline seperti pada gambar berikut:



Gambar 12. (a) Evaluasi Model dengan MSE dan RMSE (b) Evaluasi Model dengan Baseline

Dapat dilihat pada Gambar 12 (a) tingkat error MSE dan RMSE pada model yaitu 184,682 dan 13,589 sedangkan jika dibandingkan dengan tingkat error menggunakan Baseline mengalami penurunan yaitu 155,976 pada Baseline MSE, dan 12,489 pada Baseline RMSE. Ini menunjukkan bahwa model regresi yang dibuat dapat dijadikan model untuk meramal *throughput* akhir pada penelitian ini dengan cukup baik.

IV. KESIMPULAN

Pada penelitian ini terlihat bahwa perbedaan nilai peramalan yang dilakukan dengan perhitungan manual dan perhitungan menggunakan model Python, dikarenakan perhitungan manual tidak melalui tahap *preprocessing* data seperti perhitungan peramalan menggunakan model regresi yang dikembangkan dengan Python. Pengembangan model menggunakan Python juga menunjukkan perbedaan pada nilai evaluasi model regresi menggunakan MSE dan RMSE, yaitu 184,682 dan 13,589, dan nilai R^2 yaitu -0,29. Sedangkan jika dibandingkan dengan Baseline, tingkat error pada MSE dan RMSE mengalami penurunan, yakni 155,976 dan 12,489, serta nilai Baseline R^2 yaitu -0,09. Ini menunjukkan bahwa model regresi yang dibangun dapat digunakan untuk meramalkan *throughput* akhir client beberapa bulan kedepan dengan cukup baik. Peramalan ini dibutuhkan agar dapat menjadi bahan evaluasi nantinya untuk perusahaan penyedia layanan internet terkait untuk memperbaiki ataupun menjaga kualitas layanan internet yang diberikan kepada *client*. Dan dibutuhkan juga penelitian lebih lanjut dimasa depan guna mengetahui secara pasti bagaimana naik turunnya *throughput* menggunakan *machine learning* berdasarkan lainnya selain dari data yang telah tersaji, dan model ini diharapkan dapat dijadikan pengembangan untuk penelitian lebih lanjut.

REFERENSI

- [1] I. N. A. Prahara and I. R. Widiyari, "Implementasi Metode Received Signal Strength Indication dan Quality of Service Terhadap Analisis Kualitas Jaringan Wireless di CV Java Media Perdana Pati," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*. 2023. vol. 7, no. 4, pp. 528–535.
- [2] M. Z. E. Kalam, N. Made, A. Dewi, and I. M. O. Widyantara, "Analisa Kinerja Penerapan Standard Protokol Keamanan IEEE 802.11 Pada Layanan Wireless Fidelity," *Majalah Ilmiah Teknologi Elektro*. 2021. vol. 20, no. 1, pp. 89–94.
- [3] E. Fradinata, Sistem Informasi Enterprises. Syiah Kuala University Press. 2022.
- [4] N. L. W. Arya Della, R. A. N. Diaz, and K. D. P. Novianti, "Penerapan Metode Regresi Linier untuk Memprediksi Permohonan ITAS," *J. Eksplorasi Inform.*, 2021. vol. 10, no. 2, pp. 92–100.
- [5] A. Ihsan Fairuzsyifa and Y. Sulisty Nugroho, "Analisis Regresi Linier Berganda Pengaruh Minat Calon Mahasiswa di Universitas Muhammadiyah Surakarta Menggunakan Python," *J. Inform. Polinema*, 2024. vol. 10, no. 2, pp. 265–272.
- [6] M. K. Asror, and A. A. Slameto, "Analisis Perbandingan Kinerja Jaringan WLAN 2, 4 Ghz dan 5 Ghz

- pada Proses Thetering Menggunakan Metode QOS,” *Processor*. 2023. vol. 18, no. 2.
- [7] Z. Muttaqin and E. Srihartini, “Penerapan metode regresi linier sederhana untuk prediksi persediaan obat jenis tablet,” *JSiI (Jurnal Sistem Informasi)*. 2022. vol. 9, no. 1, pp. 12–16.
- [8] Tedyyana, Agus, et al. "Enhance Telecommunication Security Through the Integration of Support Vector Machines." *International Journal of Advanced Computer Science & Applications* 15.3 (2024).
- [9] A. Roihan, P. A. Sunarya, and A. S. Rafika, “Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper,” *IJCIT (Indonesian J. Comput. Inf. Technol.)*, 2020. vol.5, no.1, pp.75–82.
- [10] A. Triono, A. S. Budi, and R. Abdillah, “Implementasi Peretasan Sandi Vigenere Chipper Menggunakan Bahasa Pemograman Python,” *J. JOCOTIS - J. Sci. Inform. Robot.*, 2023. vol. 1, no. 1, pp. 1–9.
- [11] M. Ula, A. Zulfikri, A. F. Ulva, and R. A. Rizal, “Penerapan Machine Learning Clustering K-Means dan Linear Regression Dalam Penentuan Tingkat Resiko Tuberkulosis Paru,” *Indones. J. Comput. Sci.*, 2023. vol. 12, no. 1, pp. 336–348.
- [12] I. Nabillah and I. Ranggadara, “Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut,” *JOINS (Journal Inf. Syst.)*, 2020. vol. 5, no. 2, pp. 250–255.
- [13] W. S. Rudi, Y. A. Pranoto, F. X. Ariwibisono, and T. Informatika, “Kue Di Toko Karya Bahari Samarinda Berbasis Website,” 2023. vol. 7, no. 4, pp. 2451–2457.
- [14] M. A. Rozan, M. Tahir, A. P. Qirani, N. Rizqiullah, M. Veranda, R. Puji, and A. Ghaffar, “IMPLEMENTASI WEB PROXY PADA MIKROTIK UNTUK MENGOPTIMALKAN KEAMANAN JARINGAN WIRELESS LAN,” *JUKANTI (JURNAL PENDIDIKAN TEKNOLOGI INFORMASI)*, 2024. no. April, pp. 180–188.
- [15] V. A. Saputro and Raden rara Kartika Kusuma Winahyu, “Pengaruh Penggunaan Channel Bonding Terhadap Interferensi Jaringan Nirkabel,” *J. Inform. dan Rekayasa Elektron.*, 2023. vol. 6, no. 2, pp. 238–245.
- [16] P. R. Utami, “ANALISIS PERBANDINGAN QUALITY OF SERVICE JARINGAN INTERNET BERBASIS WIRELESS PADA LAYANAN INTERNET SERVICE PROVIDER (ISP) INDIHOME DAN FIRST MEDIA,” *Jurnal Ilmiah Teknologi dan Rekayasa*. 2020. pp. 125–137.