

Analisis Harga Kuota Internet Menggunakan Metode Multiple Regresi Linear

Muhammad Hafizh Al Mustofa^{1*}, Nurmalitasari²

^{1,2}Fakultas Ilmu Komputer
Jl. Bhayangkara No. 55 Tipes, Surakarta

^{1*}202021120@mhs.udb.ac.id, ²nurmalitasari@udb.ac.id

Abstrak—Perkembangan teknologi yang semakin maju ditunjukkan pada meningkatnya penggunaan internet di zaman milenial. Hal tersebut dapat terjadi karena adanya perkembangan dan kebutuhan alat komunikasi semakin pesat. Salah satu kebutuhannya adalah kuota internet. Metode yang digunakan untuk memprediksi harga kuota internet adalah regresi linear berganda. Regresi linier berganda adalah metode yang memodelkan variabel dependen dan variabel independen yang lebih dari satu, dengan menganalisis hubungan antar variabel-variabel tersebut. Dataset yang digunakan memiliki 250 data dengan 50 kolom, variabel dependennya adalah harga dengan variabel independennya adalah besar kuota GB, kecepatan per Mbps, berdasarkan jenis daerah, dan masa aktif kuota per hari. Tujuan dari penulisan adalah mengenalkan penggunaan metode regresi linier pada masyarakat terutama pada golongan mahasiswa. Hasil evaluasi linier dalam nilai MAE, MSE, RMSE, dan MAPE menunjukkan bahwa data memiliki tingkat error yang cukup rendah. Grafik prediksi vs aktual menunjukkan bahwa nilai aktual dan prediksi tidak terlalu jauh, sehingga keakuratan prediksi menjadi tinggi.

Kata kunci— Harga, Kuota Internet, dan Regresi linier berganda,

Abstract— The advancement of technology is evident in the increasing use of the internet in the millennial era. This is due to the rapid development and demand for communication tools. One of the needs is the quotas data. The method used to predict data quota prices is multiple linear regression, which is implemented in Python and Google Colab. Multiple linear regression is a method that models more than one dependent variable and an independent variable, by analyzing the relationship between those variables. The dataset used consists of 250 data points with 50 columns, where the dependent variable is the price and the independent variables are the size of the quota in GB, speed per Mbps, based on the type of region, and the quota's active duration per day. The purpose of this writing is to introduce the use of linear regression method to the community, especially students. The evaluation results of linear regression in terms of MAE, MSE, RMSE, and MAPE indicate that the data has an error rate. The prediction vs. actual graph shows that the actual and predicted values are not too far apart, indicating a high level of prediction accuracy.

Keywords: Price, Quota Data, and Multiple Linear Regression

I. PENDAHULUAN

Kemajuan teknologi dapat dilihat dari semakin maraknya penggunaan internet di kalangan masyarakat zaman sekarang. Hal tersebut diikuti oleh perkembangan alat komunikasi yang semakin pesat dan semakin murah, sehingga kebutuhan untuk komunikasi semakin meningkat. Alat komunikasi layaknya hp memiliki kelebihan yaitu dapat berkomunikasi jarak jauh dan mengakses informasi dengan lebih cepat, akan tetapi hal tersebut membutuhkan konsumsi kuota yang cukup banyak agar bisa berfungsi.

Menurut data statis yang dipublikasikan oleh cyberthreat.id, terdapat 175.4 juta pengguna internet Indonesia dan 160.0 juta pengguna social media [1]. Jumlah pengguna internet Indonesia meningkat 2,67% pada tahun 2023 dibandingkan tahun sebelumnya, dengan total 215,63 juta orang pada tahun 2022–2023, menurut survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII). Hal ini menunjukkan bahwa pengguna internet dan

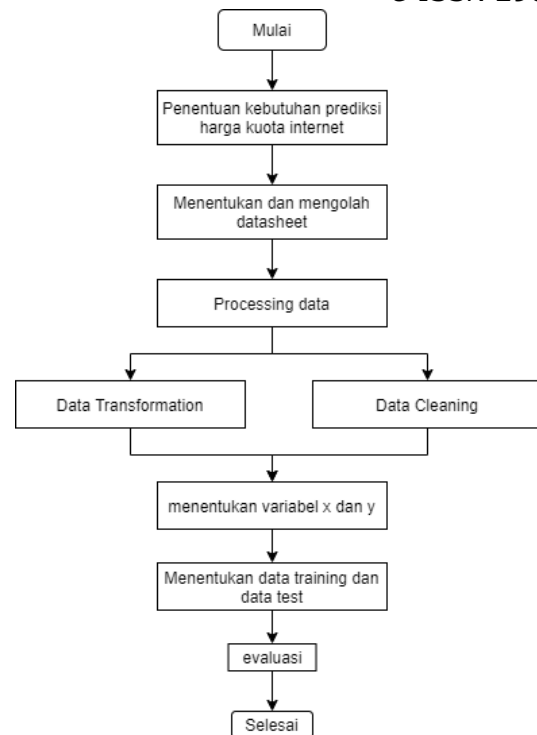
pembelian paket data yang digunakan di internet juga akan meningkat.

Karena semakin banyaknya pengguna internet di Indonesia maupun luar negeri, maka otomatis kuota yang digunakan oleh masyarakat pada zaman sekarang akan semakin meningkat. Salah satu metode yang digunakan untuk memprediksi sebuah objek adalah metode regresi linier. Regresi linier adalah suatu teknik statistik yang digunakan untuk memodelkan dan menganalisis hubungan antara satu variabel dependen (variabel terikat) dengan satu atau lebih variabel independen (variabel bebas) [2]. Tujuan utama penggunaan regresi linier adalah untuk menghitung hubungan antara parameter model dan memperkirakannya [3]. Salah satu jenis metode linier yang digunakan adalah metode linier berganda, regresi linear berganda adalah metode pendekatan statistik yang digunakan untuk memodelkan hubungan antara satu variabel dependen dan beberapa variabel independen. Tujuannya adalah untuk mengidentifikasi dan

memahami hubungan antara variabel-variabel tersebut serta membuat prediksi atau estimasi nilai variabel dependen berdasarkan variabel independen yang ada [4]. Dalam penelitian sebelumnya, yang melibatkan usia, gender, BMI, dan anak-anak, metode regresi linear digunakan. Hasil uji menunjukkan bahwa ada korelasi yang kuat antara variabel independen, yaitu biaya (changer) dan perokok (smoker), dengan nilai 0,79. Hasil ini juga menunjukkan bahwa orang yang merokok akan membayar premi asuransi lebih banyak daripada orang yang tidak merokok. Selain itu, korelasi antara biaya (changer) dan perokok (smoker) adalah positif [5], penelitian lain menggunakan penerapan regresi linear untuk memprediksi hasil nilai kuesioner mahasiswa dapat diperoleh dari variabel aspek1, aspek2, aspek3, aspek4, IPK_responden, jumlah_responden dan rata-rata yaitu Studi ini menemukan nilai dengan nilai koefisien 0,997 dan rumus regresi linear. Ini menunjukkan bahwa variabel jumlah respons tidak berhubungan dengan variabel lain [6], dan terdapat penelitian analisis kasus covid menggunakan regresi linear dan random forest [7]. Maka dari itu penggunaan *multiple regression linear* dapat digunakan pada bidang yang berbeda-beda dan memberikan data yang digunakan untuk mengambil keputusan, tujuan penulisan ini adalah menganalisis harga kuota sebagai variabel dependen dan variabel independennya adalah besar kuota per gb, kecepatan per mbps, berdasarkan jenis daerah, dan masa aktif kuota per hari. Selain itu, tujuan dari penulisan ini adalah memperkenalkan kepada masyarakat penggunaan python untuk mempermudah analisis data terutama pada golongan mahasiswa.

II. METODOLOGI PENELITIAN

a. Alur penelitian



Gambar 1. Alur penelitian

Alur penelitian terdiri dari berbagai langkah, seperti memilih topik penelitian, menentukan masalah penelitian, menentukan tujuan penelitian, memilih metode penelitian, mengumpulkan data, menganalisis data, menginterpretasikan temuan, dan membuat laporan penelitian [8]. Pengertian lain menyatakan bahwa alur penelitian terdiri dari langkah-langkah seperti pemilihan subjek, perumusan masalah, pembatasan masalah, penentuan tujuan penelitian, penggunaan metode, pengumpulan data, analisis dan interpretasi data, dan penyusunan laporan penelitian [8]. Dalam penulisan ini alur dapat diperoleh dari penentuan factor yang mempengaruhi prediksi harga kuota internet, menentukan sumber data atau factor yang berpengaruh, memproses data, menentukan variabel x dan y untuk menghitung prediksi, menentukan data training dan data test, dan evaluasi perhitungan.

III. HASIL DAN PEMBAHASAN

3. Data yang digunakan

Hasil observasi disimpan dalam bentuk file excel dan dilakukan penghapusan data yang tidak digunakan. File excel disimpan ke dalam bentuk csv

agar bisa terbaca dengan mudah di pemrograman python. Dataset memiliki total 250 data.

4. Processing data

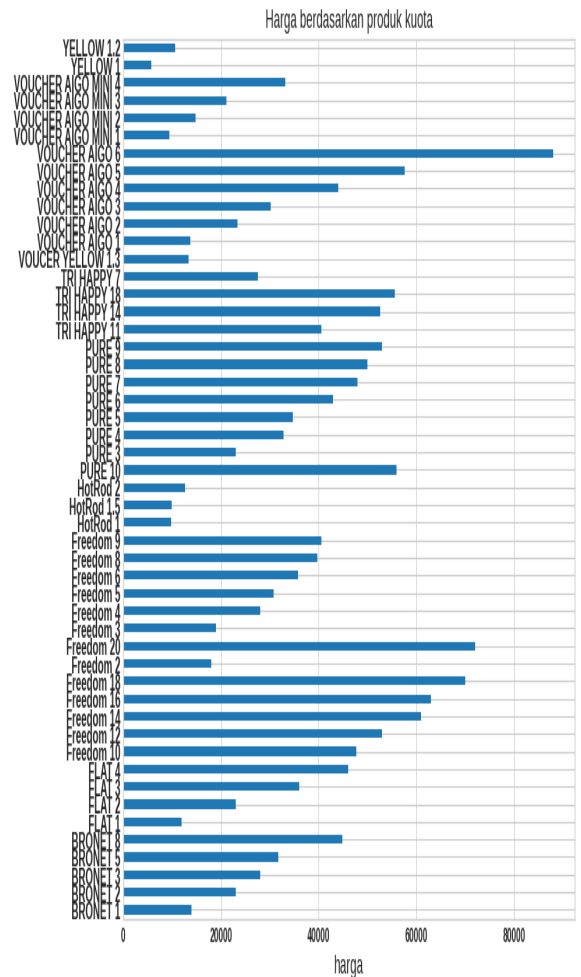
Menampilkan data yang telah dikumpulkan dalam bentuk CSV dan di import ke dalam model python.

<i>produk</i>	<i>harga</i>	<i>Besar kuota</i>	<i>Jenis daerah</i>	<i>Kecepatan mbps</i>	<i>Masa Aktif</i>
<i>Voucher Aigo 1</i>	13775	1.0	1	4.0	30
<i>Voucher Aigo 2</i>	23375	2.0	1	5.0	30
<i>Voucher Aigo 3</i>	30175	3.0	1	5.5	30

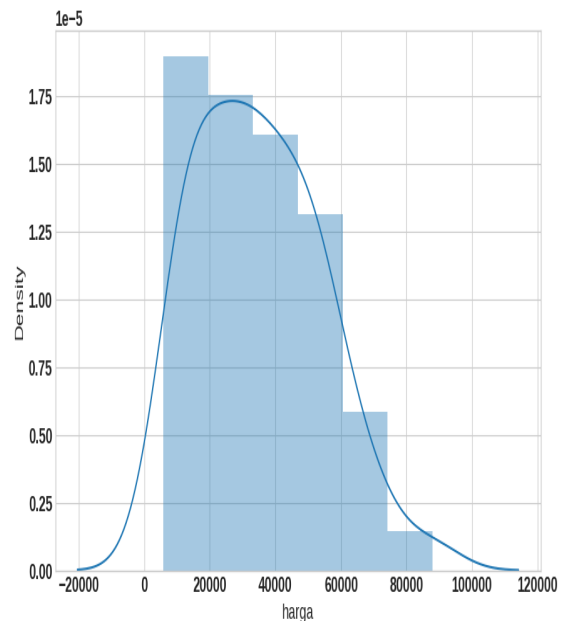
Gambar 2. Table data

5. Visualiasi data

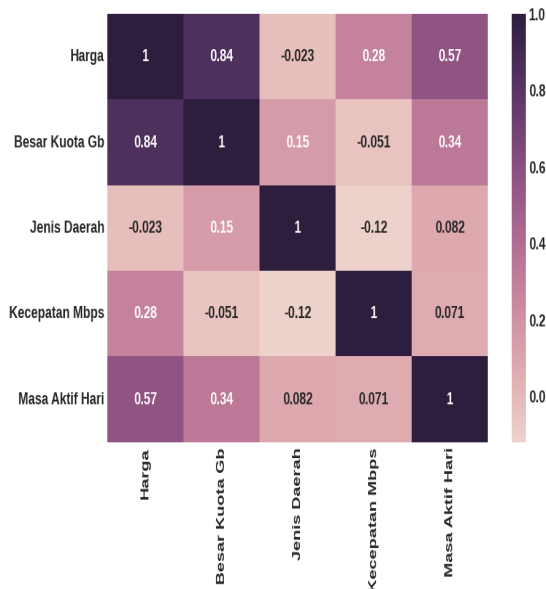
Visualisasi data adalah menggambarkan data dengan menggunakan elemen visual seperti grafik, diagram, dan ilustrasi untuk membantu analisis, pemahaman, dan komunikasi informasi [9]. Selain itu visualiasi data adalah proses membuat representasi visual dari data yang memungkinkan kita untuk menjelajahi, memahami, dan mengkomunikasikan pola dan informasi dalam data. Dari pengertian di atas, dapat disimpulkan visualisasi data adalah data yang di presentasikan dalam bentuk grafik, diagram, dan ilustrasi agar dapat memudahkan pembaca untuk menafsirkan isi data tersebut [10]. Salah satu visualisasi data yang digunakan adalah histogram. Histogram adalah jenis visualisasi data untuk merepresentasikan distribusi frekuensi dari dataset numeric.



Gambar 3. Grafik harga berdasarkan produk kuota



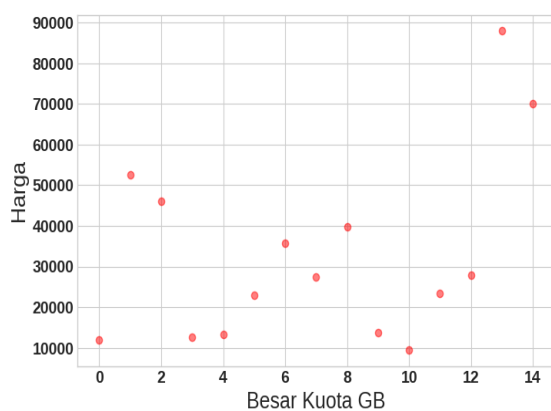
Gambar 4. Histogram harga kuota



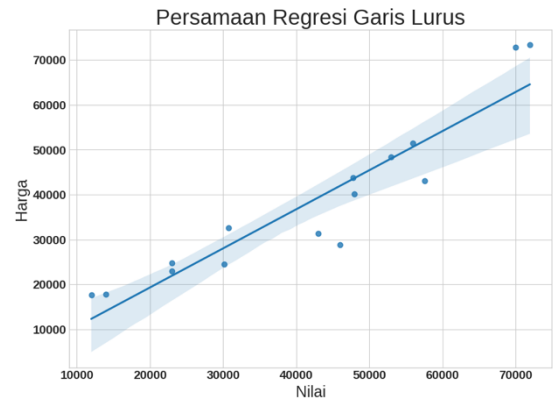
Gambar 5. korelasi variabel

Gambar diatas menunjukkan korelasi variabel-variabel tertentu. Variabel yang berpengaruh besar terhadap harga adalah besar kuota gb dengan angka nominal yaitu 0.842.

6. Pembuatan Model Linear Berganda
Model yang digunakan untuk menentukan variabel dependen (x) dan variabel independen (y). Harga sebagai varibel dependen dengan besar kuota, jenis daerah, kecepatan mbps, dan masa aktif menjadi variabel independen.



Gambar 6. Perbandingan harga dengan besar kuota



gambar 7. persamaan garis linier

7. Data training dan data test
Data training dan data set digunakan untuk mengembangkan dan mengevaluasi sebuah model. Data training digunakan untuk melatih model statistika yang berupa pola, hubungan, dan fitur-fitur tertentu, sedangkan data testing digunakan untuk mengetahui performa model yang sudah dilatih sebelumnya untuk mengukur sebuah prediksi. Komposisi pembagian ini adalah 80% dari data yang ada untuk data training dan 20% untuk data test [12].
8. Uji normalitas
Uji normalitas digunakan untuk menguji sebuah data agar bisa distribusi secara normal, salah satu metode uji normalitas adalah *jarque-bera*. Nilai jarque bera yang di dapat adalah sebesar 0.3181971427534504. Nilai minimal data distribusi normal adalah 0,05 maka hasil jarque bera diatas, termasuk dalam pendistribusian data yang normal.
9. Uji Statistik
Uji statistik Shapiro-Wilk didasarkan pada hipotesis nol bahwa data diambil dari populasi yang terdistribusi secara normal. Metode ini menghasilkan statistik uji W dan nilai p-value yang menunjukkan seberapa baik hipotesis nol didukung oleh jumlah data yang dikumpulkan. Hasil yang diperoleh adalah $pvalue=0.1051243459119072$, $pvalue=3.8327189031406e-$

06,pvalue=5.194955472417475e-07,p
value=5.194954724174754-07,
pvalue=1.5432672501036748e-11.

kualitas model yang lebih baik. Nilai
RMSE yang diperoleh adalah
7331.000915815389.

10. Evaluasi linear

Pengujian model linear regression digunakan pada pengujian data secara menyeluruh, metode perhitungan akurasi sendiri akan diberlakukan. Penerapan metode ini dilakukan dengan perhitungan nilai total jumlah prediksi yang benar pada data[11]. Evaluasi linear adalah proses mengukur dan menganalisis kualitas dan kinerja model regresi linier dalam memprediksi nilai variabel target berdasarkan variabel prediktor yang ditentukan. Evaluasi ini bertujuan untuk menilai model linear dalam memberikan keakuratan prediksi dan bersifat relevan terhadap data yang ada.

- Nilai MAE mengukur rata-rata dari selisih absolut antara nilai prediksi dan nilai sebenarnya. MAE mengabaikan arah kesalahan, artinya hanya memperhitungkan besarnya selisih. Nilai MAE, nilai MAE yang diperoleh adalah 5636.525999441807.
- Nilai MAPE diukur mengukur persentase rata-rata dari selisih absolut antara nilai prediksi dan nilai sebenarnya. Nilai MAPE yang diperoleh 0.26227708235182656.
- Nilai MSE diukur dari persentase rata-rata dari selisih absolut antara nilai prediksi dan nilai sebenarnya. Nilai MSE yang diperoleh 53743574.42768607
- Nilai RMSE
Nilai RMSE adalah akar kuadrat dari MSE. Nilai RMSE sering digunakan karena menghasilkan skala yang serupa dengan data asli. Nilai RMSE yang lebih kecil menunjukkan tingkat kesalahan yang lebih rendah dan

- Perhitungan linear dengan Librystatsmodel
Evaluasi menggunakan statsmodel dapat diperoleh hasil dibawah

<i>Dep. Variable:</i>	harga
<i>Model:</i>	OLS
<i>Method:</i>	Least Squares
<i>Date:</i>	Fri, 23 Jun 2023
<i>Time:</i>	14:36:48
<i>No.</i>	50
<i>Observations:</i>	
<i>Df Residuals:</i>	46
<i>Df Model:</i>	4
<i>R-squared</i>	0.977
<i>(uncentered):</i>	
<i>Adj. R-squared</i>	0.975
<i>(uncentered):</i>	
<i>F-statistic:</i>	489.1
<i>Prob (F-</i>	4.76e-37
<i>statistic):</i>	
<i>Log-</i>	-506.18
<i>Likelihood:</i>	
<i>AIC:</i>	1020.
<i>BIC:</i>	1028.

Dengan nilai Coef

coef	std err	t	P> t	[0.025	0.975]
Besar	2706.19	175.47	15.4	0.000	2352.9
Kuota	46	1	22	0	89
GB					
Jenis	-5826.29	1749.804	-3.33	0.002	-9348.4
Darah	1660.40	281.31	5.90	0.000	1094.1
Kecapatan	34	7	2	0	42

Mbps					
Masa Aktif Hari	536.2067	83.106	6.452	0.000	368.923

Nilai uji normalitas

Durbin-Watson:	0.842
Jarque-Bera (JB):	0.089
Prob(JB)	0.956
Cond. No.	56.0
Durbin-Watson:	0.842
Omnibus:	0.033
Prob(Omnibus) :	0.984
Skew:	0.048
Kurtosis:	2.816

Gambar 7. Hasil OLS

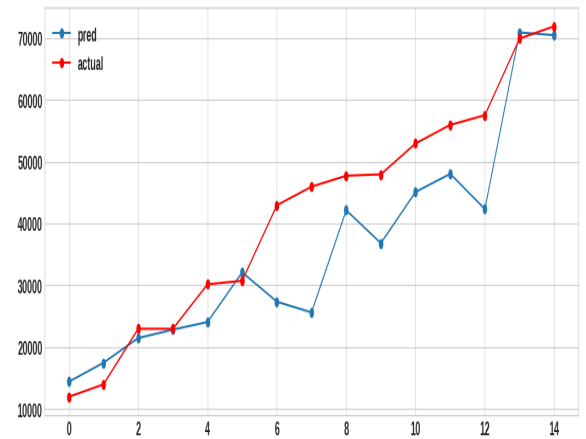
11. Hasil Prediksi

Setelah menemukan nilai dari MAE, MSE, MAPE, dan RMSE maka dapat menilai keakuratan prediksi.

Hasil prediksi menggunakan rumus $8275.74 + \text{Besar Kuota GB} * 3072.45 + \text{Kecepatan Mbps} * 3948.98 + \text{Jenis Daerah} * 2114.85$.

Adanya implementasi prediksi dari harga kuota dengan besar kuota 10 gb, kecepatan 20 mbps, dan jenis daerah adalah local adalah sebesar 128.431.

Hasil perbandingan data actual dan prediksi dapat di presentasikan dalam grafik dalam gambar 8.



Gambar 8. Grafik prediksi vs actual

Dari grafik diatas dapat ditafsirkan bahwa nilai prediksi (biru) tidak terlalu jauh dengan nilai actual (merah) sehingga keakuratan dari hasil prediksi cukup tinggi.

IV. KESIMPULAN

Metode regresi linier dapat digunakan untuk memprediksi harga kuota dengan baik, dengan cara menentukan datasheet yang digunakan dan memproses data. Hasil korelasi yang didapat adalah bahwa variabel besar kuota per gb sangat berpengaruh terhadap variabel harga karena memiliki nilai tertinggi yaitu 0.842. Adanya hasil uji normalitas dengan nilai 0.31 menunjukkan bahwa data dapat distribusi dengan normal. Selain itu, dengan adanya nilai RMSE, MAE, MSE, dan MAPE maka dapat diperoleh hasil prediksi dengan linear berganda. Pada tabel actual dan prediksi menunjukkan bahwa nilai actual dan prediksi tidak berbeda jauh, sehingga nilai keakuratan menjadi cukup tinggi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih terhadap pihak-pihak dalam membantu penulisan jurnal, terutama pada pihak dosen yang senantiasa membantu dalam konsultasi penulisan.

REFERENSI

- [1] <https://www.bing.com/search?pglt=2083&q=data+pengguna+internet+cyberthreat+id&cvid=5911271ae48e40be9df4dfc0d90397e4&aqs=edg e..69i57j018.9328j01&FORM=ANNTA1&PC=U531>
- [2] Joseph F. Hair, William C. Black, Barry J. Babin, Rolph E. Anderson(2013). *Multivariate Data Analysis*. Pearson Education Limited,London.
- [3] Gujarati, D.N. and Porter, D.C. (2009). *Basic Econometrics*. 5th Edition, McGraw Hill Inc. New York.
- [4] Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Regression Models*, McGraw-Hill/Irwin. New York.
- [5] Muhammad Sholeh, Suraya, Dina Andayati (2022). Machine Linear untuk Analisis Regresi Linier Biaya Asuransi Kesehatan dengan Menggunakan Python Jupyter Notebook. *Jurnal Edukasi Dan Penelitian Informatika*.
- [6] Muhammad Sholeh (2022). Penerapan Regresi Linear Ganda Untuk Memprediksi Hasil Nilai Kuesioner Mahasiswa Dengan Menggunakan Python. *Jurnal Dinamika Informatika*. Semarang
- [7] Alma Hidayanti,Amril Mutoi Siregar, Santi Arum Puspita Lestari, Yana Cahyana(2022). Model Analisis Kasus Covid-19 Di Indonesia Menggunakan Algoritma Regresi Linier Dan Random Forest.Petir.
- [8] Sugiyono (2017). *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Alfa Beta. Bandung
- [9] Arikunto S (2010). *Prosedur Penelitian*. Rineka Cipta. Jakarta.
- [10] Edward R. Tufte (1997). *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press
- [11] Hadley Wickham (2009). *ggplot2: elegant graphics for data analysis*. Springer.
- [12] Yayan Adrianova Eka Tuah(2020). Implementasi Linear Sederhana Untuk Prediksi Gaji Berdasarkan Pengalaman Lama Bekerja. *Jutech*
- [13] W. Musu et al. "Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4 . 5;". in *PROSIDING SEMINAR ILMIAH SISTEM INFORMASI DAN TEKNOLOGI INFORMASI*, 2021, pp. 186–195.