
KOMPARASI TEKNIK *UNDERSAMPLING* DAN *OVERSAMPLING* PADA REGRESI LOGISTIK BINER DALAM MENDUGA FAKTOR DETERMINAN BERHENTI MEROKOK PENDUDUK LANJUT USIA

Reni Amelia¹⁾, Indahwati²⁾, Erfiani³⁾, Anwar Fitrianto⁴⁾, Akbar Rizki⁵⁾
Prodi Statistika dan Sains Data IPB University dan Badan Pusat Statistik^{1,2,3,4,5)}
IPB University^{1,2,3,4,5)}

Kampus IPB Baranangsiang, Jl. Raya Pajajaran, Bogor, Jawa Barat-Indonesia^{1,2,3,4,5)}

email: reniamelia@apps.ipb.ac.id¹⁾, indah.stk@gmail.com²⁾, erfiani@apps.ipb.ac.id³⁾, anwarstat@gmail.com⁴⁾, akbar.ritzki@apps.ipb.ac.id⁵⁾

Abstrak

Teknik *resampling* adalah salah satu teknik *pre-processing* untuk menyeimbangkan distribusi data sehingga mengurangi efek distribusi kelas atau kategori yang tidak seimbang. Teknik *resampling* yang biasa digunakan adalah *random oversampling* dan *random undersampling*. Dalam penelitian ini, *random oversampling* digunakan untuk menyeimbangkan data dengan cara *oversampling* secara acak pada kelas minoritas (penduduk lansia yang berhenti merokok). *Random undersampling* digunakan untuk menyeimbangkan data dengan cara *undersampling* (mengeliminasi) secara acak kelas mayoritas (penduduk lansia yang masih merokok). Data yang telah diproses dengan *resampling* selanjutnya dilakukan pemodelan dengan model regresi logistik biner. Model regresi logistik biner dengan *random undersampling* merupakan model terbaik karena memiliki *balanced accuracy* terbesar. Peubah yang signifikan memengaruhi berhenti merokok adalah pendidikan, pekerjaan, akses internet, dan usia lansia.

Kata Kunci: *Resampling, Undersampling, Oversampling, Regresi Logistik Biner, Berhenti Merokok*

1. Pendahuluan

Perilaku merokok dilihat dari berbagai sisi sangat merugikan bagi pelakunya. *World Health Organization (WHO)* menyatakan bahwa merokok berakibat fatal bagi kesehatan bahkan dapat menyebabkan kematian. [1] Efek buruk dari merokok ini terjadi pada semua kalangan usia, termasuk penduduk lanjut usia.[2] Berdasarkan data Survei Sosial Ekonomi Nasional (SUSENAS) yang dilakukan Badan Pusat Statistik (BPS) tahun 2020, sekitar 23.55 persen lansia merokok. [3] Semakin lama seseorang merokok, ketergantungan terhadap rokok akan semakin meningkat. [4] Burns menyatakan bahwa berhenti merokok merupakan cara paling efektif untuk mengubah risiko penyakit pada lansia.[5]

Berbagai kebijakan disusun untuk mendukung proses berhenti merokok baik di dunia maupun di Indonesia. WHO telah menyusun *World Health Organization Framework Convention on Tobacco Control (WHO FCTC)* untuk melindungi generasi sekarang dan yang akan datang terhadap efek buruk dari merokok. [2] Dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024 Indonesia, layanan berhenti merokok merupakan salah satu strategi kebijakan pengendalian tembakau. Ketersediaan data kesehatan yang beragam dan spesifik akan meningkatkan perencanaan kebijakan kesehatan (Kemenkes 2017).[6] Perlu dikaji faktor determinan berhenti merokok pada lansia.

Dalam mengkaji determinan berhenti merokok pada lansia, digunakan model regresi logistik biner. Regresi logistik biner digunakan untuk memodelkan hubungan antara peubah respon yang terdiri dari dua kategori dengan satu atau lebih peubah penjelas (Hosmer dan Lemeshow 2000). [7] Peubah bebas yang digunakan pada pemodelan regresi logistik biner adalah peubah yang signifikan memiliki hubungan dengan status merokok lansia pada uji asosiasi (uji *Chi-square*).

Sebelum proses pemodelan, dilakukan juga *pre-processing* data dengan menggunakan teknik *resampling*, yaitu salah satu teknik *pre-processing* untuk menyeimbangkan distribusi data sehingga mengurangi efek distribusi kelas atau kategori yang tidak seimbang (Jian et al. 2016). [8] Ketidakseimbangan data memberikan dampak yang buruk pada hasil klasifikasi dimana kelas minoritas sering disalahklasifikasikan sebagai kelas mayoritas.

Proses *resampling* dilakukan karena data lansia yang merokok jauh lebih kecil dibandingkan data lansia yang masih merokok. Data sampel penduduk lanjut usia yang diperoleh terdiri dari sekitar 11 persen lansia Indonesia berhenti merokok tembakau sedangkan sisanya masih merokok. Teknik *resampling* yang digunakan adalah *random oversampling* dan *random undersampling*. *Random oversampling* digunakan untuk menyeimbangkan data dengan cara *oversampling* secara acak pada kelas minoritas (penduduk lansia yang berhenti merokok). *Random undersampling* digunakan untuk menyeimbangkan data dengan cara *undersampling* (mengeliminasi)

secara acak kelas mayoritas (penduduk lansia yang masih merokok)

Tujuan dari penelitian ini: (1) mengetahui karakteristik penduduk lanjut usia (lansia) yang berhenti merokok; (2) mengaplikasikan metode regresi logistik biner untuk mengidentifikasi faktor apa saja yang menjadi determinan berhenti merokok lansia; (3) menentukan metode regresi logistik biner dengan proses *resampling* (*undersampling* dan *oversampling*) atau metode regresi logistik biner tanpa proses *resampling* yang merupakan model terbaik dalam mengklasifikasikan lansia yang berhenti merokok.

2. Landasan Teori

Teknik *Resampling*

Ketidakeimbangan data terjadi pada saat suatu kelas atau kategori tertentu memiliki data yang lebih banyak dibandingkan dengan kategori lainnya. [8] Ketidakeimbangan data ini perlu ditangani karena memengaruhi akurasi pada proses klasifikasi data. [9] Teknik *resampling* merupakan salah satu solusi untuk menangani data yang tidak seimbang. Terdapat proses *random undersampling* dan *oversampling*. *Random undersampling* merupakan proses *sampling* yang dilakukan dengan mengeliminasi sebagian data pada kelas mayoritas secara acak. [10] *Random oversampling* merupakan proses *sampling* yang dilakukan dengan menambahkan jumlah data pada kelas minoritas secara acak.

Uji Asosiasi

Sebelum dilakukan model regresi logistik biner, dilakukan uji asosiasi antara peubah bebas dengan peubah tidak bebas. Pengujian dilakukan dengan menggunakan hipotesis berikut:

H_0 : Tidak ada asosiasi antara peubah bebas dan peubah tidak bebas

H_1 : Ada asosiasi antara peubah bebas dan peubah tidak bebas

Statistik uji *Chi-Square* yang digunakan adalah pada persamaan (1).

$$\chi_{obs}^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \dots\dots\dots(1)$$

Tolak H_0 jika $\chi_{obs}^2 > \chi_{(r-1)(k-1)}^2$

Keterangan:

O_{ij} = frekuensi observasi pada kategori baris ke – i kolom ke – j

E_{ij} = frekuensi harapan pada kategori baris ke – i kolom ke – j

r = jumlah kategori pada baris

k = jumlah kategori pada kolom

Regresi Logistik Biner

Regresi logistik biner digunakan untuk memodelkan hubungan antara peubah respon yang terdiri dari dua kategori dengan satu atau lebih peubah penjelas (Hosmer dan Lemeshow 2000).[7] Model regresi logistik biner selengkapannya ada pada persamaan (2).

$$g(x) = \ln \left[\frac{\pi(x)}{1-\pi(x)} \right] = \beta_0 + \beta^T x \dots\dots\dots(2)$$

Keterangan:

$$\pi(x) = \frac{\exp(\beta_0 + \beta^T x)}{1 + \exp(\beta_0 + \beta^T x)}$$

$$g(x) = \text{logit } \pi(x)$$

β_0 = parameter intercept

β^T = vektor yang berisi nilai parameter atau koefisien regresi

x = vektor yang berisi nilai variabel bebas

Pendugaan parameter pada model regresi logistik biner menggunakan metode kemungkinan maksimum dan diselesaikan dengan iterasi *Newton Raphson*. [11] Dalam pengujian parameter model dilakukan dengan *likelihood ratio test* atau uji G (uji simultan) dan statistik uji Wald (uji parsial).

Pemilihan Model Terbaik

Balanced accuracy merupakan salah satu tolok ukur menilai kebaikan model pada data yang tidak seimbang. Rumus menghitung *balanced accuracy* melibatkan nilai pada Tabel 1.

Tabel 1. Ketepatan Klasifikasi

Prediksi	Kenyataan	
	Masih Merokok	Sudah Tidak Merokok
Masih Merokok	a	b
Sudah Tidak Merokok	d	e

Rumus *balanced accuracy*:

$$\text{Balanced Accuracy} = \frac{\text{sensitivity} + \text{specificity}}{2} \dots\dots\dots(3)$$

Keterangan:

$$\text{Sensitivity} = \frac{a}{a + d}$$

$$\text{Specificity} = \frac{e}{b + e}$$

Berhenti Merokok Lansia

Merokok sangat berakibat fatal bagi kesehatan bahkan dapat menyebabkan kematian. Burns menyatakan bahwa berhenti merokok merupakan cara paling efektif untuk mengubah risiko penyakit pada lansia.[5] Terdapat beberapa faktor yang memengaruhi seseorang berhenti merokok, termasuk lansia. Tabel 2 menjelaskan beberapa literatur yang menjelaskan hubungan beberapa peubah dengan status berhenti merokok lansia.

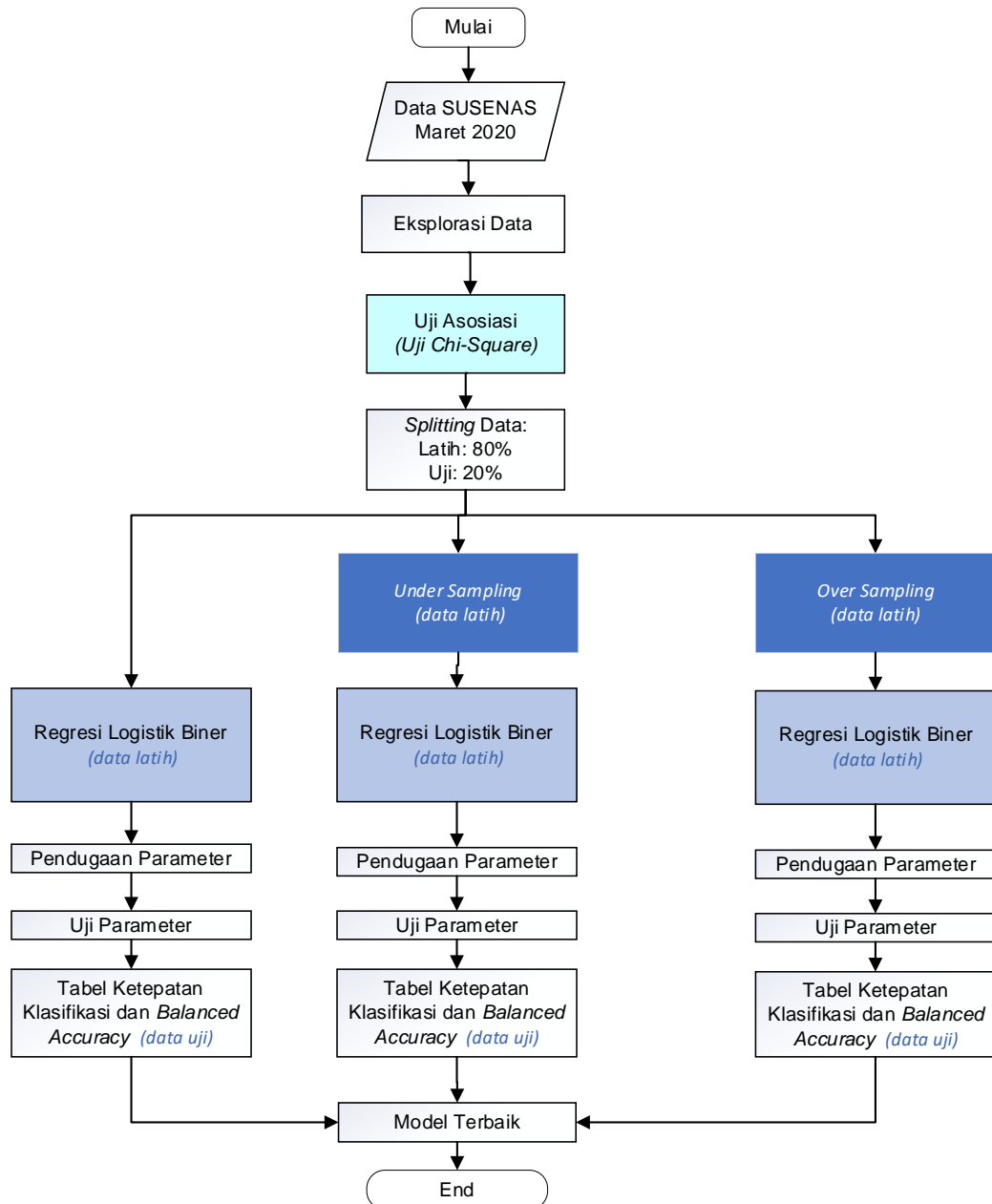
Tabel 2. Peubah Penelitian dan Referensi

No	Peubah Penelitian	Referensi	Kesimpulan
1	Status Berhenti Merokok(Y)		
2	Tingkat pendidikan lansia	Sirait <i>et al.</i> (2002) Rahim <i>et al.</i> (2016) Juliansyah <i>et al.</i> (2018) Sihombing <i>et al.</i> (2020)	Semakin tinggi pendidikan akan meningkatkan kecenderungan untuk berhenti (tidak) merokok. [12,13,14,15]
3	Jenis kelamin	Sirait <i>et all.</i> (2002) Rahim <i>et al.</i> (2016) Sukmono <i>et al.</i> (2019) Sadarang RAI <i>et al.</i> (2021)	Perempuan memiliki kecenderungan yang lebih tinggi untuk berhenti (tidak) merokok dibandingkan laki-laki [12,13,16,17]
4	Status perkawinan	Ramsey <i>et al.</i> (2018) Reskiaddin <i>et al.</i> (2021)	Seseorang yang belum kawin memiliki kecenderungan yang lebih tinggi untuk berhenti (tidak) merokok dibandingkan yang kawin [18,19]
5	Status bekerja	Rahim <i>et al.</i> (2016) Reskiaddin <i>et al.</i> (2021)	Seseorang yang tidak bekerja memiliki kecenderungan yang lebih tinggi untuk berhenti (tidak) merokok dibandingkan yang bekerja (memiliki penghasilan sendiri) [13,19]
6	Tipe daerah tempat tinggal	Sirait <i>et al.</i> (2002) Rahim <i>et al.</i> (2016)	Seseorang yang tinggal di perkotaan memiliki kecenderungan yang lebih tinggi untuk berhenti (tidak) merokok dibandingkan yang tinggal di perdesaan [12,13]
7	Status akses internet	Lestari <i>et al.</i> (2013) Brown <i>et al.</i> (2014) Taylor <i>et al.</i> (2017)	Seseorang yang mengakses internet (media promosi berhenti merokok) memiliki kecenderungan yang lebih tinggi untuk berhenti (tidak) merokok dibandingkan yang tidak mengakses internet (media promosi berhenti merokok) [20, 21, 22]
8	Umur	1. Hiscock <i>et al.</i> (2015) 2. Qiu <i>et al.</i> (2020)	Seseorang yang memiliki usia yang lebih tua memiliki kecenderungan yang lebih

No	Peubah Penelitian	Referensi	Kesimpulan
			tinggi untuk berhenti (tidak) merokok dibandingkan yang lebih muda. [23, 24]

3. Metode Penelitian

Gambar 1 memperlihatkan *flowchart* tahapan dalam penelitian ini.



Gambar 1. *Flowchart* Tahapan Penelitian

Secara umum, tahapan penilitan ini adalah sebagai berikut:

1. Melakukan eksplorasi terhadap data awal

Penelitian ini menggunakan data sekunder hasil Survei Sosial Ekonomi Nasional (SUSENAS) yang dilakukan Badan Pusat Statistik (BPS) pada Maret 2020. Populasi yang menjadi fokus dalam penelitian ini adalah penduduk lanjut usia (60 tahun ke atas) yang pernah merokok tembakau di Indonesia. Total sampel

(SUSENAS) yang digunakan sebanyak 30.342 lansia (3.486 lansia berhenti merokok dan 26.856 lansia masih merokok).

2. Melakukan uji asosiasi

Sebelum dilakukan pemodelan, dilakukan uji asosiasi dengan uji *Chi-Square* untuk menguji apakah ada asosiasi antara setiap peubah bebas dengan peubah status berhenti merokok (peubah tidak bebas).

3. Membagi data menjadi 80% data latih dan 20% data uji

Data sampel 30.342 lansia yang pernah merokok tembakau dibagi menjadi dua, yaitu 80 persen data *training* (data latih) sekitar 24.273 lansia dan 20 persen data *testing* (data uji) sekitar 6.069 lansia. Data *training* digunakan untuk membentuk model sedangkan data *testing* digunakan untuk menguji model yang telah dibentuk. Dari 24.273 lansia yang menjadi data *training*, 2.794 lansia yang dikategorikan sebagai berhenti merokok sedangkan 21.479 lansia yang dikategorikan sebagai masih merokok.

4. Melakukan pemodelan regresi logistik biner (diawali dengan *undersampling*, *oversampling*, dan tanpa *resampling*)

Model regresi logistik biner digunakan untuk mengidentifikasi faktor apa saja yang menjadi determinan berhenti merokok lansia. Dalam penelitian ini juga digunakan *pre-processing* data dengan menggunakan teknik *resampling*, *random oversampling* dan *random undersampling*. Teknik *resampling* diterapkan pada data *training* (data latih) untuk pembentukan model regresi logistik biner. Proses *undersampling* dilakukan dengan mengeliminasi kategori masih merokok menjadi 2.794 (sama dengan jumlah lansia yang berhenti merokok). Total dataset baru hasil *undersampling* sebanyak 5588. Pada *oversampling*, kategori minoritas dinaikkan secara acak sehingga jumlahnya sama dengan kategori mayoritas menjadi 21.479. Total dataset hasil *oversampling* sebanyak 42.958.

Pada pembentukan persamaan regresi logistik biner, dilakukan pendugaan parameter menggunakan metode kemungkinan maksimum dan diselesaikan dengan *iterasi Newton Raphson* menggunakan *software R*. Setelah itu dilakukan pengujian apakah parameter tersebut signifikan. Model yang telah terbentuk selanjutnya dilakukan evaluasi. Evaluasi dilakukan dengan menyusun tabel ketepatan klasifikasi dan menghitung *balanced accuracy*.

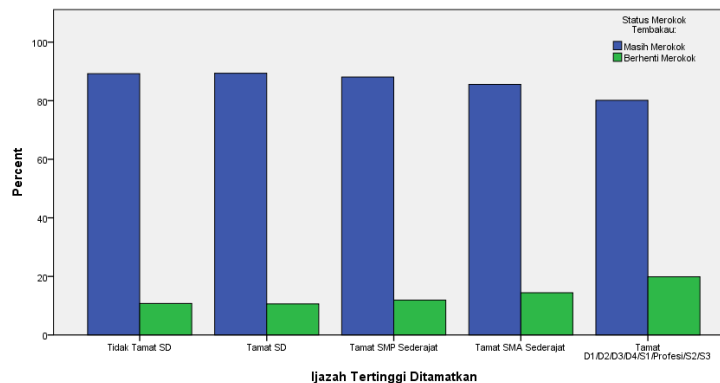
5. Menentukan model terbaik

Model yang terbaik adalah model yang memiliki *balanced accuracy* tertinggi. Selanjutnya model ini yang akan diinterpretasikan dan digunakan untuk mengklasifikasikan status berhenti merokok pada lansia.

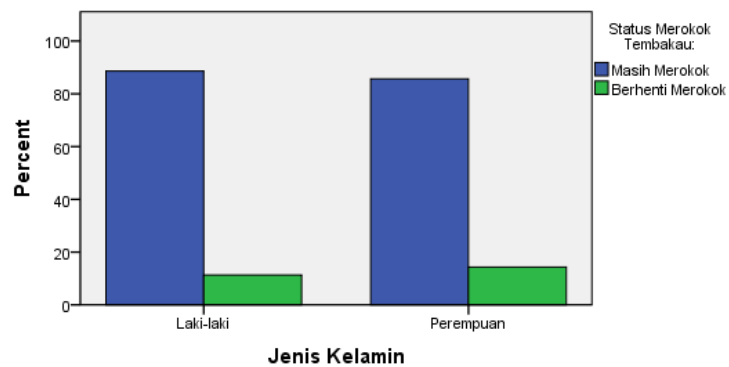
4. Hasil Penelitian

1. Hasil Eksplorasi terhadap Data Awal

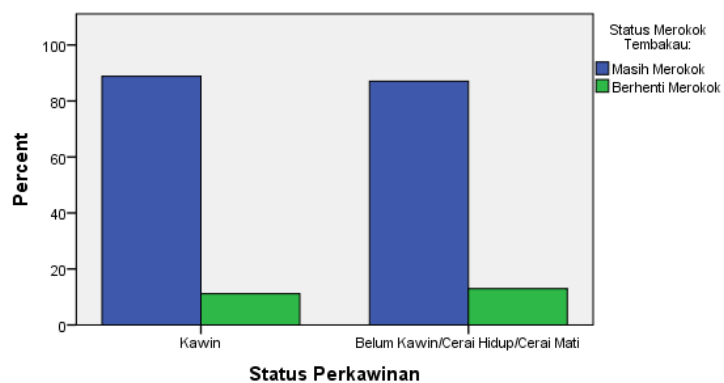
Dari data sampel 30.342 lansia yang pernah merokok tembakau hasil SUSENAS Maret 2020, proporsi penduduk lanjut usia yang berhenti merokok dibandingkan lansia yang masih merokok berdasarkan ijazah tertinggi yang ditamatkan semakin meningkat seiring makin tinggi jenjang pendidikannya (Gambar 2). Proporsi berhenti merokok perempuan lebih besar dibandingkan laki-laki (Gambar 3). Proporsi berhenti merokok lansia yang belum kawin/cerai hidup/cerai mati lebih besar dibandingkan lansia yang kawin (Gambar 4). Proporsi berhenti merokok lansia yang tidak bekerja lebih besar dibandingkan yang bekerja (Gambar 5). Proporsi lansia berhenti merokok di perkotaan lebih besar dibandingkan perdesaan (Gambar 6). Proporsi lansia berhenti merokok untuk lansia yang mengakses internet lebih besar dibandingkan yang tidak mengakses internet (Gambar 7). Median usia lansia yang berhenti merokok lebih besar dibandingkan median usia lansia yang masih merokok (Gambar 8).



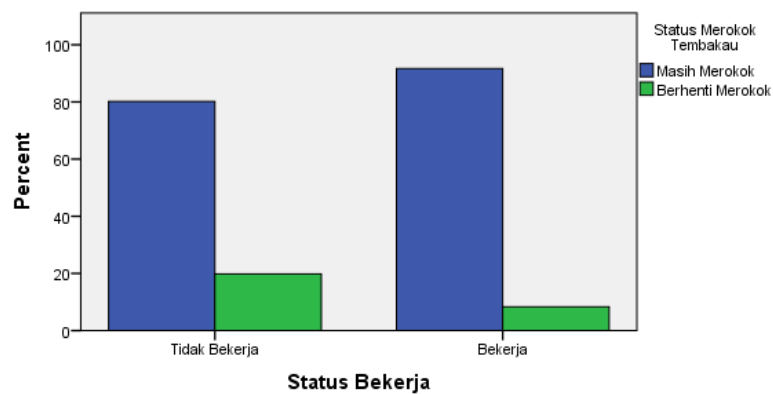
Gambar 2. Persentase Lansia Menurut Ijazah Tertinggi yang Ditamatkan dan Status Merokok Tembakau di Indonesia Tahun 2020



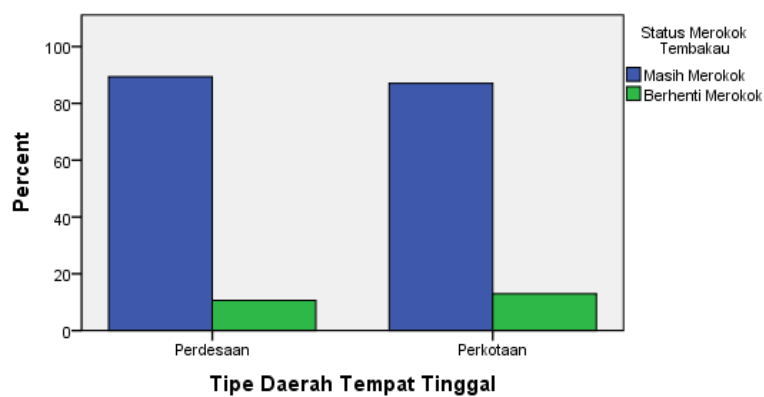
Gambar 3. Persentase Lansia Menurut Jenis Kelamin dan Status Merokok Tembakau di Indonesia Tahun 2020



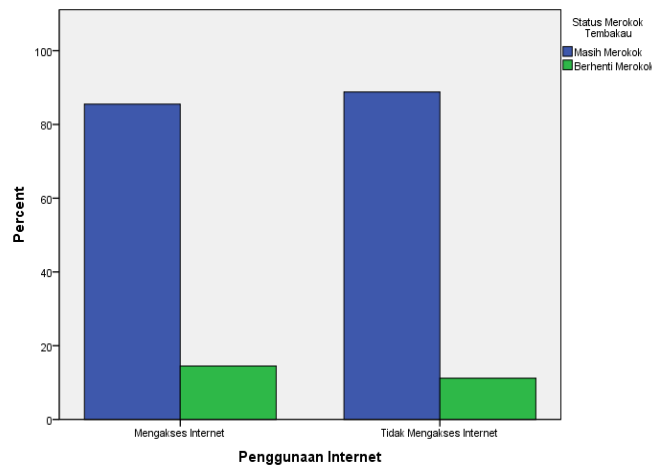
Gambar 4. Persentase Lansia Menurut Status Perkawinan dan Status Merokok Tembakau di Indonesia Tahun 2020



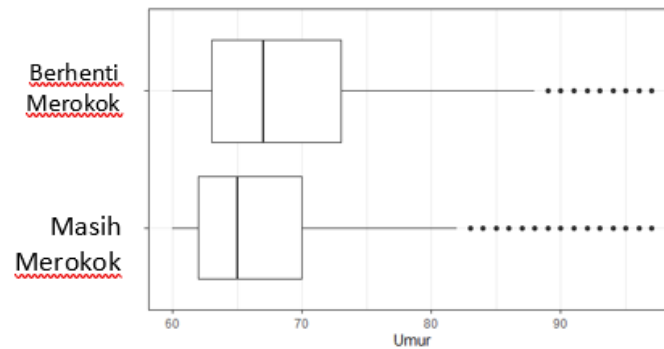
Gambar 5. Persentase Lansia Menurut Status Bekerja dan Status Merokok Tembakau di Indonesia Tahun 2020



Gambar 6. Persentase Lansia Menurut Status Tipe Daerah Tempat Tinggal dan Status Merokok Tembakau di Indonesia Tahun 2020



Gambar 7. Persentase Lansia Penggunaan Internet dan Status Merokok Tembakau di Indonesia Tahun 2020



Gambar 8. Usia Lansia Menurut Status Merokok Tembakau di Indonesia Tahun 2020

2. Hasil Uji Asosiasi

Hasil uji asosiasi dengan menggunakan uji *Chi-Square* menunjukkan bahwa jenis kelamin, ijazah tertinggi yang ditamatkan, status perkawinan, status bekerja, tipe daerah tempat tinggal, status akses internet, dan umur memiliki asosiasi dengan status berhenti merokok pada lansia (Tabel 4).

Tabel 4. *Chi-square Value*, Derajat Bebas, dan *P-Value* Hasil Uji Asosiasi (Uji *Chi-Square*) antara Peubah Bebas dengan Peubah Tidak Bebas (Status Berhenti Merokok)

Peubah Bebas	<i>Chi-Square Value</i>	Derajat Bebas	<i>P-Value</i>
Jenis Kelamin	113.44	4	< 2.2e-16
Ijazah Tertinggi yang Ditamatkan	11.77	1	0.0006
Status Perkawinan	14.11	1	0.0002
Status Bekerja	792.75	1	< 2e-16
Tipe Daerah Tempat Tinggal	36.80	1	1.309e-09
Status Akses Internet	25.52	1	4.378e-07
Umur	360.63	37	2.2e-16

Sumber: Output Software R Hasil Olah Data Sampel Susenas Maret 2020

3. Hasil Membagi Data Menjadi 80% Data Latih dan 20% Data Uji

Terdapat 24.273 lansia yang menjadi data latih (*training*) dan 6.069 lansia data *testing* (data uji). Dari 24.273 lansia yang menjadi data *training*, 2.794 lansia yang dikategorikan sebagai berhenti merokok sedangkan 21.479 lansia yang dikategorikan sebagai masih merokok.

4. Melakukan pemodelan regresi logistik biner (diawali dengan *undersampling*, *oversampling*, dan tanpa *resampling*)

Tabel 5 menyajikan hasil pemodelan regresi logistik biner dengan teknik *resampling* dan tanpa *resampling*.

Tabel 5. Hasil Model Regresi Logistik Biner

Variabel	Model Regresi Logistik Biner					
	Tanpa <i>Resampling</i>		Dengan <i>Random Oversampling</i>		Dengan <i>Random Undersampling</i>	
	Koefisien Penduga Parameter	<i>p_value</i>	Koefisien Penduga Parameter	<i>p_value</i>	Koefisien Penduga Parameter	<i>p_value</i>
Intercept	-3.87	<2e-16	-1.721	<2e-16	-2.166	1.2e-10
Ijazah Tertinggi Yang Ditamatkan						
Tamat SD sederajat vs Tidak Tamat SD	0.071	0.148	0.029	0.213	0.086	0.183
Tamat SMP sederajat vs Tidak Tamat SD	0.200	0.008	0.141	0.000	0.276	0.008
Tamat SMA sederajat vs Tidak Tamat SD	0.250	0.001	0.223	2.96e-09	0.314	0.003
Tamat Perguruan Tinggi vs Tidak Tamat SD	0.513	5.8e-07	0.457	< 2e-16	0.528	0.001
Jenis Kelamin (Laki-laki vs Perempuan)	0.004	0.963	0.007	0.885	-0.079	0.537
Status Perkawinan (Kawin vs Tidak Kawin/Cerai Hidup/Cerai Mati)	0.096	0.077	0.082	0.002	0.042	0.573
Status Bekerja (Bekerja vs Tidak Bekerja)	-0.849	<2e-16	-0.853	<2e-16	-0.849	<2e-16
Tipe Daerah Tempat Tinggal (Perdesaan vs Perkotaan)	-0.031	0.486	-0.0445	0.039	-0.066	0.280
Status Akses Internet (Meangakses Internet vs Tidak Mengakses Internet)	0.096	0.207	0.123	0.001	0.271	0.013
Umur	0.033	<2e-16	0.032	<2e-16	0.038	<2e-16

Sumber: Output *Software R* Hasil Olah Data Sampel Susenas Maret 2020

Tabel 5 menunjukkan bahwa peubah ijazah tertinggi yang ditamatkan (tamat SMP sederajat vs tidak tamat SD, tamat SMA sederajat vs tidak tamat SD, dan tamat Perguruan Tinggi vs tidak tamat SD), status bekerja, dan umur lansia signifikan ($p_value < 0.05$) memengaruhi status berhenti merokok pada lansia pada taraf alpha 5% pada semua jenis model (model regresi logistik biner dengan dan tanpa *resampling*). Peubah status mengakses internet signifikan memengaruhi status berhenti merokok pada lansia pada taraf alpha 5% pada model regresi logistik biner dengan melakukan *resampling* (*oversampling* dan *undersampling*) sedangkan peubah status perkawinan lansia signifikan memengaruhi status berhenti merokok pada lansia pada taraf alpha 5% pada model regresi logistik biner dengan melakukan *oversampling*.

5. Model Terbaik

Model yang terbaik adalah model yang memiliki *balanced accuracy* tertinggi. Dari ketiga model regresi logistik biner, model yang didahului dengan proses *random undersampling* memberikan nilai *balanced accuracy* terbesar (62.44 persen) dibandingkan dengan model tanpa *resampling* (50.00 persen) dan dengan *oversampling* (62.20 persen). Selanjutnya, model regresi logistik biner dengan proses *random sampling* dipilih menjadi model terbaik dan diinterpretasikan nilai *odds*-nya (Tabel 6). Persamaan regresi logistik biner yang terbentuk adalah:

$$\text{StatusBerhentiMerokok} = -2.166 + 0.276 * \text{IjazahTertinggiSD} + 0.314 * \text{IjazahTertinggiSMP} + 0.528 * \text{IjazahTertinggiSMA} - 0.849 * \text{StatusBekerja} + 0.271 * \text{StatusAksesInternet} + 0.038 * \text{Umur}$$

Tabel 6. Nilai Estimasi Koefisien Penduga Parameter dan Eksponensialnya (*Odds Ratio*) Pada Model Regresi Logistik Biner dengan *Undersampling*

Karakteristik	Estimasi Koefisien	Exp(Estimasi Koefisien)= <i>Odds Ratio</i>
<i>Intercept</i>	-2.166	0.115
Ijazah Tertinggi Yang Ditamatkan		
Tamat SMP sederajat vs Tidak Tamat SD	0.276	1.318
Tamat SMA sederajat vs Tidak Tamat SD	0.314	1.369
Tamat Perguruan Tinggi vs Tidak Tamat SD	0.528	1.696
Status Bekerja (Bekerja vs Tidak Bekerja)	-0.849	0.428
Status Akses Internet (Meangakses Internet vs Tidak Mengakses Internet)	0.271	1.311
Umur	0.038	1.039

Sumber: Output *Software R* Hasil Olah Data Sampel Susenas Maret 2020

Interpretasi dari *odd ratio*, yaitu:

- $Exp(\text{Intecept}) = 0.115 \rightarrow$ tanpa memperhatikan peubah bebasnya, peluang lansia berhenti merokok adalah 0.115 kali dibandingkan peluang untuk masih merokok.
- $Exp(\text{Tamat SMP sederajat vs Tidak Tamat SD}) = 1.318 \rightarrow$ *Odds* lansia berhenti merokok jika lansia memiliki ijazah terakhir tamat SMP Sederajat adalah 1.318 kali dibandingkan *odd* yang sama jika lansia tersebut memiliki ijazah tidak tamat SD dengan peubah bebas lainnya konstan.
- $Exp(\text{Tamat SMA sederajat vs Tidak Tamat SD}) = 1.369 \rightarrow$ *Odds* lansia berhenti merokok jika lansia memiliki ijazah terakhir tamat SMA Sederajat adalah 1.3692 kali dibandingkan *odd* yang sama jika lansia tersebut memiliki ijazah tidak tamat SD dengan peubah bebas lainnya konstan.
- $Exp(\text{Tamat Perguruan Tinggi vs Tidak Tamat SD}) = 1.696 \rightarrow$ *Odds* lansia berhenti merokok jika lansia memiliki ijazah terakhir tamat D1/D2/D3/S1/Profesi/S2/S3 adalah 1.696 kali dibandingkan *odd* yang sama jika lansia tersebut memiliki ijazah tidak tamat SD dengan peubah bebas lainnya konstan.
- $Exp(\text{StatusBekerja}) = 0.428 \rightarrow$ *Odds* lansia berhenti merokok jika lansia bekerja adalah 0.428 kali dibandingkan *odd* yang sama jika lansia tersebut tidak bekerja dengan peubah bebas lainnya konstan.
- $Exp(\text{StatusAksesInternet}) = 1.311 \rightarrow$ *Odds* lansia berhenti merokok jika lansia tersebut mengakses internet adalah 1.311 kali dibandingkan *odd* yang sama jika lansia tersebut tidak mengakses internet dengan peubah bebas lainnya konstan.
- $Exp(\text{Umur}) = 1.039 \rightarrow$ *Odds* lansia berhenti merokok akan meningkat 1.039 kali jika usia lansia naik sebesar satu tahun dengan peubah bebas lainnya konstan.

5. Kesimpulan

Hasil penelitian ini menunjukkan model regresi logistik biner dengan dilakukan *random undersampling* merupakan model terbaik dalam mengklasifikasikan lansia yang berhenti merokok. Peubah yang berhubungan dengan berhenti merokok adalah pendidikan, pekerjaan, akses internet, dan usia lansia. Karakteristik lansia yang cenderung untuk berhenti merokok adalah lansia yang pendidikannya lebih tinggi (minimal tamat SMP), tidak bekerja, mengakses internet, dan usia yang lebih tua. Rekomendasi yang diberikan antara lain perlu adanya

sosialisasi resiko merokok dengan lebih menekankan pada lansia yang jenjang pendidikannya tamat SD ke bawah, bekerja, dan lansia muda. Peningkatan promosi kesehatan terkait resiko merokok pada media internet juga perlu ditingkatkan. Penelitian selanjutnya dapat dilakukan dengan menggunakan peubah lain yang diduga memengaruhi lansia berhenti merokok serta menggunakan teknik *resampling* lainnya seperti SMOTE (*Synthetic Minority Oversampling Technique*).

6. Daftar Pustaka

- [1] World Health Organization (WHO). 2011. WHO REPORT on the global TOBACCO epidemic, 2011: Warning about the dangers of tobacco. Italy: World Health Organization
- [2] World Health Organization (WHO). 2012. WHO Global Report: Mortality Attributable of tobacco. Switzerland: World Health Organization
- [3] Badan Pusat Statistik. Statistik Penduduk Lanjut Usia 2020. 2020. Jakarta: Badan Pusat Statistik
- [4] Risdiana, Nurvita dan Proboningrum, Bikassari W. (2019). Perbedaan Tingkat Ketergantungan Merokok Antara Perokok Remaja Dengan Perokok Lansia. *Indonesian Journal of Nursing Practices*. 3(1): 28-34. doi: 10.18196/ijnp.3190
- [5] Burns, DM. (2000). Cigarette Smoking among the Elderly: Disease Consequences and the Benefits of Cessation. *American Journal of Health Promotion*. 14(6): 357-361. doi: 10.4278/0890-1171-14.6.357
- [6] Kementerian Kesehatan RI (Kemenkes). 2017. Analisis Lansia 2017. Jakarta: Pusat Data dan Informasi Kementerian Kesehatan RI
- [7] Hosmer DW, Lemeshow S. 2000. *Applied Logistic Regression*. 2nd ed. New York: J Wiley
- [8] Jian C, Gao J, Ao Y. (2016). A New Sampling Method for Classifying Imbalanced Data Based on Support Vector Machine Ensemble. *Journal Neurocomputing*. 193: 115-122. doi: 10.1016/j.neucom.2016.02.006
- [9] Thanathamath P, Lursinap C. (2013). Handling Imbalanced Datasets with Synthetic Boundary Data Generation Using Bootstrap Resampling and AdaBoost Techniques. *Pattern Recognition Letters. Elsevier BV* 34(12): 1339-1347. doi: 10.1016/j.patrec.2013.04.019
- [10] Prusa J, Khoshgoftaar TM, Dittman DJ, Napolitano A. (2015). Using Random Undersampling to Alleviate Class Imbalance on Tweet Sentiment Data. *IEEE Xplore*. doi: 10.1109/IRI.2015.39
- [11] Agresti A. 2002. *Categorical Data Analysis*. New York: J Willey
- [12] Sirait AM, Pradono Y, Toruan IL. (2002). Perilaku Merokok di Indonesia. *Bul. Penel. Kesehatan*, 30(3): 139-152.
- [13] Rahim FK, Suksaraj T, Jayasvasti I. (2016). Social Determinant of Health of Adult Smoking Behavior: Differences between Urban and Rural Areas in Indonesia. *Kesmas: National Public Health Journal*. 11(2): 51-55. doi:0.21109/kesmas.v11i2.1237
- [14] Juliansyah E, Rizal A. (2018). Faktor Umur, Pendidikan, dan Pengetahuan Dengan Perilaku Merokok Di Wilayah Kerja Puskesmas Sungai Durian, Kabupaten Sintang. *visikES Jurnal Kesehatan Masyarakat*. 17(1): 92-107. doi:10.33633/visikes.v17i01.1853
- [15] Sihombing PR, Arsani AM. (2020). Pengaruh Tingkat Pendidikan, Tingkat Kesejahteraan dan Penghasilan Terhadap Konsumsi Rokok Harian dari Penduduk Dewasa di Indonesia Tahun 2015. *Bappenas Working Papers*. 3(1): 75-87. doi:10.47266/bwp.v3i1.57
- [16] Sukmono R, Salam R. (2019). Pengaruh Faktor Demografi terhadap Status Merokok Remaja sebagai Dampak Alokasi Dana Bagi Hasil Cukai Hasil Tembakau (DBHCHT). *Jurnal Perspektif Bea dan Cukai*, 3(2): 51-70. doi: 10.31092/jpbc.v3i2.435
- [17] Sadarang RAI. (2021). Factors Associated With Quitting Smoking in Indonesia. *J Prev Med Public Health*, 54(2): 137-144. doi:10.3961/jpmp.20.293
- [18] Ramsey MW, Chen-Sankey JC, Reese J, Choi K. (2019). Association between marital status and cigarette smoking: Variation by race and ethnicity. *Prev Med*, 119: 48-51. doi:10.1016/j.ypmed.2018.12.010.
- [19] Reskiaddin LO, Supriyati. (2021). Proses Perubahan Perilaku Berhenti Merokok: Studi Kualitatif Mengenai Motif, Dukungan Sosial dan Mekanisme Coping. *Perilaku dan Promosi Kesehatan: Indonesian Journal of Health Promotion and Behavior*. 3(1): 58-70. doi: 10.47034/ppk.v3i1.4142
- [20] Lestari VD, Umamah M, Pramasari AL, Dharmawan Y. (2013). Smoking Effect Video Learning Berbasis Mobile Sebagai Media Penyuluhan Kesehatan Anti Rokok. *Jurnal Ilmu Mahasiswa*, 3(1): 15-20. <https://ejournal.undip.ac.id/index.php/jim/article/view/10866>
- [21] Brown J, Michie S, Geraghty AWA, Yardley L, Gardner B, Shahab L, Stapleton JA, West R. (2014). Internet-based intervention for smoking cessation (StopAdvisor) in people with low and high socioeconomic status: a randomised controlled trial. *The Lancet Respiratory Medicine* 2(12): 997-1006. doi:10.1016/S2213-2600(14)70195-X

- [22] Taylor GMJ, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J.(2017). Internet-based interventions for smoking cessation (Review). *Cochrane Database of Systematic Reviews*, 2017(7): CD007078: 1-129. doi:10.1002%2F14651858.CD007078.pub5
- [23] Hiscock R,Dobbie F, Bauld L.(2015). Smoking Cessation and Socioeconomic Status:An Update of Existing Evidence from a National Evaluation of English Stop Smoking Services. *Hindawi Publishing Corporation: BioMed Research International*, 205:ID274056 : 1-10.doi:10.1155/2015/274056
- [24] Qiu D, Chen T, LiuT, Song F.2020. Smoking cessation and related factors in middle-aged and older Chinese adults:Evidence from a longitudinal study. *PLoS ONE*, 15(10): 1-9. doi:10.1371/journal.pone.0240806