

PENERAPAN METODE *RANDOM FOREST* UNTUK KLASIFIKASI WANITA USIA SUBUR DI PERDESAAN DALAM MENGGUNAKAN INTERNET (SDKI 2017)

Intan Maulida Khoirun Nisa
Politeknik Statistika STIS

Rani Nooraeni
Politeknik Statistika STIS

ABSTRAK, Internet adalah salah satu media penyebarluasan pengetahuan mengenai alat kontrasepsi. Alat kontrasepsi diketahui sebagai pengendali fertilitas. Faktannya, masih terdapat 32 persen desa di Indonesia yang belum dapat mengakses internet. Berdasarkan data SDKI 2017, tingkat fertilitas di desa relatif lebih tinggi daripada di kota. Fertilitas yang tinggi di daerah pedesaan disebabkan oleh rendahnya pengetahuan mengenai penggunaan alat kontrasepsi. Berdasarkan data SDKI pula, kebutuhan alat atau cara KB belum terpenuhi di pedesaan. Oleh karena itu, penelitian ini bertujuan untuk mengetahui faktor-faktor yang memengaruhi WUS di pedesaan dalam menggunakan internet menggunakan model *random forest*. Model *random forest* dapat digunakan untuk memprediksi variabel prediktor. Model yang dibentuk menggunakan 80 persen dari data total dengan jumlah data sebanyak 35.939 data dan data *testing* yang digunakan sebanyak 8.984 data. Terdapat tiga variabel yang digunakan dalam penelitian ini yaitu umur, tingkat pendidikan yang ditamatkan, dan indeks kekayaan. Seluruh variabel yang didapat dari SDKI tahun 2017 dengan beberapa modifikasi. Dalam model yang terbentuk, variabel yang memiliki kontribusi terbesar adalah tingkat pendidikan yang ditamatkan.

Kata Kunci: *random forest, wanita usia subur, pengguna internet*

1. PENDAHULUAN

Perkembangan teknologi yang sangat pesat mengharuskan masyarakat untuk mengakses internet sebab internet memudahkan masyarakat dalam mendapatkan informasi. Menurut Kemenkominfo (2015), meskipun layanan dan fasilitas untuk menggunakan internet di wilayah desa sudah diterapkan sejak tahun 2009, namun masih banyak penduduk di wilayah pedesaan yang belum mencoba untuk menggunakan internet. Keberadaan internet dapat menambah informasi bagi siapapun termasuk wanita usia subur. Secara rata-rata angka fertilitas total di pedesaan selalu lebih tinggi dibandingkan angka fertilitas di perkotaan. Pada tahun 2017, angka fertilitas total di pedesaan adalah sebesar 2,6,

sedangkan di perkotaan memiliki angka kelahiran total sebesar 2,3.

Terdapat perbedaan pengetahuan mengenai kesehatan reproduksi pada siswa SMA yang berasal dari pedesaan dan perkotaan, salah satu penyebabnya adalah perbedaan jumlah sumber informasi [2]. Sedangkan untuk membatasi tingginya fertilitas diperlukan pengetahuan mengenai penggunaan alat kontrasepsi yang bisa diperoleh dari internet. Pemberian konseling KB dapat memengaruhi keputusan pasangan usia subur untuk memilih alat kontrasepsi sebagai penundaan kehamilan [2]. Segala jenis alat kontrasepsi berpengaruh untuk mengurangi angka kelahiran [4]. Di desa masih terdapat banyak wanita usia subur yang tingkat pendidikannya di bawah SMA, sehingga lebih dibutuhkan internet untuk mengakses informasi yang lebih luas mengenai alat kontrasepsi.

Berdasarkan data survey APJI (Asosiasi Penyelenggara Data Internet) tahun 2018, pengguna internet pedesaan di Indonesia hanya sebesar 32 persen. Oleh karena itu, wilayah pedesaan di Indonesia dijadikan sebagai area sasaran untuk melakukan penelitian pengguna internet. Faktor yang mempengaruhi perilaku konsumsi media internet secara parsial adalah usia [5]. Orang yang berusia tua cenderung lebih rendah dalam penggunaan internet untuk mencari informasi dibanding yang berusia muda.

Berdasarkan latar belakang yang telah dipaparkan di atas, maka peneliti ingin mengetahui klasifikasi wanita usia subur di pedesaan berdasarkan variabel umur, tingkat pendidikan, dan indeks kekayaan di Indonesia. Metode Random Forest dipilih untuk penelitian ini karena tepat digunakan pada data yang timpang (*imbalanced data*) dan memiliki kesalahan (*error*) yang lebih kecil.

2. TINJAUAN PUSTAKA

Variabel-variabel yang dijadikan amatan dalam penelitian ini adalah umur, tingkat pendidikan yang ditamatkan, dan indeks kekayaan.

UMUR DENGAN PENGGUNAAN INTERNET

Berdasarkan KBBI, umur atau usia adalah satuan waktu yang mengukur waktu keberadaan suatu benda atau makhluk baik yang hidup maupun yang mati. Berdasarkan data hasil survei APJII tahun 2017 menunjukkan bahwa variabel umur memengaruhi terhadap penggunaan internet, yang menunjukkan bahwa semakin tua seseorang kecenderungan untuk mengakses internet semakin kecil. Berdasarkan survei tersebut didapatkan bahwa proporsi pengguna internet di Indonesia untuk usia di bawah 34 tahun berkisar 67 persen. Sementara sisanya berada diatas usia 34 tahun. Hal ini didukung oleh sebuah jurnal penelitian berjudul "Analisis Perbedaan Frekuensi Penggunaan Internet Berdasarkan Faktor Demografi" dengan membagi menjadi beberapa kelompok umur. Untuk usia 25 tahun ke bawah memiliki proporsi sebesar 69 persen, sedangkan sisanya di usia 25 tahun ke atas menunjukkan terjadi perbedaan proporsi penggunaan internet untuk variabel usia.

TINGKAT PENDIDIKAN DENGAN PENGGUNAAN INTERNET

Penelitian berkaitan dengan Pengaruh Usia, Tingkat Pendidikan, dan Jenis Kelamin terhadap Konsumsi Media mengungkap bahwa usia, tingkat pendidikan, dan jenis kelamin merupakan faktor demografis yang berpengaruh terhadap pemilihan media yang akan dikonsumsi[8]. Dimana jika dilihat dari sudut pandang tingkat pendidikannya saja, maka dapat dikatakan bahwa semua responden, yang berada pada jenjang pendidikan apapun adalah pengguna internet, namun mereka yang berusia 40 tahun ke atas dengan mayoritas tamatan pendidikannya adalah sarjana, magister, dan doktoral tidak memilih internet sebagai konsumsi media yang utama.

INDEKS KEKAYAAN DENGAN PENGGUNAAN INTERNET

Dalam penelitian yang dilakukan oleh Wahid (2005), menyatakan bahwa hambatan utama

mengenai tingkat adopsi internet yang masih rendah dipengaruhi oleh biaya akses yang masih sangat mahal. Setiap orang memiliki persepsi yang berbeda tentang mahal tidaknya suatu biaya tergantung keadaan ekonomi atau tingkat kesejahteraan yang dimiliki. *Demographic and Health Survey*, salah satu metode yang dikembangkan perihal status sosial ekonomi suatu rumah tangga ialah *asset index* atau *wealth index*[6]. Indeks kekayaan (*wealth index*) merupakan karakteristik latar belakang yang digunakan dalam laporan sebagai pendekatan untuk mengukur standar hidup rumah tangga dalam jangka panjang. Metode pengukuran indeks ini dikembangkan oleh Filmer dan Pritchett[7]. Indeks ini disusun dari variabel-variabel yang menggambarkan kepemilikan barang dan karakteristik perumahan, sehingga sering pula dikenal dengan *asset index*. Keduanya memiliki arah pengertian yang sama, *asset* dikarenakan indeks ini disusun dari kepemilikan aset oleh setiap rumah tangga, sedangkan *wealth* mengarah kepada kepemilikan aset tertentu oleh suatu rumah tangga menunjukkan derajat kekayaan rumah tangga tersebut.

3. METODOLOGI

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2017. Sampel dalam penelitian ini adalah penduduk usia subur usia 15-49 tahun. Sampel yang digunakan ini berjumlah 44.923 sampel yang terkonsentrasi pada daerah pedesaan.

Dalam penelitian ini variabel dependen yang digunakan yaitu status penggunaan internet, yang memiliki 2 kategori yaitu:

Y = 0 untuk tidak pernah menggunakan internet

Y = 1 untuk pernah menggunakan internet

Variabel independen yang digunakan didasarkan beberapa penelitian yang relevan. Adapun variabel independen yang digunakan dalam penelitian ini disajikan dalam tabel berikut.

No	Variabel	Keterangan
1.	Usia	0 = 30-49 1 = 15-29

2. Pendidikan	0 = tinggi 1 = rendah
3. Wealth Index	0 = kaya 1 = miskin 2 = menengah

Prosedur Analisis

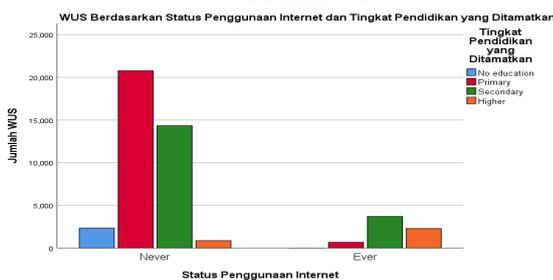
Analisis yang digunakan dalam penelitian ini adalah dengan pendekatan *Random forest*. Skema Random Forest pertama kali dicetuskan oleh Breiman untuk membangun prediktor dengan sekumpulan pohon keputusan yang berkembang secara acak pada subruang data. *Output* akhir dari metode ini adalah akan dihasilkannya satu set “pohon” keputusan yang berasal dari modus setiap pohon yang ada[9]. Variabel dependen y terdiri dari 2 kategori yaitu sukses dan gagal. Adapun langkah-langkah dalam melakukan analisis dengan metode *random forest* adalah sebagai berikut.

1. Mempersiapkan seluruh variabel baik variabel dependen maupun variabel independent.
2. Melakukan pengkodean pada setiap kategorik variabel.
3. Melakukan pengubahan tipe data pada variabel kategorik melalui fungsi *as.factor()*
4. Melakukan pemisahan data dengan proporsi 4 : 1 untuk data yang dibuat menjadi model dan data yang dibuat untuk pengujian akurasi tebakan model.
5. Membuat model *random forest* dengan data model (*data training*).
6. Melakukan pengujian menggunakan data pengujian (*data testing*).
7. Membuat tabel konfusi (*confussion matrix*).
8. Menghitung akurasi tebakan benar.

4. PEMBAHASAN

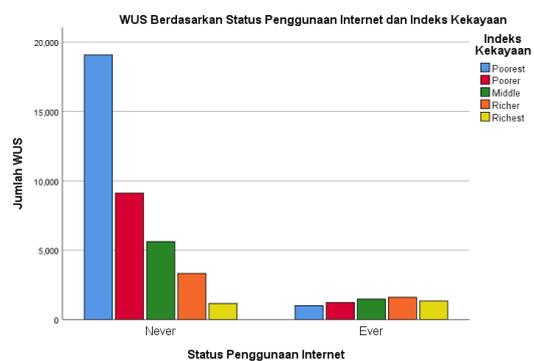
Analisis Deskriptif

Grafik 4.1



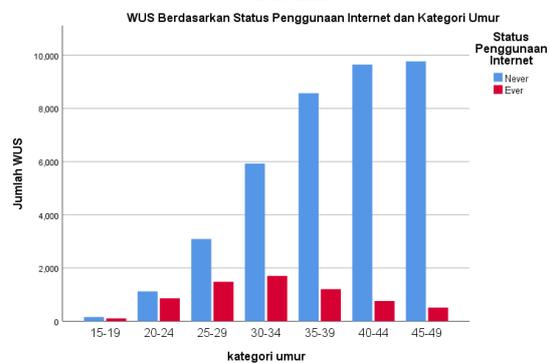
Berdasarkan grafik di atas, dapat dilihat bahwa Wanita Usia Subur yang ada di perdesaan lebih banyak yang berstatus bukan pengguna internet. Jika dilihat dari status penggunaannya, WUS yang menggunakan internet paling banyak pada pendidikan tertinggi Sekolah Menengah. Sedangkan pada WUS yang bukan pengguna internet, paling banyak berasal dari mereka yang hanya menamatkan pendidikan dasar. Hal ini berbeda dengan temuan jurnal terkait yang mengatakan bahwa semua penduduk dengan jenjang pendidikan apapun termasuk pengguna internet (Nur, Aulia, 2014).

Grafik 4.2



Dari grafik tersebut, dapat dilihat bahwa WUS yang menggunakan internet tidak terdapat perbedaan yang signifikan jika dikelompokkan berdasarkan indeks kekayaannya. Sedangkan pada WUS yang bukan pengguna internet, paling banyak berasal dari mereka yang indeks kekayaannya berstatus miskin.

Grafik 4.3

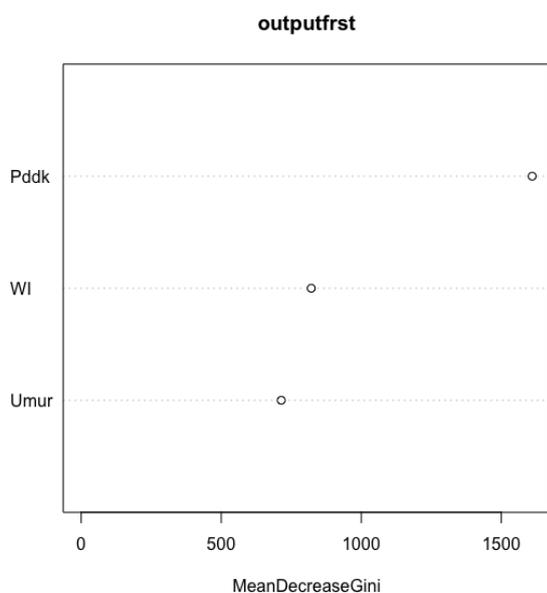


Berdasarkan grafik di atas, kelompok umur Wanita Usia Subur yang menggunakan internet cenderung pada usia 25 hingga 39 tahun.

Sedangkan pada WUS yang tidak menggunakan internet, semakin tua usia WUS, semakin banyak yang tidak menggunakan internet. Hal ini sejalan dengan temuan jurnal terkait yang mengatakan bahwa, penduduk yang berusia 40 tahun ke atas tidak memilih internet sebagai konsumsi media yang utama (Aulia, 2014).

Analisis Inferensia

Grafik 4.4



Nilai *Mean Decrease Gini* menunjukkan seberapa besar kontribusi suatu variabel. Semakin tinggi *Mean Decrease Gini* berarti semakin besar kontribusi variabel tersebut. Berdasarkan model *random forest* di atas dapat disimpulkan bahwa variabel tingkat pendidikan sangat penting karena memiliki pengaruh yang paling signifikan terhadap variabel penggunaan internet wanita usia subur di daerah pedesaan. Hal ini ditunjukkan oleh nilai *Mean Decrease Gini* yang paling tinggi daripada variabel yang lain, menunjukkan bahwa semakin tinggi pendidikan seseorang sangat berpengaruh terhadap pengetahuan dan wawasan mengenai teknologi sehingga kebutuhan untuk mengakses internet lebih tinggi dibandingkan dengan orang yang memiliki tingkat pendidikan lebih rendah. Hasil penelitian ini, sejalan dengan temuan jurnal terkait dari Aulia (2014). Menurut Aulia (2014), semua responden yang berada pada jenjang pendidikan apapun adalah pengguna internet. Kemudian variabel dengan kontribusi kedua

adalah variabel *Wealth Index*. Sedangkan variabel dengan pengaruh terendah adalah variabel umur. Hal tersebut menunjukkan bahwa kecenderungan seseorang untuk mengakses internet tidak terlalu memandang perbedaan umur.

```
Confussion matrix
> prediksiap2=predict(outputfrst,trainapgw)
> cm6=table(trainapgw$pikode,prediksiap2)
> cm6
  prediksiap2
    0      1
0 29729  901
1  2823 2486
> sum(diag(prop.table(cm6)))
[1] 0.89638
```

Pada *output confusion matrix* di atas menunjukkan angka 0.89638, yang berarti model tersebut mampu memprediksi dengan tepat sebesar 89 persen data *training*. Data ini diambil sejumlah 80 persen dari total observasi dan digunakan untuk membentuk model.

```
> cm5=table(testapgw$pikode,prediksiapg)
> cm5
  prediksiapg
    0      1
0  7420  237
1   711  616
> sum(diag(prop.table(cm5)))
[1] 0.8944791
```

Data testing di atas diambil 20 persen dari total observasi dan digunakan untuk memvalidasi apakah model tersebut sudah baik atau belum. *Output confusion matrix* di atas menunjukkan bahwa observasi yang tidak diprediksi dengan tepat oleh model proporsinya hanya sebesar 8 persen dari total data *testing*. Sedangkan angka 0.89638 menunjukkan model tersebut mampu memprediksi dengan tepat sebesar 89 persen data testing.

5. KESIMPULAN

Pembahasan hasil menunjukkan bahwa Wanita Usia Subur di wilayah pedesaan di Indonesia dapat diklasifikasikan menurut faktor umur, tingkat pendidikan yang ditamatkan, dan indeks kekayaan. Tingkat pendidikan yang ditamatkan

memiliki kontribusi pengaruh terbesar, selanjutnya diikuti oleh indeks kekayaan, dan umur. *Output confusion matrix* menunjukkan bahwa observasi yang tidak diprediksi dengan tepat oleh model proporsinya hanya sebesar 8 persen dari total data *testing*. Sedangkan angka 0.89638 menunjukkan model tersebut mampu memprediksi dengan tepat sebesar 89 persen data *testing*.

Faktor Demografi. *Jurnal Eksekutif Volume 10 No 1*.

- [9] Wahid, Fathul. (2005). *Apakah perempuan Indonesia Terbelakang dalam Adopsi Internet?*: Temuan Empiris*. Yogyakarta: Universitas Islam Indonesia.

6. DAFTAR PUSTAKA

- [1] Chen, Chao (2004). *Using Random Forest in Learn Imbalanced Data*. UC Berkeley: Department of Statistic
- [2] Susanto, A. (2015). Faktor-Faktor yang Memengaruhi Penggunaan Internet Masyarakat Desa Pasar VI Kualanamu, Deli Serdang, Sumatera Utara. *Jurnal Penelitian Pos dan Informatika, Volume 5 No 1*.
- [3] Sari, Silviana Kartika dkk. (2010). Hubungan Konseling Keluarga Berencana (Kb) Dengan Pengambilan Keputusan Pasangan Usia Subur (Pus) Dalam Penggunaan Alat Kontrasepsi. *Bidan Prada, Jurnal Ilmiah Kebidanan, Vol. 1 No. 1 Edisi Desember 2010*
- [4] Felicia, Zelia dkk. (2015). Pengaruh Penggunaan Berbagai Jenis Alat Kontrasepsi terhadap Kelahiran di Kota Surabaya. *Jurnal Statistika Vol. 8 No. 1*
- [5] Nur, Aulia. (2014). *Pengaruh Usia, Tingkat Pendidikan, dan Jenis Kelamin terhadap Perilaku Konsumsi Media*. Semarang: Universitas Diponegoro.
- [6] Rahani, Rini. (2012). *Variabel-Variabel yang Memengaruhi Status Kekayaan Rumah Tangga dengan Kepala Rumah Tangga Janda Bekerja Berdasarkan Wealth Index di Indonesia Tahun 2017*. Jakarta: Sekolah Tinggi Ilmu Statistik.
- [7] Filter, D., & Prichett, L. H. (2012). Estimating Wealth Effect Without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India. *Demography 38 (1)*.
- [8] Kristiani, N. (2013). Analisis Perbedaan Frekuensi Penggunaan Internet Berdasarkan