

Model Regresi Linear Berganda Untuk Prediksi Nilai Kualitas Tidur Berdasarkan Gaya Hidup

Sidarta David Setia¹, Masparudin ^{2*}, Kaharuddin ³, Musliadi KH ⁴

^{1,2,3}Teknik Perangkat Lunak, Universitas Universal

⁴Teknik Informatika, Universitas Universal

*Corresponding author E-mail: masparudin.mahmud@gmail.com

Article Info

Article history:

Received 05-07-2026

Revised 26-07-2026

Accepted 30-07-2026

Keyword:

Quality of Sleep, Multiple Linear Regression, Machine Learning, Lifestyle, Model Interpretation.

ABSTRACT

Sleep disorders significantly impact individuals' quality of life and chronic health. Previous research utilizing the Sleep Health and Lifestyle Dataset predominantly focused on Classification tasks—predicting discrete diagnostic categories such as Insomnia or Sleep Apnea. Although classification accuracy has been high, this method fails to provide a continuous, quantitative assessment of the severity of Quality of Sleep (QOS). This study aims to address this limitation by developing and interpreting a Multiple Linear Regression (MLR) model to predict the numeric Quality of Sleep (QOS) score on a 1-10 scale based on lifestyle, demographic, and biometric factors. The MLR model was applied following pre-processing, which included One-Hot Encoding for categorical variables and the removal of diagnostic variables to prevent data leakage. Evaluation results demonstrate that the model achieved excellent performance, confirmed by a high Coefficient of Determination (R^2) of 0.957 and a very low Mean Absolute Error (MAE) of 0.145 units. Quantitative analysis of the regression coefficients identified Sleep Duration as the most dominant positive predictor and Stress Level as the most significant negative predictor of QOS. These findings provide an important contribution in the form of an interpretable mathematical equation, which can be utilized by clinicians to make measurable, evidence-based intervention recommendations, shifting the focus from diagnosis to quantitative management and prevention.



Copyright © 2026. This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.

I. PENDAHULUAN

Kualitas tidur (Quality of Sleep atau QoS) merupakan hal yang penting bagi kesehatan fisik dan mental. Gangguan kualitas tidur adalah isu yang sering terjadi di masyarakat yang dapat dipengaruhi kuat oleh pola hidup, peningkatan tingkat stres, dan aktivitas fisik yang minim. Permasalahan ini tidak hanya mengganggu fungsi kognitif, tetapi juga berkorelasi erat dengan peningkatan risiko penyakit kronis seperti hipertensi dan diabetes [1], [2], [3]. Beberapa literatur klinis secara konsisten menyoroti bahwa modifikasi gaya hidup, seperti kebugaran jasmani melalui aktivitas fisik [4], [5] serta manajemen tingkat stres harian [6], [7], memiliki hubungan yang kuat terhadap fluktuasi kualitas tidur pasien. Oleh karena itu, identifikasi faktor yang

memengaruhi QOS secara kuantitatif sangat penting untuk intervensi kesehatan yang efektif.

Dalam ranah data mining, data terkait kesehatan tidur dan gaya hidup telah menjadi subjek dari beberapa penelitian terdahulu yang berfokus dominan pada tugas klasifikasi, menggunakan algoritma seperti Random Forest, K-Nearest Neighbors, Support Vector Machine (SVM), hingga Neural Network [8], [9], [10], [11], [12]. Model klasifikasi tersebut, meskipun akurat (mencapai 84% - 89,69%) dalam memprediksi kategori diagnostik diskrit seperti Insomnia atau Sleep Apnea, tidak dirancang untuk penilaian tingkat keparahan kualitas tidur yang berkelanjutan.

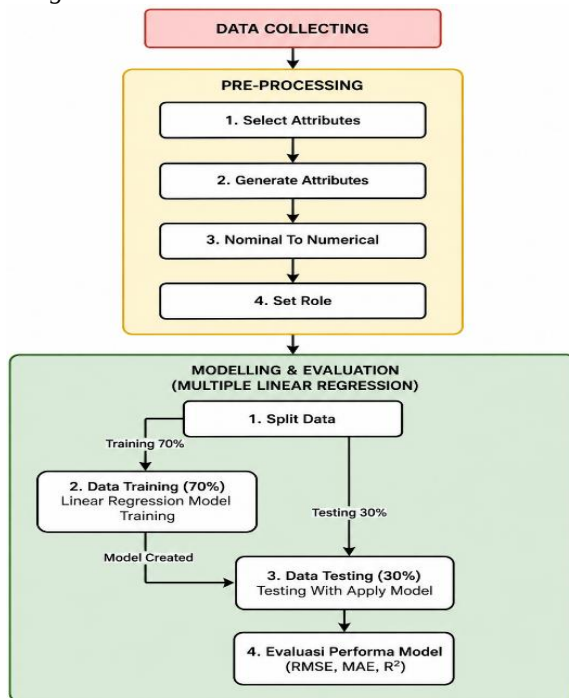
Fokus pada klasifikasi memiliki keterbatasan dalam konteks pencegahan dan manajemen kesehatan. Model ini tidak

memberikan penilaian kuantitatif terhadap tingkat keparahan kualitas tidur pada skala numerik (1-10) [13]. Hal ini dapat menciptakan kesenjangan dalam upaya menyediakan intervensi berbasis bukti, di mana klinisi membutuhkan informasi mengenai sejauh mana setiap faktor gaya hidup memengaruhi skor QOS seseorang.

Untuk mengatasi keterbatasan tersebut, penelitian ini bertujuan untuk mengusulkan pengembangan dan interpretasi Model Regresi Linear Berganda (MLR) [14]. Model MLR bertujuan untuk memprediksi nilai numerik dari Quality of Sleep (QOS) pada skala 1–10. Pendekatan regresi dipilih karena kemampuannya dalam menyediakan persamaan matematis yang dapat diinterpretasikan secara kuantitatif, yang sangat diperlukan dalam domain kesehatan [14].

II. METODE

Metode pada penelitian ini secara umum terdiri dari 3 tahapan utama yaitu *Data Collecting*, *Pre-Processing* dan *Modelling & Evaluation*.



Gambar 1. Alur Penelitian

A. Tahap Pengumpulan dan Input Data (Data Collecting)

Alur penelitian diawali dengan fase Data Collecting, di mana dataset sekunder Sleep Health and Lifestyle yang diperoleh dari repositori Kaggle disiapkan. Eksplorasi data mentah ini diintegrasikan ke dalam sistem menggunakan operator Read CSV pada RapidMiner. Operator ini bertugas untuk memetakan struktur data awal, mengidentifikasi tipe data bawaan (seperti text, integer, atau real), serta memastikan bahwa pemisah kolom (delimiter) terbaca dengan presisi sehingga tidak terjadi pergeseran nilai antar-variabel sebelum masuk ke komputasi berikutnya.

B. Tahap Pra-pemrosesan Data (Pre-Processing)

Setelah dataset berhasil dimuat, alur berlanjut ke tahap pra-pemrosesan yang krusial untuk menjamin kualitas data (*data cleansing*) dan kesesuaian format algoritma. Tahapan pra-pemrosesan ini dibagi menjadi empat sub-proses yang saling berurutan:

- 1) **Penyaringan Fitur (*Select Attributes*):** Operator Select Attributes diterapkan untuk mereduksi dimensi data yang tidak relevan. Pada langkah ini, atribut Person ID dan Sleep Disorder secara sengaja dieliminasi dari ruang fitur. Eksklusi Person ID dilakukan karena atribut tersebut hanya berupa indeks pengenalan unik yang tidak memiliki signifikansi statistik terhadap variabilitas kualitas tidur. Sementara itu, penghapusan Sleep Disorder dilakukan untuk meminimalisasi risiko kebocoran data (*data leakage*), karena keberadaan label gangguan tidur secara langsung dapat mendistorsi objektivitas model dalam memprediksi skala kualitas tidur murni.
- 2) **Rekayasa Atribut (*Generate Attributes*):** Atribut Blood Pressure pada dataset asli bertipe kategorikal/string dengan format fraksi tekanan darah (misalnya "120/80"). Karena algoritma regresi tidak dapat memproses format teks tersebut, operator Generate Attributes digunakan untuk melakukan ekstraksi matematis. Melalui fungsi transformasi berbasis string pembagi (separator "/"), variabel tersebut dipecah dan ditransformasikan menjadi dua atribut numerik baru yang independen, yaitu tekanan darah sistolik (Systolic) dan tekanan darah diastolik (Diastolic).
- 3) **Konversi Tipe Data (*Nominal to Numerical*):** Dataset ini masih mengandung variabel kategorikal teks seperti Gender, Occupation, dan BMI Category. Untuk menjembatani kebutuhan algoritma Regresi Linear Berganda yang mensyaratkan seluruh masukan berupa matriks numerik, digunakan operator Nominal to Numerical. Operator ini mengeksekusi teknik One-Hot Encoding, yang bekerja dengan mengubah setiap nilai unik dari variabel nominal menjadi kolom baru berskala biner (nilai 0 atau 1), sehingga interaksi antarkategori dapat dipahami secara matematis oleh komputer.
- 4) **Penetapan Peran Variabel (*Set Role*):** Sub-proses terakhir dalam pra-pemrosesan adalah mendefinisikan orientasi variabel menggunakan operator Set Role. Di sini, atribut Quality of Sleep (QOS) secara eksplisit diubah perannya menjadi label (variabel dependen target yang akan diprediksi). Sedangkan seluruh variabel independen tersisa (seperti usia, durasi tidur,

tingkat stres, langkah harian, serta variabel hasil pecahan sistolik dan diastolik) secara otomatis ditetapkan perannya sebagai regular (fitur prediktor).

C. Tahap Pemodelan dan Evaluasi (Modelling & Evaluation)

Fase akhir dari alur penelitian ini berfokus pada konstruksi model matematis dan pengujian keandalannya. Proses pemodelan dipisahkan secara ketat untuk menjamin validitas pengujian melalui rangkaian tahapan berikut:

- 1) Pembagian Data (*Split Data*): Aliran data dari tahap pra-pemrosesan masuk ke operator Split Data untuk dibagi menjadi dua subset yang saling bebas (mutually exclusive). Pembagian dilakukan secara acak menggunakan rasio 70% (0,7) yang dialokasikan sebagai data latih (Training 70%) dan 30% (0,3) sisanya dialokasikan sebagai data uji (Testing 30%).
- 2) Pelatihan Model (*Linear Regression Model Training*): Subset data latih sebesar 70% dialirkan menuju sub-proses Data Training untuk melatih model menggunakan algoritma Regresi Linear Berganda. Demi menghasilkan estimasi parameter yang kokoh, konfigurasi parameter diatur secara ketat, meliputi: pemilihan fitur berbasis metode M5 Prime, nilai toleransi minimum sebesar 0,05, dan konstanta regularisasi Ridge sebesar $1.0E-8$ untuk mencegah terjadinya overfitting. Selain itu, fungsi use bias diaktifkan untuk menghitung konstanta intersep, serta opsi eliminate colinear features dijalankan guna mendeteksi sekaligus menghapus fitur yang mengalami multikolinieritas parah. Output dari fase pelatihan ini menghasilkan sebuah model prediksi yang matang (Model Created).
- 3) Pengujian Model (*Testing with Apply Model*): Pada tahapan ini, model yang telah terbentuk (Model Created) dialirkan dan dipertemukan dengan subset data uji sebesar 30% (Testing 30%) di dalam operator Apply Model. Proses ini menantang model untuk memprediksi nilai *Quality of Sleep* pada data uji yang sebelumnya tidak pernah dilihat atau dipelajari oleh model selama proses pelatihan, sehingga kemampuan generalisasi model dapat diukur secara objektif.
- 4) Evaluasi Performa Model: Alur pemodelan ditutup dengan pengukuran tingkat akurasi pada operator Performance (Regression). Evaluasi keandalan prediksi diukur secara komprehensif berdasarkan tiga metrik utama, yaitu *Root Mean Squared Error* (RMSE) untuk mengetahui deviasi rata-rata kuadrat dari error prediksi, *Absolute Error* (MAE) untuk

menghitung rata-rata absolut jarak kesalahan prediksi terhadap nilai riil, serta *Squared Correlation* (Koefisien Determinasi atau R^2) untuk mengukur persentase variabilitas kualitas tidur yang berhasil dijelaskan secara linear oleh kombinasi seluruh fitur prediktor dalam model.

D. Persamaan Model

Hubungan kausalitas dan interaksi linear antar-variabel dalam penelitian ini dimodelkan menggunakan pendekatan statistik multivariat melalui algoritma Regresi Linear Berganda. Di dalam alur komputasi, rumusan matematis ini secara aktif diterapkan oleh sistem pada fase pengujian (Testing with Apply Model) menggunakan bobot parameter yang telah dioptimalkan pada fase pelatihan (Model Training). Secara teoretis, struktur model estimasi untuk memprediksi nilai kualitas tidur (Quality of Sleep) diformulasikan melalui Persamaan (1):

$$QOS = \beta_0 + \sum_{i=1}^n \beta_i X_i + \epsilon \quad (1)$$

Dimana

- QOS merepresentasikan nilai estimasi kualitas tidur sebagai variabel dependen target yang diukur pada skala 1 hingga 10
- β_0 mewakili konstanta atau titik potong intersep, yaitu nilai estimasi kualitas tidur ketika seluruh variabel prediktor bernilai nol
- β_i menunjukkan koefisien regresi atau bobot kontribusi untuk masing-masing variabel independen ke- i .
- X_i merupakan nilai riil dari fitur-fitur prediktor (seperti *Age*, *Sleep Duration*, *Stress Level*, *Daily Steps*, serta fitur hasil ekstraksi *Systolic* dan *Diastolic*).
- n menyatakan jumlah keseluruhan dimensi atribut independen yang lolos seleksi fitur.
- ϵ menggambarkan komponen *residual error* atau variansi kesalahan acak yang tidak dapat dijelaskan oleh model.

Persamaan umum (1) ini berfungsi sebagai landasan matematis awal pada metode penelitian. Nilai riil dari konstanta (β_0) dan koefisien bobot (β_i) akan diekstraksi secara otomatis berdasarkan hasil minimisasi *error* oleh operator *Linear Regression* di RapidMiner, yang kemudian akan dianalisis dan dibahas secara komprehensif pada bagian hasil dan pembahasan

III. HASIL DAN PEMBAHASAN

Pada bab ini dipaparkan hasil-hasil empiris dari setiap tahapan komputasi *data mining* yang telah dieksekusi menggunakan RapidMiner Studio, diikuti dengan pembahasan mendalam mengenai performa model Regresi

Linear Berganda (*Multiple Linear Regression*) dalam memprediksi kualitas tidur (*Quality of Sleep*).

A. Hasil Tahap Pra-pemrosesan Data (Pre-Processing)

Tahap pra-pemrosesan data memegang peranan krusial untuk mentransformasikan dataset mentah *Sleep Health and Lifestyle* agar memenuhi asumsi serta persyaratan matematis dari algoritma Regresi Linear Berganda. Melalui rekonstruksi fitur yang ketat, diperoleh beberapa hasil signifikan pada sub-proses berikut:

1. **Pembersihan Data (*Data Cleansing*):** Dataset awal yang berjumlah 374 baris dianalisis secara menyeluruh untuk mendeteksi anomali data. Hasil inspeksi menunjukkan bahwa tidak ditemukan adanya nilai yang hilang (*missing values*) maupun data duplikat pada seluruh baris dan kolom. Oleh karena itu, integritas ukuran sampel asli sebanyak 374 data dapat dipertahankan sepenuhnya tanpa memerlukan tindakan reduksi sampel.
2. **Penyaringan dan Eksklusi Variabel Diagnostik:** Sejalan dengan desain metodologi untuk menghindari *data leakage*, atribut *Person ID* dan *Sleep Disorder* resmi dieliminasi dari ruang fitur. Langkah penapisan ini memastikan bahwa model prediktif kualitas tidur (*QOS*) yang dibangun murni didasarkan pada interaksi faktor gaya hidup, demografi, dan biometrik harian pengguna, bukan dari bias penanda klinis gangguan tidur.
3. **Rekayasa Atribut Tekanan Darah:** Variabel *Blood Pressure* yang semula bertipe string (teks) berhasil diekstraksi secara matematis. Melalui fungsi pembagi string, komponen fraksi tekanan darah dipecah secara akurat menjadi dua variabel numerik kontinu yang baru, yaitu tekanan darah sistolik (*Systolic*) dan tekanan darah diastolik (*Diastolic*). Transformasi ini memungkinkan aspek hemodinamik tubuh subjek terbaca sebagai parameter independen yang valid oleh algoritma regresi.
4. **Transformasi Variabel Kategorikal (*One-Hot Encoding*):** Untuk mengonversi variabel non-numerik ke dalam bentuk matriks biner yang dapat dipahami oleh mesin, dilakukan prosedur *Nominal to Numerical* melalui *One-Hot Encoding*. Variabel *Occupation* yang memiliki 11 kategori unik ditransformasikan secara linear menjadi 10 variabel biner. Variabel *Gender* (2 kategori) dikonversi menjadi 1 variabel biner penanda dikotomi. Sementara itu, variabel *BMI Category* yang memiliki 3 tingkatan kategori dikompresi menjadi 2 variabel biner independen.

Melalui serangkaian rekonstruksi pra-pemrosesan di atas, ruang fitur konvergen secara optimal. Setelah proses eliminasi dan ekspansi biner selesai, jumlah total atribut

independen (prediktor) final yang siap diumpankan ke dalam arsitektur pelatihan model adalah sebanyak 9 atribut regular.

B. Hasil Eksperimen dan Evaluasi Performa Regresi Linear

Setelah fase persiapan data selesai, data divalidasi menggunakan skenario *Split Data* berasio 70:30, di mana model dilatih pada fase *Linear Regression Model Training* dan diuji pada fase *Testing with Apply Model*. Evaluasi keandalan dan akurasi prediksi model regresi linear ini diukur secara kuantitatif melalui tiga metrik komparatif utama yang disajikan secara terperinci pada Tabel 1.

Tabel 1. Perbandingan Metrik Evaluasi Performa Model Regresi Linear Berganda

Metrik Evaluasi	Nilai Hasil (<i>Testing</i>)	Nilai Hasil (<i>Training</i>)	Keterangan
Koefisien Determinasi (R^2)	0,957	0,980	Menunjukkan 95,7% variabilitas QOS pada <i>testing</i> dan pada <i>training</i> menunjukkan 98%.
Mean Absolute Error (MAE)	0,145 +/- 0,208	0,108 +/- 0,130	Menunjukkan rata-rata kesalahan absolut QOS pada <i>testing</i> sekitar 0,14-0,20 unit, sedangkan pada <i>training</i> sekitar 0,10-0,13 unit berdasarkan skala 1-10
Root Mean Squared Error (RMSE)	0,253 +/- 0,000	0,169 +/- 0,000	Menunjukkan rata-rata kesalahan pada outlier menunjukkan performa model yang sangat baik apa bila nilai rendah

C. Analisis Koefisien Determinasi (R^2)

Berdasarkan data empiris pada Tabel 1, pengujian model pada fase *Training* menghasilkan nilai Koefisien Determinasi (R^2) sebesar 0,980. Angka ini mengindikasikan bahwa sebesar 98,0% variabilitas atau naik-turunnya nilai kualitas tidur (*QOS*) pada data latih mampu dijelaskan secara simultan dan linear oleh kombinasi 9 fitur prediktor yang digunakan. Ketika diuji menggunakan data independen pada fase *Testing*, model tetap menunjukkan performa yang sangat konsisten dengan raihan nilai R^2 sebesar 0,957. Hal ini membuktikan bahwa 95,7% variansi kualitas tidur pada data uji berhasil diprediksi dengan tepat oleh model, sementara 4,3% sisanya dipengaruhi oleh faktor-faktor luar yang tidak dicakup dalam penelitian ini. Selisih yang sangat tipis (hanya sekitar 2,3%) antara nilai R^2 *Training* dan *Testing* menegaskan bahwa model regresi yang dibangun terhindar dari kondisi *overfitting* maupun *underfitting*, serta memiliki kemampuan generalisasi yang luar biasa kokoh pada data baru.

D. Analisis Efisiensi Kesalahan Prediksi (MAE dan RMSE)

Evaluasi terhadap margin kesalahan model dilakukan secara komprehensif menggunakan metrik MAE dan RMSE untuk melihat akurasi prediksi secara riil pada skala target kualitas tidur (1 sampai 10).

1. *Mean Absolute Error (MAE)*: Nilai MAE pada fase *Training* tercatat sebesar $0,108 \pm 0,130$, sedangkan pada fase *Testing* mengalami sedikit kenaikan alami menjadi $0,145 \pm 0,208$. Secara praktis, nilai MAE *Testing* sebesar $0,145$ mengindikasikan bahwa rata-rata deviasi atau jarak kesalahan antara nilai kualitas tidur yang diprediksi oleh persamaan model terhadap nilai riil di lapangan hanya berkisar sekitar $0,14$ hingga $0,20$ unit saja. Mengingat target evaluasi berada pada rentang skala 1-10, margin kesalahan absolut di bawah $0,2$ unit ini mencerminkan tingkat presisi prediksi yang sangat tinggi.
2. *Root Mean Squared Error (RMSE)*: Metrik RMSE memberikan penalti yang lebih besar terhadap kesalahan prediksi berskala besar atau *outlier*. Nilai RMSE yang diperoleh adalah sebesar $0,169 \pm 0,000$ pada fase *Training* dan $0,253 \pm 0,000$ pada fase *Testing*. Nilai RMSE *Testing* yang berada pada angka rendah ($0,253$) mengonfirmasi bahwa model tidak hanya memiliki performa rata-rata yang baik, tetapi juga sangat stabil dan jarang menghasilkan *outlier error* (kesalahan prediksi ekstrem yang fatal) saat dihadapkan pada karakteristik data pengujian yang bervariasi. Nilai standar deviasi kesalahan yang mencapai $\pm 0,000$ pada kedua fase menunjukkan tingkat konvergensi dan konsistensi kesalahan yang homogen di seluruh sampel pengujian.

Secara keseluruhan, integrasi hasil tinggi pada metrik R^2 yang mendekati sempurna ($> 95\%$) serta rendahnya nilai indeks kesalahan (MAE dan RMSE < 0.3) membuktikan bahwa rumusan matematis Regresi Linear Berganda yang dieksekusi melalui alur kerja RapidMiner Studio ini sangat andal, valid, dan memiliki tingkat signifikansi yang tinggi dalam memprediksi kualitas tidur berdasarkan parameter gaya hidup dan biometrik harian.

Analisis Koefisien Regresi dan Interpretasi Klinis

Setelah keandalan model tervalidasi melalui metrik performa (error rate yang rendah dan R^2 yang tinggi), tahap krusial selanjutnya adalah membongkar struktur black-box dari algoritma untuk mengetahui bobot pengaruh masing-masing fitur. Ekstraksi parameter dari model Regresi Linear Berganda menghasilkan konstanta intersep β_0 dan koefisien bobot β_i untuk setiap atribut gaya hidup. Rangkuman nilai koefisien dari prediktor-prediktor utama dapat dilihat pada Tabel 2.

Tabel 2. Bobot Koefisien Regresi Linear Berganda terhadap Skor Kualitas Tidur

Variabel Prediktor (X_i)	Koefisien (β_i)	Keterangan Pengaruh terhadap QOS
Konstanta (Intersep / β_0)	8.033	Nilai dasar QOS jika semua variabel bernilai 0

Age	0.054	Positif	
Sleep Duration	0.269	Positif (dominan)	(Prediktor)
Physical Activity Level	0.0003	Positif	
Stress Level	-0.374	Negatif (signifikan)	(Prediktor)
Heart Rate	-0.019	Negatif	
Daily Steps	0.00002	Positif (Sangat kecil)	
Blood Pressure (Systolic)	-0.003	Negatif	
Blood Pressure (Diastolic)	-0.011	Negatif	
Gender: Male	0.457	Positif (Female)	(Dibandingkan)
BMI Category: Obese	-0.138	Negatif (Normal)	(Dibandingkan)
BMI Category: Overweight	-0.554	Negatif (Normal)	(Dibandingkan)

Berdasarkan Tabel 2, dapat diobservasi bahwa fitur-fitur gaya hidup memiliki arah dan magnitudo pengaruh yang bervariasi terhadap kualitas tidur (QOS). Berbeda dengan model klasifikasi yang hanya memberikan label diagnostik, nilai koefisien pada regresi ini memungkinkan pembentukan persamaan matematis yang terukur (mengacu pada Persamaan 1), yang secara ringkas diekspresikan sebagai berikut:

$$QOS = 8.033 + 0.269 (\text{Sleep Duration}) - 0.374 (\text{Stress Level}) + 0.054 (\text{Age}) - 0.019 (\text{Heart Rate}) - 0.011 (\text{Diastolic}) \quad (2)$$

Persamaan empiris di atas memberikan wawasan kuantitatif yang berharga. Secara spesifik, Sleep Duration (Durasi Tidur) teridentifikasi sebagai prediktor positif paling dominan dengan bobot $+0,269$. Hal ini mengindikasikan bahwa setiap penambahan satu jam durasi tidur akan meningkatkan kualitas tidur sebesar $0,269$ poin, dengan asumsi variabel lain konstan. Sebaliknya, Stress Level (Tingkat Stres) menjadi faktor penggerus paling signifikan dengan koefisien negatif sebesar $-0,374$, yang berarti setiap kenaikan satu tingkat stres akan menurunkan skor kualitas tidur secara drastis. Temuan terkuantifikasi ini menjawab keterbatasan model-model sebelumnya, karena kini klinisi memiliki landasan matematis untuk menakar intervensi—misalnya, memprioritaskan manajemen stres dan perbaikan durasi tidur—untuk mengembalikan kualitas tidur pasien ke angka optimal.

Lebih lanjut, temuan dari persamaan regresi ini tidak hanya kuat secara matematis, tetapi juga tervalidasi oleh literatur klinis terdahulu. Signifikansi variabel Stress Level sebagai prediktor negatif paling dominan sejalan dengan temuan observasional pada penelitian [2] dan [7], yang menegaskan bahwa tingginya stres secara langsung memicu degradasi kualitas tidur yang dapat berujung pada komplikasi kesehatan

kronis. Selanjutnya, meskipun bobot regresi untuk aktivitas fisik (*Physical Activity Level* dan *Daily Steps*) tergolong kecil secara desimal, arah koefisiennya yang bernilai positif secara konsisten mendukung rekomendasi klinis yang menyatakan bahwa olahraga dan kebugaran jasmani berkontribusi pada perbaikan kualitas tidur [4], [5]. Hal ini membuktikan bahwa persamaan MLR yang dihasilkan mampu menerjemahkan fenomena biologis dan gaya hidup ke dalam parameter komputasional yang valid.

IV. KESIMPULAN

Penelitian ini telah berhasil mengimplementasikan model Regresi Linear Berganda (*Multiple Linear Regression*) untuk memprediksi nilai numerik *Quality of Sleep* (QOS) pada skala 1–10 menggunakan *Sleep Health and Lifestyle Dataset*. Berdasarkan hasil eksperimen yang dieksekusi melalui alur kerja RapidMiner Studio, model prediktif yang dibangun menunjukkan performa yang sangat tangguh dan konsisten. Hal ini dibuktikan secara empiris oleh capaian nilai Koefisien Determinasi (R^2) sebesar 0,957 pada data pengujian, serta didukung oleh tingkat kesalahan prediktif yang sangat rendah, dengan nilai *Mean Absolute Error* (MAE) sebesar 0,145 dan *Root Mean Squared Error* (RMSE) sebesar 0,253.

Seluruh tujuan utama dari penelitian ini telah terpenuhi secara komprehensif. Melalui analisis koefisien regresi secara kuantitatif, penelitian ini berhasil mengidentifikasi bahwa variabel *Sleep Duration* (Durasi Tidur) bertindak sebagai faktor determinan positif yang paling signifikan, sedangkan variabel *Stress Level* (Tingkat Stres) terbukti menjadi kontributor negatif utama yang paling kuat dalam memengaruhi fluktuasi kualitas tidur seseorang.

DAFTAR PUSTAKA

- [1] N. Ulfah and A. Syahlani, "Hubungan gaya hidup dan kualitas tidur dengan kejadian hipertensi remaja," *Heal. Sci. J.*, vol. 16, no. 1, pp. 158–165, 2025, doi: 10.34305/jikbh.v16i01.1504.
- [2] M. Siregar, M. Deliana, Dedi, and P. Sari, "Hubungan Tingkat Stress dan Gaya Hidup Terhadap Kualitas Tidur pada Pasien Hipertensi," *J. Keperawatan Dirgahayu*, vol. 6, no. 2, pp. 29–36, Jan. 2025, doi: 10.52841/jkd.v6i2.464.
- [3] B. Saputra and M. Daniati, "Hubungan Konsumsi Alkohol dan Kualitas Tidur Terhadap Kejadian Hipertensi," *J. Keperawatan Hang Tuah (Hang Tuah Nurs. J.)*, vol. 2, no. 1, pp. 49–62, 2021.
- [4] D. Supriadi, M. B. Santoso, and N. P. Supriantini, "Penatalaksanaan kualitas tidur pada lansia dengan melakukan aktivitas olahraga jalan kaki," *Holistik J. Kesehat.*, vol. 17, no. 4, pp. 294–303, 2023.
- [5] S. D. Gunarsa and S. Wibowo, "Hubungan Kualitas Tidur dengan Kebugaran Jasmani Siswa," *J. Pendidik. Olahraga dan Kesehat.*, vol. 09, no. 01, pp. 43–52, 2021.
- [6] S. Sza *et al.*, "Penerapan Decision Tree dan Random Forest dalam Deteksi Tingkat Stres Manusia Berdasarkan Kondisi Tidur," *J. Teknol. Inf. dan Ilmu Komput. Vol.*, vol. 11, no. 5, pp. 1043–1050, 2024, doi: 10.25126/jtiik.2024117993.
- [7] I. W. G. Saraswasta, F. A. R. Tegu, A. A. A. E. Primayanthi, and R. T. Sucitra, "Hubungan Tingkat Stres dan Gaya Hidup Terhadap Kualitas Tidur pada Lansia di Desa Kota Batu," *Community Publ. Nurs.*, vol. 13, no. 3, pp. 307–314, 2025.
- [8] I. A. Hidayat, "Classification of Sleep Disorders Using Random Forest on Sleep Health and Lifestyle Dataset," *J. Dinda*, vol. 3, no. 2, pp. 71–76, 2023.
- [9] D. Fitriyani, M. Amelia, and S. S. Yuliana, "Penerapan Algoritma Random Forest Untuk Klasifikasi Gangguan Tidur Berdasarkan Pola Kehidupan Sehari-hari," *J. Apl. Inform. dan Multimed.*, vol. 1, no. 1, pp. 21–26, 2025.
- [10] N. Khasanah, D. Uki, E. Saputri, F. Aziz, and T. Hidayat, "Studi Perbandingan Algoritma Random Forest dan K-Nearest Neighbors (KNN) dalam Klasifikasi Gangguan Tidur," *Comput. Sci.*, vol. 5, no. 1, pp. 17–25, 2025.
- [11] D. Sari, "Prediksi Gangguan Tidur pada Sleep Health and Lifestyle Menggunakan Support Vector Machine dan Neural Network," *J. vokasi Inform.*, vol. 4, no. 1, pp. 36–42, 2024.
- [12] M. Maulidah and N. Hidayati, "Prediksi Kesehatan Tidur dan Gaya Hidup Menggunakan Machine Learning," *CONTEN Comput. Netw. Technol. Vol.*, vol. 4, no. 1, pp. 81–86, 2024.
- [13] D. A. Putri and M. Yasin, "Data Mining dengan Menggunakan Metode Time Series Analysis pada Pola Waktu Tidur," *Zetroem*, vol. 07, no. 02, pp. 65–72, 2025.
- [14] N. R. Haryana, R. Rosmiati, and E. M. Purba, "Generation Z Lifestyle in Aspect of Eating Behavior , Stress , Sleep Quality and Its Relation to Nutritional Status : Literature Review," *J. Gizi Kerja dan Produkt.*, vol. 4, no. 2, pp. 267–282, 2023.