

## **PREDIKSI DAN ANALISIS FAKTOR GAYA HIDUP TERHADAP SKOR KESEHATAN MENGGUNAKAN REGRESI LINEAR DI RAPIDMINER**

Aisyah Qurrata Ayun<sup>1</sup>, Ulva Dwita<sup>2</sup>, Yuhelni Mala Sari<sup>3</sup>

<sup>1,2,3</sup> Program Studi Informatika, Universitas Adzkia

aisyaquraa@gmail.com

### **ABSTRAK**

Kesehatan merupakan aset penting dalam kehidupan manusia yang sangat dipengaruhi oleh pola hidup sehari-hari. Penelitian ini bertujuan untuk melakukan prediksi dan analisis terhadap skor kesehatan individu berdasarkan faktor-faktor gaya hidup seperti usia, indeks massa tubuh (BMI), frekuensi olahraga, kualitas diet, durasi tidur, status merokok, dan konsumsi alkohol. Metode yang digunakan adalah regresi linear dengan bantuan perangkat lunak RapidMiner sebagai alat analisis. Hasil pengujian menunjukkan bahwa semua variabel memiliki pengaruh yang signifikan secara statistik terhadap skor kesehatan ( $p < 0,05$ ). Model terbaik diperoleh saat proporsi data pelatihan sebesar 90% dan data pengujian sebesar 10%, dengan nilai RMSE terendah sebesar 5.495. Analisis regresi menunjukkan bahwa durasi tidur, frekuensi olahraga, dan kualitas diet memiliki pengaruh positif terhadap skor kesehatan, sedangkan usia, BMI, kebiasaan merokok, dan konsumsi alkohol berdampak negatif. Temuan ini mendukung pentingnya pola hidup sehat dalam meningkatkan kualitas kesehatan, serta membuktikan bahwa regresi linear dapat menjadi pendekatan efektif dalam sistem pendukung keputusan berbasis data di bidang kesehatan.

**Kata Kunci:** skor kesehatan, gaya hidup, regresi linear, RapidMiner, prediksi kesehatan

*JRSKM is licensed under a Creative Commons 4.0 International License.*



### **1. PENDAHULUAN**

Kesehatan merupakan aset penting dalam kehidupan manusia. [1] Seiring berkembangnya gaya hidup modern, pola makan tidak sehat, kurangnya aktivitas fisik, kebiasaan merokok, serta pola tidur yang tidak teratur menjadi faktor yang memengaruhi penurunan kualitas kesehatan masyarakat. [2] Oleh karena itu, upaya untuk memahami dan mengukur pengaruh gaya hidup terhadap kesehatan menjadi penting untuk dilakukan, terutama dengan memanfaatkan teknologi dan data yang tersedia. [3]

Dalam era digital dan data saat ini, metode analisis data seperti regresi linear dapat digunakan untuk membangun model prediksi kesehatan berdasarkan berbagai variabel gaya hidup. [4] Regresi linear merupakan metode statistik yang digunakan untuk memodelkan hubungan antara satu variabel dependen dengan satu atau lebih variabel independen melalui suatu persamaan linear. Model ini memungkinkan kita untuk memprediksi nilai variabel dependen berdasarkan nilai variabel independen yang diketahui. [5] Dengan menggunakan alat bantu seperti RapidMiner, proses analisis dapat dilakukan secara efisien dan terstruktur. [6] Regresi linear tidak hanya memungkinkan untuk memprediksi skor kesehatan seseorang berdasarkan faktor gaya hidup, tetapi juga memberikan gambaran mengenai seberapa besar pengaruh masing-masing faktor tersebut terhadap hasil akhir.

Berdasarkan penelitian sebelumnya, regresi linear digunakan untuk memprediksi jumlah perokok dan dampaknya terhadap kesehatan masyarakat. Hasilnya dengan menggunakan regresi linear untuk memprediksi

jumlah perokok, dapat memodelkan hubungan antara variabel-variabel yang berpotensi mempengaruhi jumlah perokok dengan variabel dependen yang merupakan jumlah perokok itu sendiri. Regresi dapat digunakan untuk menganalisis faktor-faktor yang mempengaruhi perilaku merokok dan merumuskan strategi pencegahan yang lebih efektif untuk mengurangi prevalensi merokok dan meningkatkan kesehatan masyarakat. [7]

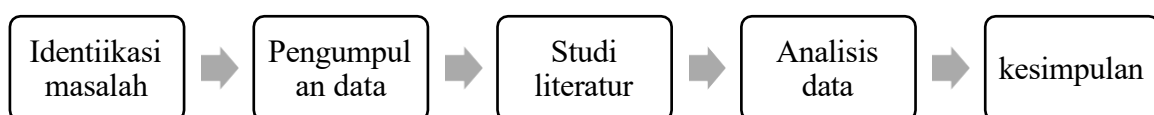
Regresi linear juga mampu memprediksi suhu badan sapi menggunakan *data smart* ternak dari PT Telkom Indonesia. Model ini mampu memahami dan mengukur masing-masing faktor lingkungan terhadap suhu tubuh sapi. Hasil pengujian model lumayan tinggi yang menunjukkan kemampuan model dalam memprediksi suhu tubuh sapi dengan baik dan menjelaskan variasi dalam data. [8] Kemudian regresi linear juga digunakan untuk menganalisis faktor ekonomi dan sosial dalam kejadian *stunting* pada anak, regresi linear bertujuan untuk menganalisis variabel independen, yaitu pendapatan keluarga, tingkat pendidikan orang tua, dan akses terhadap layanan kesehatan, dengan variabel dependen, yaitu status *stunting* anak berdasarkan indeks *Z-score*. Kesimpulannya, hasil ini mendukung bahwa model regresi efektif dalam menjelaskan faktor-faktor yang memengaruhi status *stunting* anak berdasarkan indeks *Z-score*. [9]

Metode regresi linear juga digunakan untuk menganalisis faktor-faktor yang mempengaruhi harga cabai merah di kota Manado Provinsi Sulawesi Utara. Data yang digunakan adalah data harga cabai per bulan di Kota Manado periode tahun 2020-2022. Variabel yang digunakan dalam penelitian ini adalah produksi, luas lahan, curah hujan, dan inflasi. Maka hasil yang diperoleh yaitu faktor-faktor yang mempengaruhi harga cabai merah di Kota Manado adalah luas lahan dan inflasi. [10]

Dari uraian diatas maka penelitian ini bertujuan untuk melakukan prediksi dan analisis terhadap skor kesehatan individu berdasarkan faktor-faktor gaya hidup seperti usia, indeks massa tubuh (BMI), frekuensi olahraga, kualitas diet, jam tidur, kebiasaan merokok, dan kebiasaan mengkonsumsi alkohol. Dengan menganalisis koefisien regresi yang dihasilkan, dapat diketahui faktor mana yang paling dominan dalam menentukan skor kesehatan seseorang. Hasil dari penelitian ini diharapkan dapat memberikan wawasan dan rekomendasi praktis dalam mendukung pola hidup sehat.

## 2. METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan deskriptif-prediktif. Metode yang digunakan adalah regresi linear, yang bertujuan untuk menganalisis hubungan antar variabel gaya hidup dengan skor kesehatan, serta membangun model prediksi skor kesehatan menggunakan data numerik. Dalam prosesnya penelitian ini melakukan beberapa tahapan, sebagai berikut:



Gambar2. Tahpan Penelitian

### 1. Identifikasi Masalah

Pada tahap ini, peneliti mengidentifikasi isu utama bahwa skor kesehatan seseorang sangat dipengaruhi oleh gaya hidup sehari-hari. Namun, belum banyak yang secara kuantitatif memprediksi skor kesehatan berdasarkan faktor-faktor tersebut.

2. Pengumpulan Data

Data sekunder berupa dataset sintetik tentang gaya hidup dan kesehatan diperoleh dalam format .csv. *Dataset* ini berisi variabel-variabel yang diduga berpengaruh terhadap skor kesehatan.

3. Studi Literatur

Peneliti menelusuri jurnal, artikel, dan penelitian terdahulu untuk memahami hubungan gaya hidup dengan kesehatan, serta melihat bagaimana regresi linear telah digunakan dalam analisis prediktif di bidang kesehatan.

4. Analisis Data

Proses ini dilakukan menggunakan RapidMiner, dengan tahapan:

1. Preprocessing data
2. Pembuatan model regresi linear
3. Evaluasi model
4. Interpretasi hasil koefisien variabel untuk mengetahui seberapa besar dan arah pengaruh tiap faktor terhadap skor kesehatan.

5. Kesimpulan

Pada akhir penelitian, diperoleh kesimpulan tentang variabel gaya hidup mana yang paling signifikan berpengaruh terhadap skor kesehatan. Selain itu, model prediksi yang dibangun dapat digunakan sebagai alat bantu untuk memperkirakan skor kesehatan berdasarkan pola hidup individu.

## 2.1. Faktor Gaya Hidup

Gaya hidup mencerminkan pola perilaku sehari-hari. Gaya hidup yang baik adalah gaya hidup yang memberikan dampak positif dan bermanfaat bagi hidup. Sebaliknya, gaya hidup tidak sehat seperti kurangnya aktivitas fisik, konsumsi makanan tidak sehat, pola istirahat yang tidak teratur, diet yang buruk, dan stres akan berdampak negatif pada kesehatan. [11]

Dalam penelitian ini, faktor gaya hidup yang dianalisis meliputi usia, indeks massa tubuh (BMI), frekuensi olahraga, kualitas diet, jam tidur, dan status merokok. Faktor-faktor ini dipilih karena memiliki pengaruh signifikan terhadap skor kesehatan individu.

### 2.1.1. Usia (*Age*)

Usia merupakan faktor penting dalam menentukan status kesehatan. Seiring bertambahnya usia, risiko terkena penyakit kronis seperti hipertensi, diabetes, dan penyakit jantung meningkat. Oleh karena itu, usia dijadikan variabel independen untuk menganalisis pengaruhnya terhadap skor kesehatan. [12]

### 2.1.2. Indeks Masa Tubuh (BMI)

BMI adalah indikator yang digunakan untuk menilai status berat badan seseorang. BMI yang tinggi atau rendah dapat menunjukkan adanya risiko kesehatan, seperti obesitas atau malnutrisi. Dalam penelitian ini, BMI digunakan untuk melihat hubungan antara status berat badan dengan skor kesehatan. [13]

### 2.1.3. Frekuensi Olahraga (*Exercise Frequency*)

Aktivitas fisik secara teratur dapat meningkatkan kebugaran dan menurunkan risiko penyakit. Namun, frekuensi olahraga yang terlalu tinggi atau rendah dapat berdampak negatif pada kesehatan. Oleh karena itu, frekuensi olahraga dianalisis untuk mengetahui pengaruhnya terhadap skor kesehatan. [14]

#### 2.1.4. Kualitas Diet (*Diet Quality*)

Kualitas diet mencerminkan seberapa baik pola makan seseorang dalam memenuhi kebutuhan gizi. Diet yang seimbang dan kaya nutrisi dapat meningkatkan kesehatan, sedangkan diet yang buruk dapat meningkatkan risiko penyakit. Variabel ini digunakan untuk menilai pengaruh pola makan terhadap skor kesehatan. [15]

#### 2.1.5. Jam Tidur (*Sleep Hours*)

Tidur yang cukup dan berkualitas penting untuk pemulihan tubuh dan fungsi kognitif. Kurang tidur dapat menyebabkan berbagai masalah kesehatan, termasuk gangguan metabolisme dan penurunan sistem imun. Oleh karena itu, jam tidur dianalisis dalam kaitannya dengan skor kesehatan. [16]

#### 2.1.6. Status Merokok (*Smoking Status*)

Merokok diketahui sebagai faktor risiko utama berbagai penyakit serius, termasuk kanker dan penyakit jantung. Status merokok (perokok atau tidak) digunakan untuk melihat sejauh mana kebiasaan ini mempengaruhi skor kesehatan individu. [17]

## 2.2. Regresi Linear

Regresi linear adalah salah satu metode statistik yang digunakan untuk memodelkan hubungan antara satu variabel dependen dengan satu atau lebih variabel independen. Metode ini berfungsi untuk memprediksi nilai variabel dependen berdasarkan nilai-nilai dari variabel independennya. Dalam penelitian ini, regresi linear digunakan untuk memprediksi Skor Kesehatan (*Health\_Score*) berdasarkan faktor-faktor gaya hidup seseorang.

### 2.2.1. Prediksi dan Analisis dengan Regresi Linear

Dalam konteks penelitian ini, regresi linear digunakan untuk dua tujuan, yaitu:

1. Prediksi (*prediction*) menggunakan nilai-nilai variabel independen untuk memperkirakan skor kesehatan seseorang. Hal ini berguna dalam sistem pendukung keputusan kesehatan berbasis data. [18]
2. Analisis (*analysis*) untuk mengetahui seberapa besar dan arah pengaruh masing-masing faktor gaya hidup terhadap skor kesehatan. Koefisien regresi akan menunjukkan apakah suatu faktor berdampak positif atau negatif, serta seberapa signifikan pengaruhnya secara statistik. [4]

### 2.2.2. Rumus Regresi Linear

#### 1. Regresi Linear Sederhana

Regresi linear sederhana adalah model regresi linear yang hanya memiliki satu peubah bebas. Secara konvensional, pada sistem koordinat Kartesius model ini digambarkan sebagai suatu garis lurus, di mana sumbu X mewakili peubah bebas dan Y peubah acak terikat. Peubah terikat Y ini kadang-kadang juga disebut sebagai respon. [19]

Rumus umum:

Y	: variabel dependen
X	: variabel independen
$\alpha$	: intersep (nilai Y saat X=0)
b	: koefisien regresi
$\varepsilon$	: eror atau residu

## 2. Regresi Linear Berganda

Regresi linear berganda adalah metode statistik yang digunakan untuk menganalisis hubungan antara satu variabel dependen (Y) dengan dua atau lebih variabel independen ( $X_1, X_2, \dots, X_n$ ). Tujuannya adalah untuk memprediksi nilai Y berdasarkan nilai-nilai  $X_1$  hingga  $X_n$ . [19]

Rumus Umum:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon$$

Y	: variabel dependen
$X_1, X_2, \dots, X_n$	: variabel-variabel independen
$\alpha$	: intersep
$b_1, b_2, \dots, b_n$	: Koefisien regresi masing-masing variabel
$\varepsilon$	: error

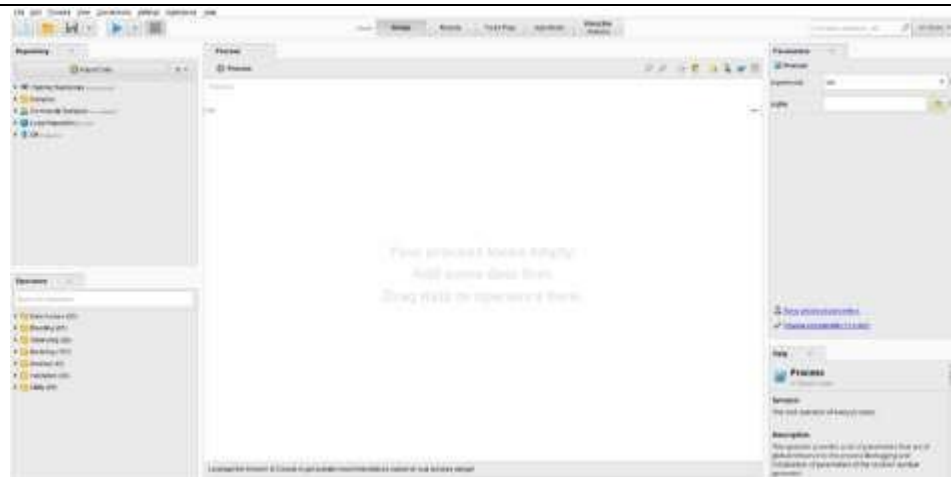
## 2.3. RapidMiner

RapidMiner adalah *platform* perangkat lunak untuk ilmu data yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran mendalam, penambangan teks, dan analisis prediktif. Platform ini digunakan dalam aplikasi bisnis dan komersial, serta untuk penelitian, pendidikan, pelatihan, *prototype* cepat, dan pengembangan aplikasi.

RapidMiner, sebelumnya dikenal sebagai YALE (*Yet Another Learning Environment*), dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Unit Kecerdasan Buatan Universitas Teknik Dortmund. Pada tahun 2006, pengembangannya dilanjutkan oleh perusahaan Rapid-I yang didirikan oleh Ingo Mierswa dan Ralf Klinkenberg. Pada tahun 2007, nama perangkat lunak ini diubah dari YALE menjadi RapidMiner. RapidMiner dirancang untuk mendukung seluruh proses ilmu data, termasuk:

1. Persiapan data untuk memuat dan mentransformasi data (ETL), prapemrosesan, dan visualisasi data.
2. Pembelajaran mesin dan analisis prediktif untuk membangun dan mengevaluasi model statistik untuk analisis prediktif.
3. Evaluasi dan implementasi model untuk menilai kinerja model dan menerapkannya dalam lingkungan produksi.

RapidMiner menggunakan antarmuka pengguna grafis (GUI) untuk merancang dan menjalankan alur kerja analitik, yang disebut "Proses", terdiri dari berbagai "Operator" yang masing-masing melakukan tugas tertentu dalam proses tersebut. [20]



Gambar 2.3. Tampilan Aplikasi Altair AI Studio RapidMiner 2025.0.0

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil Prediksi Regresi Linear

Proses diawali dengan mencari sumber *dataset* tentang faktor kesehatan. Pada penelitian ini menggunakan data yang berasal dari <https://www.kaggle.com/>. *Dataset* yang didapat sudah berbentuk format .csv. Kemudian langkah berikutnya adalah membuka aplikasi RapidMiner, diawali dengan memilih operator *Read CSV* untuk memuat data yang telah dipilih. Lalu klik *Import Configuration Wizard* dan pilih data yang akan diolah. Saat *file* dipilih akan muncul pengaturan atribut tabel.

	Age	BMI	Exercise_F...	Diet_Quality	Sleep_Hours	Smoking_S...	Alcohol_Co...	Health_Score
	real	real	integer	real	real	integer	real	real/label
1	45.961	31.997	5	56.403	7.300	0	2.835	70.542
2	38.341	29.823	6	41.838	7.012	1	7.200	57.245
3	47.772	25.298	5	76.905	8.029	1	4.098	96.334
4	58.276	21.765	2	49.757	5.803	1	3.648	61.322
5	37.190	28.491	2	44.218	7.013	0	2.840	67.176
6	37.190	26.967	2	89.125	7.786	1	1.594	100.000
7	58.951	20.476	5	81.523	7.086	0	3.623	91.386
8	48.209	28.178	3	85.902	8.388	0	0.235	94.854
9	34.366	30.248	4	67.889	4.058	1	0.327	78.356
10	46.511	22.324	4	33.890	6.153	0	5.025	64.380
11	34.439	31.587	1	63.751	5.266	1	-0.524	79.974
12	34.411	25.988	4	58.609	10.384	1	2.361	84.248
13	42.904	36.375	4	74.954	8.070	0	0.802	85.010
14	17.041	21.654	2	32.672	6.881	1	2.195	69.051
15	19.301	33.680	2	84.664	8.177	1	3.368	91.401
16	33.253	25.990	5	58.707	8.674	1	4.968	90.400
17	27.846	21.743	1	49.359	8.658	1	5.435	75.313
18	40.758	30.654	3	54.550	6.863	0	8.415	78.865

Gambar 3.1.1. Mengubah Atribut pada Tabel

Gambar 3.1 mengubah atribut pada kolom *Health\_Score* karena kolom tersebut adalah variabel dependen yang akan diprediksi. Kemudian menambahkan operator *Split Data*, operator ini digunakan untuk membagi dataset menjadi 2 bagian yaitu *data training* dan *data testing*. *Data training* digunakan untuk mempelajari pola pada *machine learning*, sementara *data testing* digunakan untuk menguji pola yang sudah ditemukan. [21] Lalu hubungkan *out* pada operator *Read CSV* dengan *exa* pada operator.



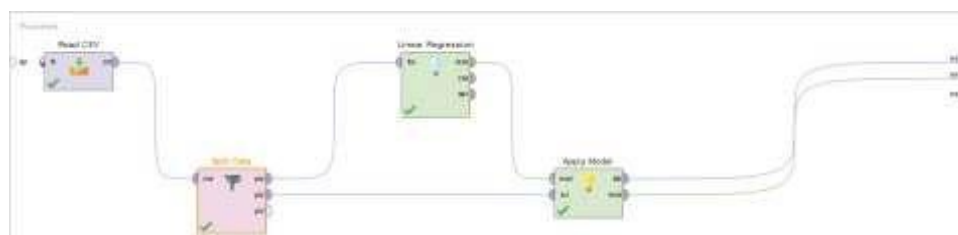
Gambar 3.1.2. Membagi Data Menjadi Data Training dan Testing

Kemudian untuk membagi data yang sesuai, pilih edit *numeration*. Pada gambar 3.2 menunjukan *data training* sebesar 90% dan sisanya adalah *data testing*. Kemudian menambahkan operator *Linear Regression*. Pada garis *par* 1 dihubungkan dengan *tra* pada *Linear Regression* dan *mod* dengan *ras* sehingga menghasilkan *Linear Regression* seperti berikut:



Gambar 3.1.3. Hasil Regresi Linear

Tahapan selanjutnya adalah menghubungkan *port* yang ada seperti pada gambar berikut:



Gambar 3.1.4. Melakukan Pengujian Data Testing

Row No.	Health_Score	prediction(H...	Age	BMI	Exercise_Fr...	Diet_Quality	Sleep_Hours	Smoking_St...	Alcohol_Con...
1	79.974	82.239	34.439	31.587	1	83.751	5.266	1	-0.524
2	94.248	89.782	34.411	25.988	4	59.609	10.384	1	2.361
3	75.313	78.566	27.848	21.743	1	49.359	8.658	1	5.435
4	98.855	89.669	43.771	22.581	0	64.906	8.383	0	1.162
5	100	96.823	44.508	25.600	4	71.704	9.746	0	2.943
6	100	102.603	22.258	24.137	0	82.841	10.461	1	0.854

Gambar 3.1.5. Hasil Pengujian Data Testing

Berdasarkan pada hasil prediksi yang diberikan RapidMiner terhadap *data testing* pada prediksi *Health\_Score*, sedangkan yang lain merupakan kolom dari faktor yang mempengaruhi skor kesehatan. Pada penelitian ini juga akan melakukan penilaian dari model regresi linear dengan menggunakan operator *performance*.

**root\_mean\_squared\_error**

root\_mean\_squared\_error: 5.495 +/- 0.000

Gambar 3.1.6. Hasil dari Pengujian Performance

Hasil dari operator *performance* dapat dilihat pada Gambar 3.1.6 yang mana nilai RSME dapat dipengaruhi oleh beberapa faktor seperti kurangnya *data training*, kesalahan dalam ansumsi jika variabel independen dan dependen dapat dijelaskan secara lurus.

Pada penelitian ini jua dilakukan pengujian mengubah *jumlah data training* untuk mengukur nilai RMSE, seperti pada tabel berikut:

Data Training	Data Testing	Nilai RMSE
90%	10%	5.495 +/- 0.000
80%	20%	5.821 +/- 0.000
70%	30%	5.837 +/- 0.000
60%	40%	5.649 +/- 0.000

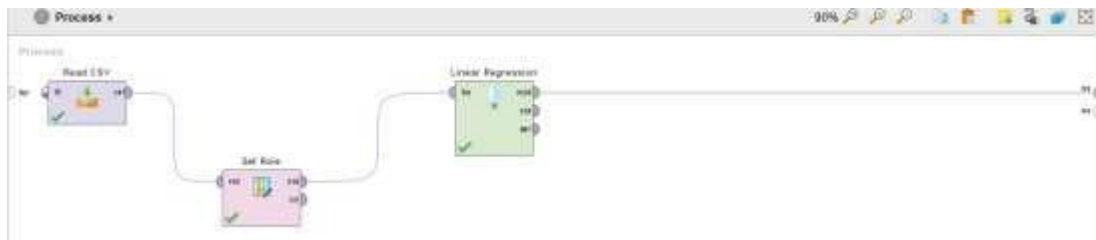
Tabel 1 Hasil Pengujian dengan Mengatur Jumlah Data Training dan Data Testing

Skenario dengan *data training* 90% dan *data testing* 10% memiliki nilai RSME terendah (5.495), menunjukan tingkat akurasi yang lebih tinggi dibandingkan skenario lainnya. Dan skenario dengan *data training* 70% dan *data testing* 30% memiliki nilai RSME yang paling tinggi (5.837). Maka dalam kasus ini, skenario optimal nampaknya adalah dengan 90% *data training* dan 10% *data testing* yang mana memberikan nilai RSME terendah dan realisitis berdasarkan data prediksi, hampir semua prediksi mendekati nilai sebenarnya meski ada beberapa yang meleset, namun ini sudah menunjukan kinerja model yang sangat baik dan dapat diandalkan.



### 3.2. Hasil Analisis Regresi Linear

Langkah awal untuk melakukan analisis regresi linear sama dengan proses prediksi diawal, hanya setelah memasukkan dataset, ambil operator *Set Role* kemudian hubungkan ke *Linear Regression*.



Gambar 3.2.1. Proses Analisis Regresi Linear

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	1.Sat	p-Value	Coll
Age	-0.238	0.019	-0.206	1.000	-15.925	0	****
BMI	-1.132	0.035	-0.814	1.000	-31.906	0	****
Exercise_Frequency	1.781	0.088	0.261	1.000	20.121	0	****
Diet_Quality	0.606	0.012	0.885	0.999	51.338	0	****
Sleep_Hours	2.427	0.116	0.270	1.000	20.832	0	****
Smoking_Status	-3.640	0.353	-0.124	0.999	-10.306	0	****
Alcohol_Consumption	-0.932	0.088	-0.142	1.000	-11.000	0	****
Intercept	64.005	1.888	?	?	33.415	0	****

Gambar 3.2.2. Hasil Analisis Regresi Linear

Berdasarkan hasil analisis regresi linear terhadap faktor-faktor gaya hidup yang memengaruhi skor kesehatan individu, diperoleh model yang menunjukkan pengaruh signifikan dari seluruh variabel independen terhadap skor kesehatan sebagai variabel dependen. Seluruh nilai p (*p-value*) untuk masing-masing variabel berada pada angka 0.000, yang berarti bahwa setiap variabel memiliki pengaruh yang signifikan secara statistik terhadap skor kesehatan ( $\alpha < 0,05$ ).

Variabel	Koefisien	Interpretasi Singkat
<i>Age</i>	-0.238	Semakin tua usia, skor kesehatan cenderung menurun.
<i>BMI</i>	-1.132	Semakin tinggi BMI, skor kesehatan menurun.
<i>Exercise_Frequency</i>	+1.781	Semakin sering olahraga, skor kesehatan meningkat.
<i>Diet_Quality</i>	+0.606	Kualitas diet lebih baik, skor kesehatan naik.
<i>Sleep_Hours</i>	+2.427	Tidur lebih lama, skor kesehatan meningkat signifikan.
<i>Smoking_Status</i>	-3.640	Merokok secara signifikan menurunkan skor kesehatan.
<i>Alcohol_Consumption</i>	-0.932	Konsumsi alkohol juga menurunkan skor kesehatan.
<i>Intercept</i>	+64.005	Skor kesehatan dasar saat semua variabel = 0 (nilai awal model).

Tabel 2. Hasil Interpretasi Variabel

1. Usia (*Age*) memiliki koefisien negatif sebesar -0.238, yang menunjukkan bahwa semakin bertambah usia seseorang, maka skor kesehatannya cenderung menurun. Hal ini selaras dengan literatur yang menyatakan bahwa usia merupakan faktor risiko terhadap munculnya penyakit kronis.
2. Indeks Massa Tubuh (BMI) juga menunjukkan pengaruh negatif dengan koefisien -1.132. Artinya, semakin tinggi nilai BMI seseorang, maka semakin rendah skor kesehatannya. Temuan ini mengindikasikan bahwa obesitas atau berat badan berlebih berdampak negatif terhadap kesehatan.

3. Frekuensi Olahraga (*Exercise\_Frequency*) memiliki pengaruh positif yang cukup kuat dengan koefisien +1.781, yang menandakan bahwa semakin sering seseorang berolahraga, maka skor kesehatannya cenderung meningkat.
4. Kualitas Diet (*Diet\_Quality*) juga berkontribusi secara positif terhadap kesehatan, dengan koefisien +0.606. Diet yang seimbang dan bergizi mendukung peningkatan skor kesehatan secara signifikan.
5. Durasi Tidur (*Sleep\_Hours*) memberikan pengaruh positif yang paling tinggi dengan koefisien +2.427. Tidur yang cukup dan berkualitas merupakan komponen penting dalam menjaga kebugaran dan sistem imun.
6. Status Merokok (*Smoking\_Status*) berpengaruh negatif cukup besar dengan koefisien -3.640, yang menunjukkan bahwa perokok cenderung memiliki skor kesehatan yang jauh lebih rendah dibandingkan non-perokok.
7. Konsumsi Alkohol (*Alcohol\_Consumption*) juga memiliki koefisien negatif sebesar -0.932, menandakan bahwa konsumsi alkohol turut menurunkan skor kesehatan.

Hasil analisis menunjukkan bahwa gaya hidup memiliki peran yang sangat besar dalam menentukan tingkat kesehatan seseorang. Faktor-faktor seperti pola tidur, olahraga, dan pola makan yang baik terbukti memberikan kontribusi positif yang signifikan terhadap skor kesehatan. Sebaliknya, kebiasaan merokok dan konsumsi alkohol secara nyata menurunkan kualitas kesehatan.

Temuan ini sejalan dengan berbagai penelitian sebelumnya yang menyebutkan bahwa perilaku gaya hidup sehat berbanding lurus dengan peningkatan status kesehatan secara umum [22];(WHO, 2022). Oleh karena itu, penting bagi individu untuk memperhatikan aspek-aspek gaya hidup yang dapat dimodifikasi demi peningkatan kualitas hidup jangka panjang.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai prediksi dan analisis faktor gaya hidup terhadap skor kesehatan menggunakan regresi linear di RapidMiner, dapat disimpulkan hal-hal berikut:

1. Model regresi linear yang dibangun melalui RapidMiner berhasil memprediksi skor kesehatan individu berdasarkan faktor-faktor gaya hidup seperti usia, indeks massa tubuh (BMI), frekuensi olahraga, kualitas diet, jam tidur, status merokok, dan konsumsi alkohol. Hasil pengujian model menunjukkan bahwa konfigurasi *data training* sebesar 90% dan *data testing* sebesar 10% menghasilkan nilai RMSE (*Root Mean Square Error*) terendah, yaitu 5.495, yang menandakan bahwa model memiliki tingkat akurasi prediksi yang baik.
2. Seluruh variabel independen memiliki pengaruh yang signifikan terhadap skor kesehatan, dengan *p-value* sebesar 0.000 pada masing-masing variabel, sehingga secara statistik layak dimasukkan dalam model.
3. Penelitian ini membuktikan bahwa pendekatan berbasis data dan teknologi seperti regresi linear dalam RapidMiner dapat digunakan secara efektif sebagai sistem pendukung keputusan dalam bidang kesehatan, khususnya untuk memahami dan memetakan pengaruh gaya hidup terhadap status kesehatan seseorang.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah berkontribusi dalam penyelesaian artikel ini. Ucapan terima kasih disampaikan kepada:

1. Bapak Muhammad Thoriq, S.Kom., M.Kom selaku dosen pengampu mata kuliah *Machine Learning* di Program Studi Informatika, Universitas Adzkia, yang telah memberikan arahan dan bimbingan selama proses penyusunan artikel ini
2. Seluruh tim pengajar dan staf Universitas Adzkia yang telah menyediakan fasilitas dan dukungan akademik yang sangat berarti;
3. Rekan-rekan mahasiswa yang telah memberikan masukan, saran, serta semangat selama proses penelitian ini;
4. Dan kepada keluarga yang senantiasa memberikan dukungan moral dan motivasi tanpa henti.

Semoga artikel ini dapat memberikan kontribusi positif dalam pengembangan ilmu pengetahuan, khususnya dalam penerapan metode regresi linear di bidang kesehatan.

## REFERENSI

- [1] H. Kurniawanto, M. A. Anggriawan, and T. Rustandi, “*Health Outcomes* Masyarakat: Mengukur Seberapa Penting Peran BPJS Kesehatan di Kota Serang,” *Al Qalam J. Ilm. Keagamaan dan Kemasyarakatan*, vol. 17, no. 6, p. 4284, 2023, doi: 10.35931/aq.v17i6.2828.
- [2] N. Zahri and N. A. A. Syifaa, “*Hubungan Antara Gaya Hidup Dan Meningkatnya Insiden Penyakit Di Kalangan Masyarakat : Analisis Terhadap Pola Makan , Aktivitas Fisik , Dan Pola Tidur,*” vol. 9, no. 3, pp. 153–162, 2025.
- [3] S. A. Putri and T. Dahlan, “Inovasi Teknologi Wearable Untuk Pemantauan Kesehatan Mandiri: Dampaknya Terhadap Peningkatan Gaya Hidup Sehat di Masyarakat,” *J. Kesehat. Masy. dan Inov. Teknol. Kesehat.*, vol. 1, no. 1, pp. 9–12, 2024.
- [4] A. Kurniadi Hermawan, A. Nugroho, and Edora, “Analisa Data Mining Untuk Prediksi Penyakit Ginjal Kronik Dengan Algoritma Regresi Linier,” *Bull. Inf. Technol.*, vol. 4, no. 1, pp. 37–48, 2023, doi: 10.47065/bit.v4i1.475.
- [5] A. Pengaruh, P. Mobile, T. Loyalitas, and K. Kartu, “1561-Article Text-4293-1-10-20220411,” vol. 8, no. April 2022, 2017, doi: 10.5281/zenodo.6446981.
- [6] S. Alfarisi, R. Astuti, and F. M. Basysyar, “Implementasi Data Mining Menggunakan Algoritma Regresi Linear Untuk Prediksi Penjualan Dan Cashflow Di Ayam Geprek Cap Cangkir,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3392–3395, 2024, doi: 10.36040/jati.v8i3.9717.
- [7] E. C. Sitohang, F. E. Ginting, and Y. M. B. Sembiring, “Prediksi Jumlah Perokok dan Dampaknya terhadap Kesehatan Masyarakat Menggunakan Regresi Linear,” *Semin. Nas. Inov. Sains Teknol. Inf. Komput.*, vol. 1, no. 2, pp. 512–516, 2024.
- [8] U. Nurul, “PENERAPAN REGRESI LINEAR UNTUK PREDIKSI SUHU BADAN SAPI MENGGUNAKAN DATA SMART TERNAK DARI PT TELKOM INDONESIA: Studi ...,” vol. 6, no.

- 2, 2024, [Online]. Available: [http://eprints.unram.ac.id/44500/2/Jurnal Tugas Akhir\\_F1D020078\\_Nurul Umami.pdf](http://eprints.unram.ac.id/44500/2/Jurnal_Tugas_Akhir_F1D020078_NurulUmami.pdf)
- [9] Joni Maulindar and Y. Istiqomah, "Analisis Faktor Ekonomi Dan Sosial Dalam Kejadian Stunting Pada Anak Dengan Metode Regresi Linear," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 2, pp. 310–317, 2024, doi: 10.29408/jit.v7i2.24022.
- [10] M. F. I. Kolibu, N. Nainggolan, and Y. A. R. Langi, "Analisis Faktor-faktor yang Mempengaruhi Harga Cabai Merah di Kota Manado Provinsi Sulawesi Utara Menggunakan Analisis Regresi Linear Berganda," *J. MIPA*, vol. 13, no. 1, pp. 32–36, 2024, doi: 10.35799/jm.v13i1.52258.
- [11] A. A. Rodiana, "Faktor-Faktor Yang Mempengaruhi Indeks Massa Tubuh Pada Remaja Di SMK PGRI 1 Kota Sukabumi," *Sekol. Tinggi Ilmu Kesehat. Sukabumi*, vol. 5, no. 01, pp. 78–86, 2022, doi: 10.34305/jmc.v5i1.1315.
- [12] P. Delyana, Pratiwi, E. Rokhmia, and G. Istiani, "Hubungan Umur Dan Jenis Kelamin Dengan Risiko Penyakit Tidak Menular (Ptm) Berdasarkan Data Skrining Kesehatan Bpjs Jakarta Selatan Tahun 2022," *J. Keperawatan dan Kesehat. Masy. Cendekia Utama*, vol. 13, no. 1, pp. 32–43, 2024, [Online]. Available: <https://www.jurnal.stikeskendekiautamakudus.ac.id/index.php/stikes/article/view/1460>
- [13] A. R. Ramadhania, A. N. Hasna, and R. K. Winata, "Hubungan Aktivitas Fisik dan Pola Makan terhadap Status Indeks Masa Tubuh Normal," vol. 3, no. 1, pp. 58–66, 2024.
- [14] R. Between, E. Frequency, E. W. Ability, A. R. Alfandary, S. Mulyani, and H. Subekti, "JURNAL," vol. 7, no. 6, pp. 1244–1256, 2024.
- [15] D. Briawan, M. F. Heryanda, and S. Sudikno, "Kualitas diet dan kontrol glikemik pada orang dewasa dengan diabetes melitus tipe dua," *J. Gizi Klin. Indones.*, vol. 18, no. 1, p. 8, 2021, doi: 10.22146/ijcn.62815.
- [16] C. Gao *et al.*, "Sleep Duration/Quality With Health Outcomes: An Umbrella Review of Meta-Analyses of Prospective Studies," *Front. Med.*, vol. 8, no. January, pp. 1–15, 2022, doi: 10.3389/fmed.2021.813943.
- [17] R. Growth, D. In, S. Rural, and A. At, "Original Research Paper Nursing," vol. 8, no. 2277, pp. 16–17, 2022.
- [18] A. Wijayadhi, M. Makmun Effendi, and S. Budi Rahardjo, "Prediksi Penyakit Jantung Dengan Algoritma Regresi Linier," *Bull. Inf. Technol.*, vol. 4, no. 1, pp. 15–28, 2023, doi: 10.47065/bit.v4i1.463.
- [19] M. Yusuf Alwy, Herman, T. H. A. Abraham, and H. Rukmana, "Analisis Regresi Linier Sederhana dan Berganda Beserta Penerapannya," *J. Educ.*, vol. 06, no. 02, pp. 13331–13344, 2024.
- [20] A. S. Ramadhantya, "Penggunaan Rapidminer Untuk Memprediksi Kelulusan Mahasiswa Dengan Algorithm Naive Bayes," vol. 10, no. 1, pp. 52–60, 2024.
- [21] P. Chang Hartono and A. Dwiyo Widianoro, "Analisis Prediksi Harga Saham Unilever Menggunakan Regresi Linier dengan RapidMiner," *J. Comput. Inf. Syst. Ampera*, vol. 5, no. 3, pp. 2775–2496, 2024, [Online]. Available: <https://journal-computing.org/index.php/journal-cisa/index>
- [22] J. S. Malo, M. H. Schafer, and A. J. Stull, "Healthy eating in life course context: Asymmetric implications of socioeconomic origins and destinations," *Soc. Sci. Med.*, vol. 372, no. October 2024, 2025, doi: 10.1016/j.socscimed.2025.117936.

