

Mengisi Gap Informasi Persentase Penduduk Miskin di Jawa Timur dengan Google Trend Index

(Filling the Information Gap on the Percentage of Poor People in East Java by Google Trend Index)

Ajiwasesa Harumeka^{1*}, Tasmilah²

¹Badan Pusat Statistik Provinsi Jawa Timur

²Badan Pusat Statistik Kota Malang

Jalan Raya Kendangsari Industri No. 43-44, Surabaya, Jawa Timur

Email : ajiwasesa@bps.go.id

ABSTRAK

Tujuan *Sustainable Development Goals* (SDGs) yang pertama adalah mengakhiri kemiskinan dalam segala bentuk dimanapun. Salah satu aspek yang penting untuk mengentaskan kemiskinan adalah tersedianya data kemiskinan yang baik. Data kemiskinan yang baik dapat digunakan untuk evaluasi program pengentasan kemiskinan pemerintah, dan dapat dibandingkan antar waktu. Badan Pusat Statistik (BPS) menghitung angka kemiskinan dua kali dalam satu tahun, yaitu pada Bulan Maret dan Bulan September. Hal tersebut menimbulkan “gap informasi”, yaitu tidak tersedianya data diantara waktu penghitungan tersebut. *Google Trend Index* (GTI) sebagai salah satu jenis *big data* dapat dimanfaatkan oleh *National Statistics Offices* (NSO) untuk mengisi “gap informasi” tersebut. Penelitian ini menggunakan GTI untuk memprediksi persentase penduduk miskin bulanan di Jawa Timur. Metode yang digunakan adalah *Feed Forward Neural Network* (FFNN). Model FFNN terbaik menghasilkan *Root Mean Square Error* (RMSE) sebesar 0,1279 dengan 1 *hidden layer* dan 1 *neuron* dalam *hidden layer* tersebut. Model terbaik tersebut digunakan untuk melakukan prediksi persentase penduduk miskin bulan Januari hingga Desember 2021.

Kata kunci: *Google Trend Index*, Kemiskinan, *Neural Network*, Prediksi

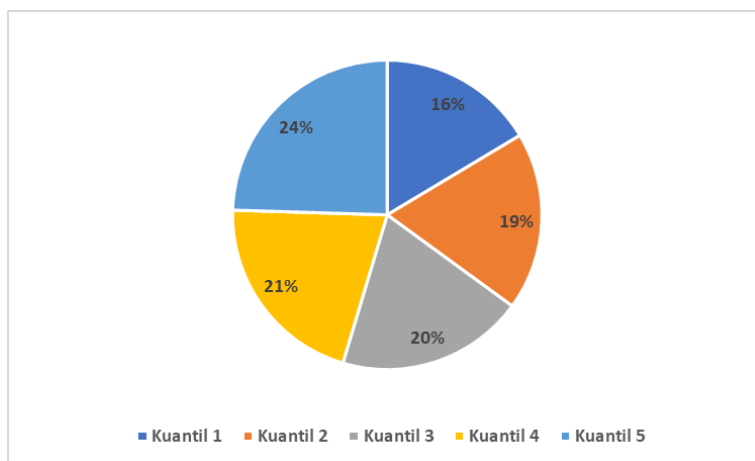
ABSTRACT

The first goal of the *Sustainable Development Goals* (SDGs) is to end poverty in all its forms everywhere. One of the important aspects for alleviating poverty is the availability of good poverty data. Good poverty data can be used to evaluate government poverty alleviation programs, and can be compared over time. The Central Statistics Agency (BPS) calculates the poverty rate twice a year, in March and September. This creates an “information gap”, namely the unavailability of data between the times of the calculation. The *Google Trend Index* (GTI) as a type of *big data* can be utilized by the *National Statistics Offices* (NSO) to fill this “information gap”. This study used GTI to predict the monthly percentage of poor people in East Java. The method used was *Feed Forward Neural Network* (FFNN). The best FFNN model produced a *Root Mean Square Error* (RMSE) of 0.1279 with 1 *hidden layer* and 1 *neuron* in the *hidden layer*. The best model was used to predict the percentage of poor people from January to December 2021.

Keywords: *Google Trend Index*, *Neural Network*, Poverty, Prediction

PENDAHULUAN

Tujuan *Sustainable Development Goals* (SDGs) yang pertama adalah mengakhiri kemiskinan dalam segala bentuk dimanapun. Salah satu aspek yang penting untuk mengentaskan kemiskinan adalah tersedianya data kemiskinan yang baik. Data kemiskinan yang baik dapat digunakan untuk evaluasi program pengentasan kemiskinan pemerintah, dan dapat dibandingkan antar waktu. Badan Pusat Statistik (BPS), sebagai lembaga non kementerian yang menyediakan data dan statistik di Indonesia, mencatat penduduk miskin di Jawa Timur pada bulan September 2021 sebesar 10,59 persen. Angka tersebut lebih tinggi dibandingkan angka nasional, yaitu sebesar 9,71 persen (BPS, 2021b). BPS menggunakan *basic needs approach* untuk mengukur angka kemiskinan tersebut. Rumah tangga dikategorikan miskin jika rata-rata pengeluaran konsumsi perkapitanya lebih kecil dari garis kemiskinan (BPS, 2021a). Pengeluaran konsumsi rumah tangga ini didapat dari Survei Sosial Ekonomi Nasional (Susenas). Sayangnya, survei ini hanya dilakukan 2 kali dalam setahun. Beberapa negara bahkan tidak dapat menyelenggarakan survei yang serupa karena mahalnya biaya yang diperlukan (ADB, 2020). Hal tersebut menimbulkan “gap informasi”, yaitu tidak tersedianya data diantara waktu pelaksanaan Susenas. “*Big Data*” dapat dimanfaatkan oleh *National Statistics Offices* (NSO) dengan biaya yang relatif lebih kecil (ADB, 2020) serta dapat mengisi gap informasi tersebut (Fasulo et al., 2017). *Google Trend Index* (GTI) merupakan salah satu jenis “*Big Data*” yang dapat digunakan.



Sumber : Susenas Maret 2021, diolah

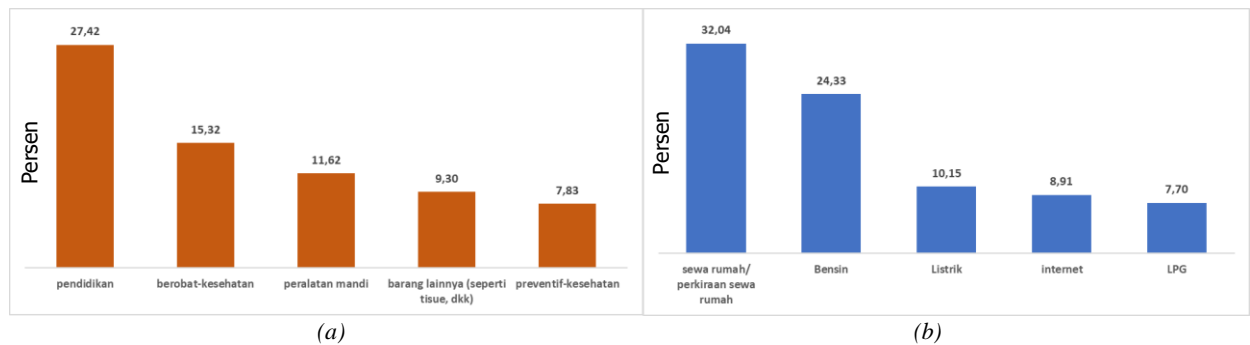
Gambar 1. Distribusi Persentase Penduduk 10 Tahun ke Atas yang Mengakses Internet menurut Kuantil Pengeluaran Perkapita

GTI merupakan layanan yang diberikan oleh *Google* yang dapat diakses secara umum (Afkham et al., 2021). Setiap orang yang terhubung internet dan melakukan pencarian informasi menggunakan *Google* direkam, kemudian diolah menjadi indeks yang disebut GTI. GTI tergantung pada literasi internet penduduk. Semakin tinggi literasi internet maka semakin besar pula peluang untuk mengakses mesin pencarian *Google*, kontribusi penduduk terhadap GTI semakin meningkat.

Literasi internet penduduk Jawa Timur cukup tinggi. BPS mencatat persentase penduduk 5 tahun ke atas yang mengakses internet sebesar 59,34 persen (BPS, 2021c). Gambar 1 menunjukkan bahwa pengguna internet di Jawa Timur tidak hanya dari kalangan ekonomi atas, ditunjukkan oleh kuantil 5, akan tetapi juga dari kalangan ekonomi menengah ke bawah, ditunjukkan oleh kuantil 1 hingga 4. Bahkan pengguna internet dari kalangan ekonomi menengah ke bawah lebih banyak dibandingkan kalangan ekonomi atas, yaitu sebesar 76 persen. Kalangan ekonomi menengah ke bawah berpeluang lebih besar untuk masuk dan keluar dari jurang kemiskinan dibandingkan kalangan ekonomi atas. Kalangan ekonomi atas relatif lebih stabil berada di luar jurang kemiskinan.

GTI dapat mengungkapkan ketertarikan masyarakat terhadap topik tertentu (Afkham et al., 2021). Misalnya seseorang menggunakan mesin pencarian *Google* dengan kata kunci “harga mobil” berarti orang itu tertarik untuk membeli atau menjual mobil. Dengan kata lain GTI dapat menggambarkan “persiapan penduduk untuk mengkonsumsi” suatu komoditas tertentu (Perez et al., 2017). Berkaitan dengan hal tersebut, GTI dapat digunakan untuk memprediksi permintaan, pengeluaran atau konsumsi komoditas tertentu. Beberapa penelitian sebelumnya menyebutkan bahwa GTI dapat digunakan sebagai pendekatan untuk memprediksi pengeluaran konsumsi rumah tangga triwulanan di Italia dengan akurasi yang baik dengan menggunakan 10 kata kunci pencarian (Fasulo et al., 2017). Hasil yang lain menyebutkan bahwa GTI dapat dijadikan sebagai pendekatan untuk memprediksi permintaan bahan bakar eceran di USA dengan menggunakan 2 kata kunci pencarian (Afkham et al., 2021). GTI juga digunakan untuk peramalan beberapa indikator lain, seperti Tingkat Penghunian Kamar (Ayuningtyas & Wirawati, 2019) dan Produk Domestik Bruto (Götz & Knetsch, 2019). Di sisi lain, dengan pendekatan *basic needs approach*, pengkategorian penduduk miskin tergantung pada nilai pengeluaran konsumsinya. Penduduk dikategorikan miskin jika pengeluaran konsumsinya di bawah garis kemiskinan. Hal tersebut memungkinkan GTI dijadikan sebagai pendekatan dalam memprediksi persentase penduduk miskin.

Permasalahan yang muncul ketika menggunakan GTI adalah pemilihan kata kunci pencarian yang akan digunakan sebagai variabel prediktor. Fasulo et al. (2017) menggunakan kata kunci pencarian yang berkaitan dengan komoditas non makanan. Berkaitan dengan hal tersebut, berdasarkan data Susenas Maret Tahun 2021, pengeluaran konsumsi non makanan penduduk Jawa Timur yang terbesar terdapat pada kelompok perumahan/fasilitas rumah tangga dan aneka barang/jasa, secara berurutan sebesar 24,4 persen dan 12,66 persen (BPS, 2021c). Gambar 2 menjelaskan bahwa di kalangan penduduk kuantil 1 sampai dengan 3, pengeluaran konsumsi dengan *share* terbesar pada kelompok perumahan/fasilitas rumah adalah biaya sewa rumah/perkiraan sewa rumah, bensin, listrik, internet, dan LPG, sedangkan pada kelompok barang/jasa meliputi biaya pendidikan, berobat, peralatan mandi, preventif kesehatan, dan barang lainnya seperti *tissue*, *pampers*, dan lain-lain. Data tersebut digunakan sebagai dasar pemilihan kata kunci pencarian.



Sumber : Susenas Maret 2021, diolah

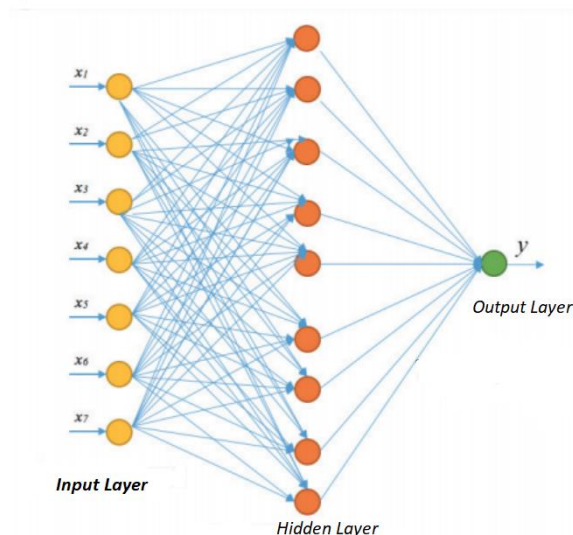
Gambar 2. 5 Komoditas Terbesar yang Dikonsumsi Penduduk Kuantil 1 Hingga Kuantil 3 di Jawa Timur: (a) Kelompok Aneka Barang/Jasa (persen), (b) Kelompok Perumahan/Fasilitas Rumah Tangga (persen)

GTI tersedia dalam *series* bulanan. Seperti yang telah disebutkan sebelumnya bahwa GTI juga dapat menggambarkan persiapan penduduk untuk mengkonsumsi suatu komoditas. Hal tersebut merupakan kelebihan yang dapat dimanfaatkan sebagai pendekatan untuk memprediksi persentase penduduk miskin bulanan di Jawa Timur. Paper ini terdiri dari 5 bab. Bab 1 berisi latar belakang penelitian. Bab 2 menjelaskan data dan metode yang digunakan. Bab 3 berisi hasil dan pembahasan. Bab 4 berisi kesimpulan. Terakhir, bab 5 terdiri dari referensi yang digunakan.

METODE

Neural Network (NN)

Neural Network (NN) merupakan metode yang diilhami oleh proses pada jaringan otak manusia. NN telah banyak digunakan untuk memprediksi cuaca, harga saham, hasil dari suatu pertandingan, seperti sepak bola, basket, dan pacuan kuda (Williams & Li, 2008). Selain itu NN juga digunakan untuk memprediksi data finansial seperti nilai tukar mata uang (Putri et al., 2021), *inflow-outflow* uang di Indonesia (Suhartono et al., 2019). Metode ini mampu mengenali pola hubungan data yang tidak linier (Suhartono et al., 2019). Tetapi juga baik untuk pola data yang linier (Bhakti, 2019).



Sumber : (Prathama et al., 2018)

Gambar 3. Arsitektur FFNN dengan 1 Hidden Layer

Terdapat 2 jenis struktur dalam NN, yaitu *Recurrent Neural Network* (RNN) dan *Feed Forward Neural Network* (FFNN). RNN merupakan NN yang dirancang untuk memproses data *sequential*. Pada RNN terdapat *looping* yang membuat informasi masa lalu tetap tersimpan. FFNN merupakan NN yang paling sederhana, informasi masuk dalam satu arah. Tidak ada proses *looping* seperti pada RNN. Penelitian ini menggunakan FFNN. Terdapat 3 layer pada FFNN, yaitu *input*, *hidden*, dan *output*. FFNN dapat memiliki lebih dari 1 *hidden layer*, atau biasa disebut dengan *Deep Neural Network* (DNN). Informasi yang diterima dari *input layer*

diproses secara berurutan pada setiap layer hingga mencapai *output layer* (Suhartono et al., 2019). Pada setiap *layer*, terdapat satu atau lebih *neuron*. Gambar 3 menunjukkan contoh arsitektur FFNN dengan 1 *hidden layer* yang memiliki 9 *neuron*, *input layer* yang memiliki 7 *neuron*, dan *output layer* yang memiliki 1 *neuron*. Dalam pemodelan, jumlah *neuron* pada *input layer* menggambarkan jumlah variabel prediktor, sedangkan jumlah *neuron* pada *output layer* menggambarkan jumlah variabel respon. Arsitektur terbaik dipilih dari ujicoba jumlah *hidden layer* dan jumlah *neuron* pada setiap *hidden layer* (Prathama et al., 2018). Jumlah *hidden layer* yang ditetapkan pada penelitian ini adalah 1 dan 2, sedangkan jumlah *neuron* pada setiap *hidden layer* akan diujicoba dari 1 hingga 5.

Neuron saling terhubung melalui penimbang. Setiap *neuron* melibatkan proses penghitungan jumlah input tertimbang kemudian mentransfer hasilnya menggunakan sebuah fungsi aktivasi. Pada *hidden layer* pertama, secara matematis proses tersebut dideskripsikan melalui persamaan 1.

$$\hat{y}_{ji} = f_j \left(\sum_{k=1}^p w_{jk} x_{ki} + b_j \right), \dots \dots \dots (1)$$

dengan

\hat{y}_{ji} = estimasi nilai observasi ke-*i* pada *neuron* ke-*j*,

i = 1, 2, ..., *n*,

n = jumlah observasi,

j = 1, 2, ..., *m*,

m = jumlah *neuron*,

p = jumlah variabel prediktor,

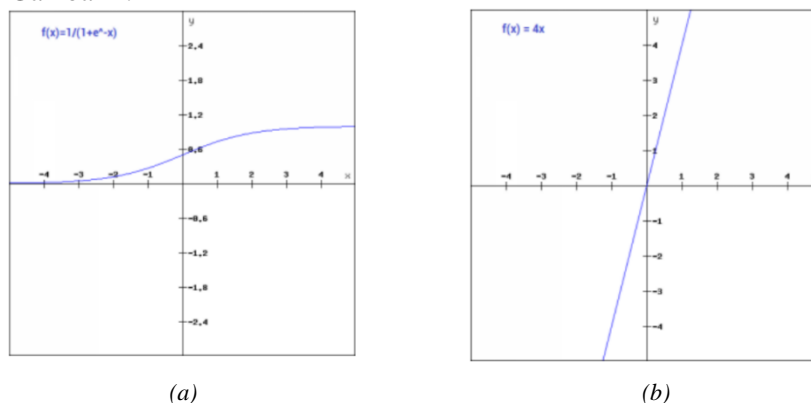
w_{jk} = penimbang variabel prediktor ke-*k* pada *neuron* ke-*j*,

x_{ki} = nilai variabel prediktor ke-*k* pada observasi ke-*i*,

b_j = nilai bias pada *neuron* ke-*j*,

f_j = fungsi aktivasi pada *neuron* ke-*j*,

Fungsi aktivasi yang digunakan di *hidden* dan *input layer* pada penelitian ini secara berturut-turut adalah fungsi *sigmoid* dan fungsi linier sebagaimana yang dilakukan oleh Suhartono et al, (2019). Fungsi sigmoid dan linier dideskripsikan oleh Gambar 4.



Sumber : (Sharma et al., 2020)

Gambar 4. Bentuk Fungsi Aktivasi : (a) fungsi sigmoid, (b) fungsi linier

Principal Component Analysis (PCA)

Seiring dengan perkembangan *Big Data*, banyak sekali dijumpai fenomena *high dimensional data*. Pada *supervise learning*, hal tersebut terjadi ketika jumlah observasi, *n*, lebih sedikit dibandingkan jumlah variabel prediktornya, *p*. Fenomena tersebut menyebabkan kesulitan untuk menemukan solusi persamaan matematikanya. Disamping itu, dimensi data yang tinggi akan mempengaruhi ketepatan prediksi (Hediyati & Suartana, 2021). Untuk menanggulanginya diperlukan reduksi dimensi. *Principal Component Analysis* (PCA) adalah salah satu metode tersebut.

PCA akan membentuk kumpulan variabel baru tanpa mengurangi karakteristik data asli secara signifikan. Metode ini mengubah sekumpulan variabel data asli yang saling berhubungan menjadi sekumpulan variabel baru yang saling bebas. Variabel baru tersebut merupakan kombinasi linier dari variabel lama. Secara umum, tahapan dalam PCA adalah sebagai berikut (Jolliffe, 2010):

1. Mendapatkan matriks kovarian, \mathbf{S} , dari matriks \mathbf{X} berukuran $n \times p$. Dimana *n* adalah jumlah observasi dan *p* adalah jumlah variabel prediktor.
2. Mendapatkan *eigen value*, λ , dan *eigen vector*, \mathbf{e} , melalui dekomposisi \mathbf{S} yang memenuhi persamaan 2.

$$A^T SA = \Delta, \dots\dots\dots (2)$$

dengan

A = matrik *orthogonal* yang kolomnya merupakan *eigen vector*, e_1, e_2, \dots, e_p .

Δ = matrik diagonal yang komponen diagonal utamanya merupakan *eigen value*, $\lambda_1, \lambda_2, \dots, \lambda_p$.

3. Memilih λ sebanyak o yang dirasa cukup untuk menjelaskan variasi variabel awal.
4. Membuat variabel baru yang merupakan kombinasi linier antara *eigen vector* dan variabel awal.

Data dan Prosedur yang Digunakan

Sesuai dengan tujuan penelitian, kami menggunakan persentase penduduk miskin yang dipublikasikan oleh Badan Pusat Statistik sebagai variabel respon. Data persentase penduduk miskin yang digunakan adalah periode semester 1 (Maret) Tahun 2012 hingga semester 2 (September) Tahun 2021. Terdapat 21 kata kunci pencarian sebagai calon variabel prediktor. Kata kunci pencarian tersebut berkaitan dengan pengeluaran penduduk dengan *share* terbesar yang telah dibahas pada latar belakang penelitian ini, antara lain :

1. Beli rumah murah
2. Jual rumah murah
3. Jual rumah
4. Rumah disewakan
5. Biaya kuliah
6. Biaya sekolah
7. Biaya Pendidikan
8. Biaya rumah sakit
9. Biaya puskesmas
10. Harga sabun mandi
11. Harga pasta gigi
12. Harga sikat gigi
13. Harga shampo
14. Harga pampers
15. Senam yoga
16. Paket zumba
17. *Medical check up*
18. Harga bensin
19. Biaya listrik
20. Paket Internet
21. Harga LPG

Periode data GTI yang digunakan tersebut mengikuti ketersediaan variabel respon, yaitu januari 2011 hingga desember 2021. Sesuai dengan tujuan penelitian, lokasi data GTI ditentukan Provinsi Jawa Timur.

Berikut ini adalah prosedur yang digunakan untuk melakukan prediksi persentase penduduk miskin bulanan :

1. Melakukan aggregasi data GTI bulanan menjadi semesteran menggunakan rata-rata satu tahun yang lalu. Pemilihan rata-rata satu tahun yang lalu didasarkan pada waktu referensi beberapa pertanyaan komoditas non makanan Susenas, yaitu setahun yang lalu. Misal, data semester 1 atau bulan Maret 2019 adalah nilai rata-rata GTI bulan April 2018 hingga Maret 2019. Agregasi tersebut menghasilkan struktur data seperti pada Tabel 1.

Tabel 1. Struktur Data

Keterangan	Persentase Penduduk Miskin (Variabel Respon)	Beli rumah murah	Jual rumah murah	...	Harga LPG
Maret 2012	Y_1	X_{1-1}	X_{1-2}	...	X_{1-21}
September 2012	Y_2	X_{2-1}	X_{2-2}	...	X_{2-21}
Maret 2013	Y_3	X_{3-1}	X_{3-2}	...	X_{3-21}
September 2013	Y_4	X_{4-1}	X_{4-2}	...	X_{4-21}
.
.
.
Maret 2021	Y_{19}	X_{19-1}	X_{19-2}	...	X_{19-21}
September 2021	Y_{20}	X_{20-1}	X_{20-2}	...	X_{20-21}

- Memilih variabel prediktor yang memiliki nilai koefisien korelasi dengan persentase penduduk miskin lebih dari 0,5 atau kurang dari -0,5. Variabel prediktor yang nilai koefisien korelasinya diluar ketentuan tersebut tidak dijadikan sebagai *dataset* yang akan dimodelkan.
- Membagi *dataset* menjadi *training* dan *testing*. Data *training* adalah data yang digunakan untuk membangun model sedangkan data *testing* adalah data yang digunakan untuk mengevaluasi model yang telah dibentuk berdasarkan data *training*. *Dataset* dengan periode tahun 2012 hingga 2020 dijadikan sebagai data *training*. Sedangkan tahun 2021 sebagai data *testing*.
- Melakukan reduksi dimensi variabel prediktor dengan menggunakan PCA.
- Menambahkan variabel *dummy* yang menjelaskan kejadian COVID19. Variabel *dummy* bernilai 0 pada bulan Maret 2012 hingga September 2019 dan bernilai 1 untuk bulan Maret 2020 hingga September 2021.
- Melakukan *preprocessing* pada data *training* menggunakan metode *standardized*.
- Melakukan pemodelan FFNN.
- Memilih parameter *jumlah hidden layer* dan *jumlah neuron* pada *hidden layer* yang optimal berdasarkan nilai *Root Mean Square Error* (RMSE) yang terkecil pada data *testing*. Nilai RMSE dapat dihitung menggunakan persamaan 3.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \dots \dots \dots (3)$$

dengan

\hat{y}_t = estimasi nilai observasi ke- t

y_t = nilai aktual observasi ke- t

- Menggunakan model FFNN terbaik untuk prediksi persentase penduduk miskin bulanan di Jawa Timur menggunakan data GTI bulanan.

HASIL DAN PEMBAHASAN

Pemilihan Variabel Prediktor

Tabel 2 menggambarkan nilai koefisien korelasi antara 21 kata kunci pencarian dan persentase penduduk miskin di Jawa Timur. Dari 21 kata kunci pencarian tersebut dipilih kata kunci pencarian yang memiliki nilai koefisien korelasi terhadap presentase penduduk miskin di Jawa Timur lebih dari 0,5 atau kurang dari -0,5. Kata kunci yang sesuai dengan ketentuan tersebut dijadikan sebagai variabel prediktor. Terdapat 2 variabel yang tidak sesuai dengan ketentuan, yaitu “Beli rumah murah” dan “Senam yoga” dengan koefisien korelasi secara berturut-turut adalah 0,235 dan -0,118. Hal itu mengindikasikan bahwa kedua kata kunci pencarian tersebut kurang mempengaruhi persentase penduduk miskin di Jawa Timur. Pada Tabel 2, kedua kata kunci pencarian tersebut ditandai dengan *font* tebal.

Tabel 2. Koefisien Korelasi 21 Kata Kunci Pencarian Terhadap Persentase Penduduk Miskin di Jawa Timur

Kata Kunci Pencarian	Koefisien Korelasi	Kata Kunci Pencarian	Koefisien Korelasi
Beli rumah murah	0,235	Harga sabun mandi	-0,801
Biaya kuliah	-0,891	Harga shampoo	-0,835
Biaya listrik	-0,678	Harga sikat gigi	-0,794
Biaya Pendidikan	0,843	Jual rumah murah	0,786
Biaya puskesmas	-0,754	Jual rumah	0,821
Biaya rumah sakit	-0,802	Medical check up	-0,856
Biaya sekolah	-0,821	Paket Internet	0,591
Harga bensin	-0,694	Rumah disewakan	0,551
Harga LPG	-0,695	Senam yoga	-0,118
Harga pampers	-0,746	Zumba	-0,678
Harga pasta gigi	-0,901		

Penentuan *Principal Component* (PC)

Jumlah *Principal Component* (PC) pada penelitian ini ditentukan berdasarkan metode *Kaiser*. Metode tersebut memilih PC yang memiliki *varians* ≥ 1 (Jolliffe, 2010; Hendro et al., 2012). PC dengan *varians* ≤ 1

mengindikasikan kurangnya informasi dari variabel asli sehingga tidak perlu disertakan. Tabel 3 menggambarkan ringkasan dari PC. PC yang digunakan pada penelitian ini ditandai dengan cetak tebal. Berdasarkan metode *Kaiser*, dipilih PC1, PC2, dan PC3. 3 PC yang digunakan tersebut mampu mewakili 94,57 persen variasi dari variabel asli.

Tabel 3. Ringkasan PC

PC	Varians	Proporsi Varians	Proporsi Varians Kumulatif
PC1	15,2119	0,8006	0,8006
PC2	1,5854	0,08344	0,88407
PC3	1,1710	0,06163	0,94570
PC4	0,4167	0,02193	0,96763
PC5	0,1585	0,00834	0,97597
PC6	0,1467	0,00772	0,98369
PC7	0,0934	0,00491	0,98860
PC8	0,0855	0,00450	0,99310
PC9	0,0698	0,00367	0,99678
PC10	0,0297	0,00156	0,99834
PC11	0,0139	0,00073	0,99907
PC12	0,0109	0,00057	0,99965
PC13	0,0035	0,00018	0,99983
PC14	0,0019	0,00010	0,99993
PC15	0,0008	0,00004	0,99997
PC16	0,0004	0,00002	1,00000
PC17	0,0001	0,00000	1,00000
PC18	0,0000	0,00000	1,00000
PC19	0,0000	0,00000	1,00000

PC1, PC2, dan PC3 merupakan variabel baru yang merupakan kombinasi linier dari 19 variabel prediktor GTI. Persamaan PC1, PC2, dan PC3 secara berturut-turut dituliskan pada persamaan 4, 5, dan 6.

$$PC1 = -0,22475 * biyakuliah - 0,2397 * biyalistrik + 0,2006 * biyapendidikan - 0,251 * biyapuskesmas - 0,239 * biayarumahsakit - 0,2528 * biyasekolah - 0,218 * hargabensin - 0,2287 * hargalpg - 0,2506 * hargapampers - 0,243 * hargapastagigi - 0,25037 * hargasabunmandi - 0,2519 * hargashampo - 0,2346 * hargasikatgigi + 0,228 * jualrumahmurah + 0,2188 * jualrumah - 0,231 * medicalcheckup + 0,1708 * paketinternet + 0,156 * rumahdisewakan - 0,2397 * zumba, (4)$$

$$PC2 = -0,1982 * biyakuliah + 0,2171 * biyalistrik + 0,3417 * biyapendidikan + 0,1103 * biyapuskesmas - 0,0341 * biayarumahsakit + 0,049 * biyasekolah + 0,113 * hargabensin + 0,289 * hargalpg + 0,0649 * hargapampers + 0,042 * hargapastagigi + 0,087 * hargasabunmandi - 0,0509 * hargashampo + 0,223 * hargasikatgigi + 0,2636 * jualrumahmurah + 0,382 * jualrumah + 0,1724 * medicalcheckup + 0,0526 * paketinternet + 0,5686 * rumahdisewakan + 0,217 * zumba, (5)$$

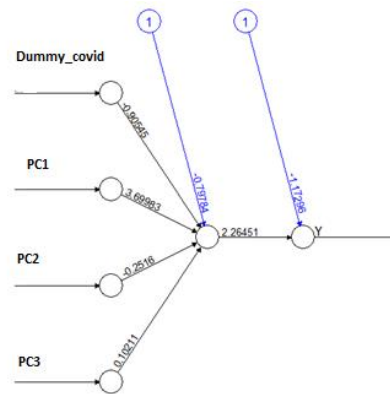
$$PC3 = 0,2856 * biyakuliah - 0,0059 * biyalistrik - 0,337 * biyapendidikan - 0,105 * biyapuskesmas + 0,2237 * biayarumahsakit + 0,0959 * biyasekolah + 0,3733 * hargabensin + 0,0255 * hargalpg - 0,0538 * hargapampers - 0,114 * hargapastagigi - 0,10 * hargasabunmandi + 0,0996 * hargashampo + 0,0553 * hargasikatgigi + 0,2449 * jualrumahmurah + 0,139 * jualrumah - 0,1395 * medicalcheckup + 0,6748 * paketinternet + 0,0347 * rumahdisewakan - 0,0059 * zumba, (6)$$

Model FFNN Terbaik

Ada 30 model FFNN yang dibentuk berdasarkan kombinasi jumlah *hidden layer* dan jumlah *neuron* di masing-masing *hidden layer*, digambarkan oleh Tabel 4. Pemilihan model terbaik didasarkan pada nilai RMSE pada data *testing*. Model terbaik adalah model FFNN dengan 1 *hidden layer* dan 1 *neuron* pada *hidden layer*nya. Model tersebut merupakan model yang paling sederhana diantara 30 model yang dibentuk. RMSE pada model tersebut sebesar 0,1279. Hal tersebut sejalan dengan penelitian yang dilakukan oleh Suhartono, et al., (2019) yang menyebutkan bahwa model yang lebih kompleks belum tentu lebih baik dibandingkan dengan model yang lebih sederhana. Arsitektur model terbaik disajikan pada Gambar 5.

Tabel 4. Model *Neural Network* dan RMSE-nya

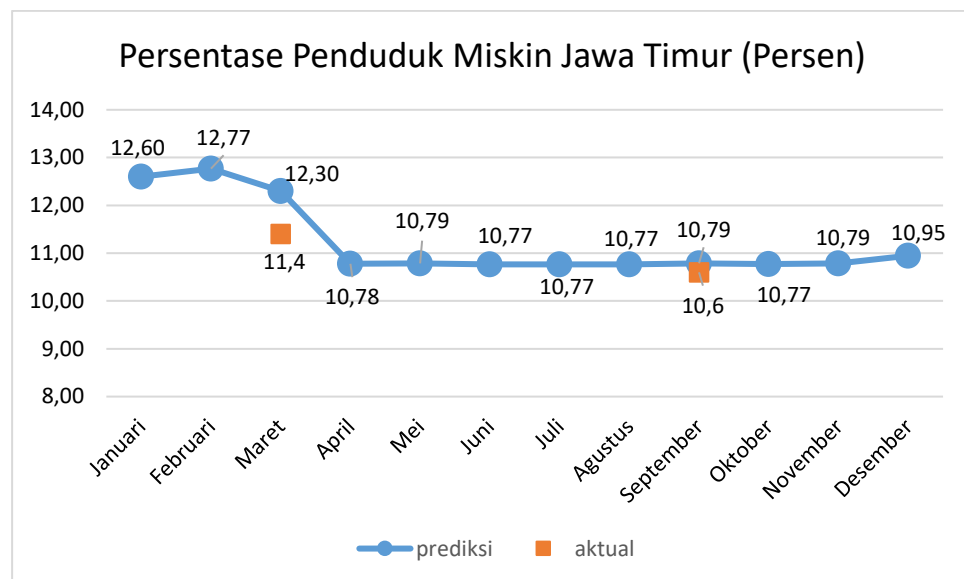
Jumlah <i>Hidden Layer</i>	Jumlah <i>Neuron</i>		RMSE <i>Testing</i>
	<i>Hidden Layer 1</i>	<i>Hidden Layer 2</i>	
1	1	-	0,1279
1	2	-	0,7266
1	3	-	1,1555
1	4	-	1,4387
1	5	-	1,2646
2	1	1	1,8409
2	1	2	0,6685
2	1	3	0,6636
2	1	4	1,8430
2	1	5	1,8530
2	2	1	1,1941
2	2	2	0,4888
2	2	3	0,9210
2	2	4	0,5173
2	2	5	1,1212
2	3	1	1,0990
2	3	2	0,3360
2	3	3	0,9754
2	3	4	0,4555
2	3	5	1,2821
2	4	1	1,1130
2	4	2	0,8734
2	4	3	1,0893
2	4	4	1,0537
2	4	5	0,7870
2	5	1	1,2086
2	5	2	0,2124
2	5	3	0,7016
2	5	4	0,9418
2	5	5	0,6448



Gambar 5. Arsitektur FFNN Terbaik untuk Persentase Penduduk Miskin di Jawa Timur

Prediksi Persentase Penduduk Miskin Bulanan

Sebelumnya telah disebutkan bahwa model FFNN terbaik adalah model yang paling sederhana, yaitu model dengan 1 *hidden layer* dan 1 *neuron*. Kemudian, model tersebut digunakan untuk melakukan prediksi persentase penduduk miskin bulanan dari bulan Januari hingga Desember 2021. Variabel prediktor yang digunakan adalah GTI berdasarkan 19 kata kunci pencarian. Masing-masing variabel dihitung rata-rata bergerak 12 bulan yang lalu agar konsisten dengan data *training* yang telah dibentuk. Selanjutnya dibentuk PC1, PC2, dan PC3 berdasarkan persamaan 4, 5, dan 6 Hasil prediksi tersebut diberikan pada Gambar 6.



Gambar 6. Prediksi Persentase Penduduk Miskin Jawa Timur Tahun 2021

Persentase penduduk miskin Jawa Timur periode Januari hingga April diprediksi berfluktuasi. Prediksi persentase penduduk miskin tertinggi berada pada bulan Februari, yaitu sekitar 12,77 persen. Bulan April hingga November, persentase penduduk miskin Jawa Timur relatif dalam situasi yang stagnan. Pada Bulan Desember 2021, persentase penduduk miskin Jawa Timur diprediksi merangkak naik menuju angka 10,95 persen. Hasil prediksi persentase penduduk miskin dari GTI juga dapat mendukung fenomena menurunnya *trend* angka kemiskinan bulan Maret hingga September 2021.

KESIMPULAN

GTI yang tersedia dalam *series* bulanan dapat digunakan untuk membangun model persentase penduduk miskin di Jawa Timur kemudian melakukan prediksi angka kemiskinan bulanan. Dengan adanya prediksi angka bulanan tersebut, gap informasi antara susenas dapat terisi. Model yang dibangun menggunakan metode FFNN menghasilkan model terbaik dengan RMSE data *testing* sebesar 0,1279. Model terbaik tersebut merupakan model yang paling sederhana diantara 30 model FFNN yang dibentuk, yaitu dengan menggunakan 1 *hidden layer* dan 1 *neuron*. Hal tersebut menunjukkan bahwa model yang lebih kompleks belum tentu lebih baik dari model yang lebih sederhana. Tentunya tidak ada satu metode yang paling baik untuk segala jenis

data. Kedepannya perlu adanya metode lain yang dapat dibandingkan dengan FFNN, seperti ARIMAX, *Time Series Regression*, dan *Support Vector Regression*.

DAFTAR PUSTAKA

- ADB. (2020). *Mapping Poverty Through Data Integration and Artificial Intelligence* (Issue September).
- Afkham, M., Ghoddusi, H., & Rafizadeh, N. (2021). Google search explains your gasoline consumption! *Energy Economics*, 99. <https://doi.org/10.1016/j.eneco.2021.105305>
- Ayuningtyas, I., & Wirawati, I. (2019). Nowcasting Tingkat Penghunian Kamar Hotel Menggunakan Google Trends (Nowcasting Occupancy Rate of Hotel Room With Google Trends). *Seminar Nasional Official Statistics 2019*, 338–343.
- Bhakti, H. D. (2019). Aplikasi Artificial Neural Network (ANN) untuk Memprediksi Masa Studi Mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Gresik. *Eksplora Informatika*, 9(1), 88–95. <https://doi.org/10.30864/eksplora.v9i1.234>
- BPS. (2021a). *Data dan Informasi Kemiskinan Kabupaten/Kota*.
- BPS. (2021b). *Indikator Makro Sosial Ekonomi Jawa Timur Triwulan II 2021*.
- BPS. (2021c). *Statistik Kesejahteraan Rakyat Provinsi Jawa Timur 2021*.
- Fasulo, A., Guandalini, A., & Terribili, M. D. (2017). Google Trends For Nowcasting Quarterly Household Consumption Expenditure. *Rivista Italiana Di Economia ...*, LXXI.
- Götz, T. B., & Knetsch, T. A. (2019). Google data in bridge equation models for German GDP. *International Journal of Forecasting*, 35(1), 45–66. <https://doi.org/10.1016/j.ijforecast.2018.08.001>
- Hediyati, D., & Suartana, I. M. (2021). Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro. *Journal of Information Engineering and Educational Technology*, 5(2), 49–54. <https://doi.org/10.26740/jieet.v5n2.p49-54>
- Hendro, G., Adj, T. B., & Setiawan, N. A. (2012). Penggunaan Metodologi Analisa Komponen Utama (PCA) untuk Mereduksi Faktor-Faktor yang Mempengaruhi Penyakit Jantung Koroner. *Seminar Nasional ScrETec*, 1–5.
- Jolliffe, I. T. (2010). Principal components analysis. *International Encyclopedia of Education*, 374–377. <https://doi.org/10.1016/B978-0-08-044894-7.01358-0>
- Perez, J. J., Gil, M., Sanchez-Fuentes, A. J., & Urtasun, A. (2017). Nowcasting Private Consumption : Traditional Indicators, Uncertainty Measures, And The Role of Internet Search Query Data. *6th World Statistics Congress*.
- Prathama, A. Y., Aminullah, A., & Saputra, A. (2018). Pendekatan Ann (Artificial Neural Network) Untuk Penentuan Prosentase Bobot Pekerjaan Dan Estimasi Nilai Pekerjaan Struktur Pada Rumah Sakit Pratama. *Jurnal Teknosains*, 7(1), 14. <https://doi.org/10.22146/teknosains.30139>
- Putri, J. A., Suhartono, S., Prabowo, H., Salehah, N. A., Prastyo, D. D., & Setiawan, S. (2021). Forecasting Currency in East Java: Classical Time Series vs. Machine Learning. *Indonesian Journal of Statistics and Its Applications*, 5(2), 284–303. <https://doi.org/10.29244/ijsa.v5i2p284-303>
- Sharma, S., Sharma, S., & Athaiya, A. (2020). Activation Functions in Neural Networks. *International Journal of Engineering Applied Sciences and Technology*, 04(12), 310–316. <https://doi.org/10.33564/ijeast.2020.v04i12.054>
- Suhartono, Ashari, D. E., Prastyo, D. D., Kuswanto, H., & Lee, M. H. (2019). Deep neural network for forecasting inflow and outflow in Indonesia. *Sains Malaysiana*, 48(8), 1787–1798. <https://doi.org/10.17576/jsm-2019-4808-26>
- Williams, J., & Li, Y. (2008). A case study using neural network algorithms: Horse racing predictions in Jamaica. *Proceedings of the 2008 International Conference on Artificial Intelligence, ICAI 2008 and Proceedings of the 2008 International Conference on Machine Learning; Models, Technologies and Applications, January*, 16–22.