

Prediksi Jumlah Perokok dan Dampaknya terhadap Kesehatan Masyarakat Menggunakan Regresi Linear

¹⁾Ernawati Carolina Sitohang, ²⁾Fitria Evaulina Ginting, ³⁾Yemia Margarettha Br Sembiring

Fakultas Ilmu Komputer, Universitas Katolik Santo Thomas Medan
Email : ernwatiCS@gmail.com

Abstrak

Penelitian ini bertujuan untuk menganalisis tren jumlah perokok dan dampaknya terhadap kesehatan masyarakat di Indonesia. Metode penelitian menggunakan data survei nasional terkait dari Badan Penelitian dan Pengembangan Kesehatan serta data terkait prevalensi penyakit terkait merokok dari Kementerian Kesehatan Indonesia. Prediksi dilakukan untuk periode 3 tahun terakhir guna mengevaluasi perubahan dalam jumlah perokok, dan dampak kesehatannya. Hasil penelitian menunjukkan bahwa meskipun terjadi penurunan jumlah perokok secara keseluruhan, terdapat peningkatan jumlah perokok. Selain itu, prevalensi penyakit terkait merokok seperti penyakit paru obstruktif kronis (PPOK) dan kanker paru juga mengalami peningkatan signifikan. Temuan ini menyoroti perlunya langkah-langkah pencegahan yang lebih efektif dan kebijakan publik yang lebih ketat untuk mengurangi prevalensi merokok dan dampak negatifnya terhadap kesehatan masyarakat.

Kata Kunci : Perokok; Kesehatan Masyarakat; Prevalensi; Kebijakan Publik

PENDAHULUAN

Merokok merupakan perilaku yang telah lama dikenal sebagai faktor risiko utama penyakit tidak menular dan menjadi masalah kesehatan global yang signifikan. Menurut World Health Organization (WHO), merokok menyebabkan lebih dari 8 juta kematian setiap tahunnya, dengan 7 juta di antaranya adalah akibat langsung dari penggunaan tembakau, sementara 1,2 juta adalah non-perokok yang terpapar asap rokok (WHO, 2020).

Di Indonesia, merokok juga merupakan masalah kesehatan masyarakat yang serius. Data dari Kementerian Kesehatan Indonesia menunjukkan bahwa prevalensi perokok di Indonesia masih tinggi, dengan lebih dari 65% pria dan 5% wanita merokok (Kementerian Kesehatan RI, 2018). Selain itu, prevalensi merokok pada remaja dan perempuan cenderung meningkat, menimbulkan keprihatinan akan dampak jangka panjangnya terhadap kesehatan masyarakat.

Dampak negatif merokok tidak hanya terbatas pada penyakit-penyakit seperti kanker paru, penyakit jantung, dan gangguan pernapasan, tetapi juga menyebabkan beban ekonomi yang besar akibat biaya perawatan kesehatan dan produktivitas yang hilang (WHO, 2019). Oleh karena itu, pengurangan prevalensi merokok dan promosi gaya hidup sehat menjadi fokus utama dalam upaya meningkatkan kesehatan masyarakat di Indonesia.

Penelitian sebelumnya telah memberikan wawasan yang berharga tentang tren jumlah perokok dan dampak kesehatannya di Indonesia, namun masih diperlukan analisis yang lebih mendalam untuk memahami dinamika perokok dan faktor-faktor yang mempengaruhinya secara lebih holistik. Oleh karena itu, penelitian ini bertujuan

untuk mengisi kesenjangan pengetahuan tersebut dengan menganalisis data survei terbaru dan mengidentifikasi implikasi kebijakan yang relevan dalam upaya mengurangi prevalensi merokok dan meningkatkan kesehatan masyarakat secara keseluruhan.

BAHAN DAN METODE

RapidMiner adalah platform analisis data yang kuat dan user-friendly yang menyediakan berbagai fitur untuk membantu pengguna dalam melakukan analisis data, pembelajaran mesin, dan penambangan data tanpa perlu menulis kode secara manual. Ini termasuk antarmuka grafis intuitif yang memungkinkan pengguna untuk membuat alur kerja analisis data dengan menarik dan menjatuhkan elemen-elemen analisis data ke dalam alur kerja tersebut. RapidMiner menyediakan berbagai algoritma pembelajaran mesin yang terintegrasi, serta alat untuk membersihkan dan mempersiapkan data sebelum dilakukan analisis lebih lanjut. Pengguna juga dapat membuat visualisasi data yang menarik dan informatif menggunakan RapidMiner. Selain itu, RapidMiner dapat diintegrasikan dengan berbagai platform dan bahasa pemrograman lainnya, seperti Python, R, dan SQL, memungkinkan pengguna untuk memperluas fungsionalitasnya sesuai kebutuhan mereka.

Linear Regression

Regresi linear adalah salah satu metode statistik yang digunakan untuk memodelkan hubungan linier antara satu atau lebih variabel independen (biasanya disebut sebagai "X" atau "variabel prediktor") dan variabel dependen (biasanya disebut sebagai "Y" atau "variabel respons"). Tujuan utama dari regresi linear adalah untuk memahami dan memprediksi nilai variabel dependen berdasarkan nilai-nilai variabel independen yang diberikan.

Pada dasarnya, regresi linear mencoba menemukan garis lurus terbaik (disebut garis regresi) yang sesuai dengan data observasi yang tersedia. Garis ini dapat digambarkan oleh persamaan matematika:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

- Y adalah variabel dependen (variabel yang ingin diprediksi).
- X adalah variabel independen (variabel yang digunakan untuk memprediksi Y).
- β_0 adalah intercept (nilai Y ketika $X = 0$).
- β_1 adalah koefisien regresi (menunjukkan perubahan dalam Y yang diakibatkan oleh perubahan satu unit dalam X).
- ϵ adalah kesalahan acak atau gangguan yang tidak dapat dijelaskan oleh model.

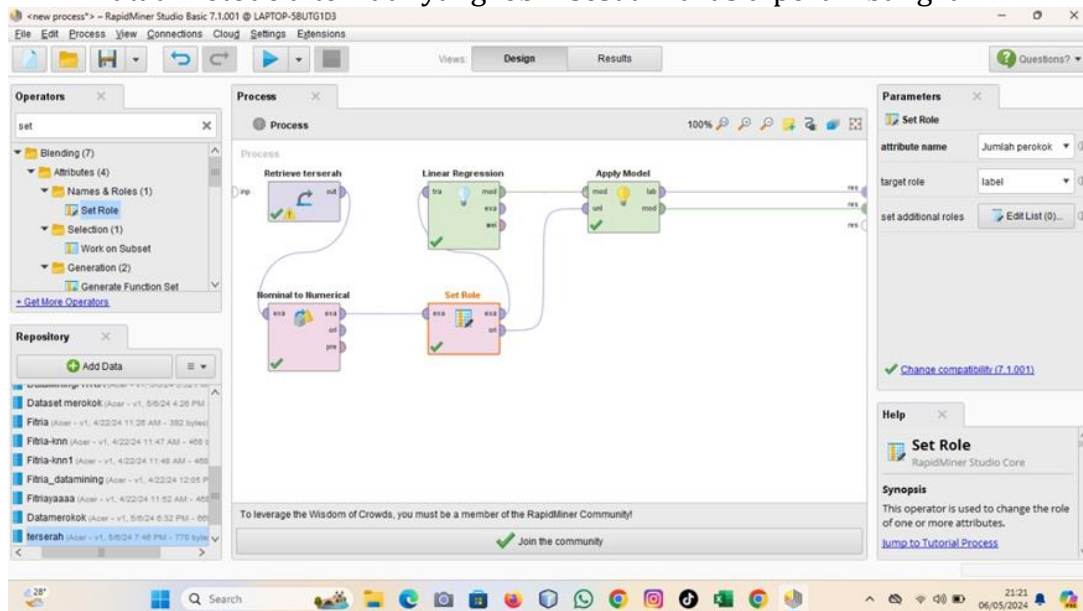
Proses fitting model regresi linear melibatkan penyesuaian parameter β_0 dan β_1 sehingga garis regresi meminimalkan jumlah kesalahan (residuals) antara nilai-nilai yang diprediksi oleh model dan nilai-nilai sebenarnya dalam data.

Regresi linear dapat digunakan untuk berbagai tujuan, termasuk:

1. Memprediksi nilai variabel dependen berdasarkan nilai-nilai variabel independen yang diberikan.
2. Memahami hubungan dan kekuatan hubungan antara variabel-variabel yang terlibat.

3. Menilai pengaruh variabel independen terhadap variabel dependen

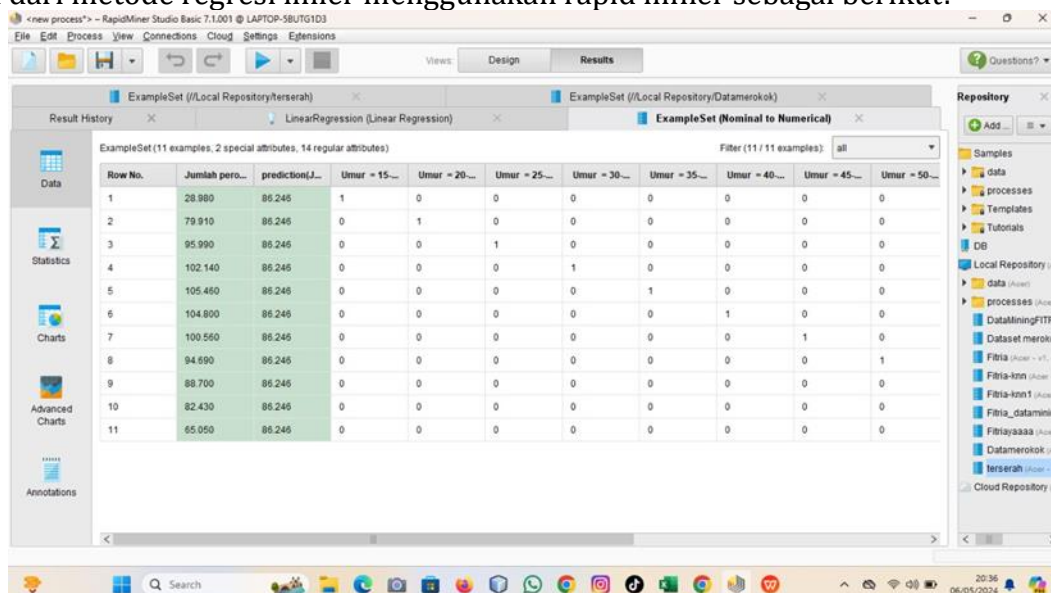
Namun, penting untuk dicatat bahwa regresi linear memiliki beberapa asumsi yang harus dipenuhi agar hasilnya valid, seperti asumsi tentang ketergantungan linear antara variabel, normalitas kesalahan, homoskedastisitas, dan independensi kesalahan. Jika asumsi-asumsi ini dilanggar, interpretasi hasil regresi harus dilakukan dengan hati-hati atau metode alternatif yang lebih sesuai harus dipertimbangkan.



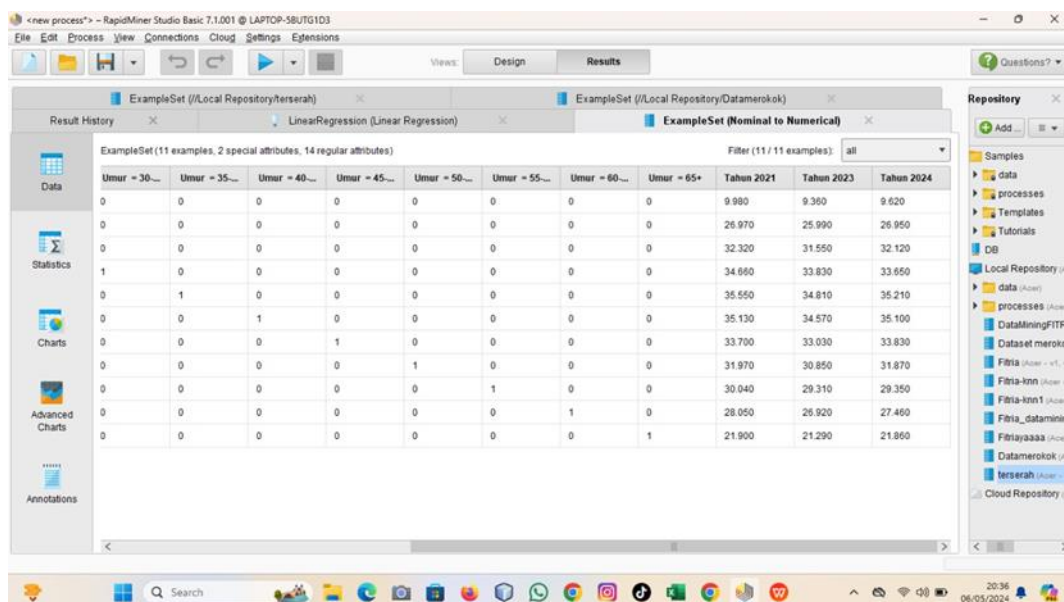
Gambar 1. Aplikasi RapidMiner

HASIL DAN PEMBAHASAN

Hasil dari metode regresi linier menggunakan rapid miner sebagai berikut:



Row No.	Jumlah perokok	prediction(J...	Umur = 15...	Umur = 20...	Umur = 25...	Umur = 30...	Umur = 35...	Umur = 40...	Umur = 45...	Umur = 50...
1	28.980	86.246	1	0	0	0	0	0	0	0
2	79.910	86.246	0	1	0	0	0	0	0	0
3	95.990	86.246	0	0	1	0	0	0	0	0
4	102.140	86.246	0	0	0	1	0	0	0	0
5	105.480	86.246	0	0	0	0	1	0	0	0
6	104.800	86.246	0	0	0	0	0	1	0	0
7	100.560	86.246	0	0	0	0	0	0	1	0
8	94.690	86.246	0	0	0	0	0	0	0	1
9	88.700	86.246	0	0	0	0	0	0	0	0
10	82.430	86.246	0	0	0	0	0	0	0	0
11	65.950	86.246	0	0	0	0	0	0	0	0



Umur = 30...	Umur = 35...	Umur = 40...	Umur = 45...	Umur = 50...	Umur = 55...	Umur = 60...	Umur = 65+	Tahun 2021	Tahun 2023	Tahun 2024
0	0	0	0	0	0	0	0	9.980	9.360	9.620
0	0	0	0	0	0	0	0	26.970	25.990	26.950
0	0	0	0	0	0	0	0	32.320	31.550	32.120
1	0	0	0	0	0	0	0	34.660	33.830	33.650
0	1	0	0	0	0	0	0	35.550	34.810	35.210
0	0	1	0	0	0	0	0	35.130	34.570	35.100
0	0	0	1	0	0	0	0	33.700	33.030	33.830
0	0	0	0	1	0	0	0	31.970	30.850	31.870
0	0	0	0	0	1	0	0	30.040	29.310	29.350
0	0	0	0	0	0	1	0	28.050	26.920	27.460
0	0	0	0	0	0	0	1	21.900	21.290	21.860

Gambar 2. Hasil Pengolahan data dengan RapidMiner

KESIMPULAN

Dalam menggunakan regresi linear untuk memprediksi jumlah perokok, kita dapat memodelkan hubungan antara variabel-variabel yang berpotensi mempengaruhi jumlah perokok dengan variabel dependen yang merupakan jumlah perokok itu sendiri. Variabel independen dapat mencakup faktor-faktor seperti demografi (usia, jenis kelamin, pendidikan), faktor ekonomi (pendapatan, pekerjaan), dan faktor sosial (lingkungan, promosi kesehatan).

Dengan menggunakan regresi linear, kita dapat mengestimasi koefisien untuk setiap variabel independen yang menunjukkan seberapa besar pengaruhnya terhadap jumlah perokok. Hasil regresi dapat digunakan untuk membuat prediksi tentang jumlah perokok di masa depan berdasarkan nilai-nilai variabel independen yang diberikan.

Namun, penting untuk diingat bahwa regresi linear memiliki beberapa asumsi yang harus dipenuhi untuk menghasilkan hasil yang valid. Oleh karena itu, sebelum melakukan analisis regresi, perlu dilakukan pemeriksaan terhadap asumsi-asumsi tersebut dan tindakan koreksi yang diperlukan jika asumsi tersebut dilanggar.

Dengan demikian, kesimpulan menggunakan regresi linear untuk memprediksi jumlah perokok adalah bahwa dengan memahami dan menganalisis faktor-faktor yang mempengaruhi perilaku merokok, kita dapat menggunakan model regresi untuk membuat prediksi tentang jumlah perokok di masa depan dan merumuskan strategi pencegahan yang lebih efektif untuk mengurangi prevalensi merokok dan meningkatkan kesehatan masyarakat secara keseluruhan.

UCAPAN TERIMA KASIH

Kami ingin mengucapkan terima kasih atas kesempatan untuk belajar lebih dalam tentang masalah jumlah perokok menggunakan regresi linear. Metode ini memberikan wawasan yang berharga tentang faktor-faktor yang mempengaruhi perilaku merokok dan membantu kami memprediksi tren masa depan. Dengan pemahaman yang lebih baik tentang hubungan antara variabel-variabel yang terlibat, kami dapat merumuskan



strategi yang lebih efektif dalam menangani masalah ini dan meningkatkan kesehatan masyarakat secara keseluruhan. Terima kasih atas bimbingan dan dukungan dalam penelitian ini.

DAFTAR PUSTAKA

1. Adhayanti, R. 2007. Hubungan Tingkat Pengetahuan Bahaya Rokok Bagi Kesehatan Terhadap Perilaku Merokok
2. Abazinab, S., & Dereje, N. (2015). Prevalence and Predictors of Cigarette Smoking among Adolescents of Ethiopia: School Based Cross Sectional Survey. *Journal of Child and Adolescent Behaviour*, 03(01). <https://doi.org/10.4172/2375-4494.1000182>
3. Atari, D. O. (2014). Gender differences in the prevalence and determinants of tobacco use among school-aged adolescents (11 – 17 years) in Sudan and South Sudan. *Pan African Medical Journal*, 18, 118. <https://doi.org/10.11604/pamj.2014.18.118.3202>