

## Prediksi Tingkat Kemiskinan Provinsi Gorontalo Menggunakan Metode Gabungan K-Means dan Generalized Regression Neural Network

Saprina Mamase<sup>1)</sup>, Ruly S. Sinukun<sup>2)</sup>

<sup>1)</sup>Teknik Informatika, Politeknik Gorontalo

Jl. Muchlis Rahim, Panggulo Barat, Kec. Botupingge, Kab. Bone Bolango, Gorontalo

E-mail : rina@poligon.ac.id<sup>1)</sup>

### ABSTRAK

Penurunan tingkat kemiskinan menjadi hal penting dan menjadi fokus utama pemerintah Provinsi Gorontalo, karena merupakan salah satu daerah termiskin di Indonesia. Provinsi Gorontalo memiliki data persentasi penduduk miskin yang bersifat non linear. Salah satu aspek penting untuk mendukung strategi penanggulangan kemiskinan adalah tersedianya data kemiskinan yang akurat dan tepat sasaran. Sistem pengukuran yang diusulkan dalam penelitian ini menggunakan metode yang sangat cocok untuk data non linear. Generalized Regression Neural Network (GRNN) merupakan metode prediksi *probabilistic neural network* yang bersifat stabil dan cocok diterapkan pada data kemiskinan penduduk yang memiliki karakteristik yang beragam dan bersifat *non linear*. Peningkatan penggunaan jumlah data latih pada GRNN dapat mengakibatkan masalah *overfitting* sehingga berdampak pada penurunan performa hasil Prediksi. Dalam mengatasi hal tersebut, penelitian ini menggunakan metode K-Means dalam proses pemilihan data latih yang sesuai sehingga akan menghasilkan struktur jaringan yang lebih efisien atau model GRNN dengan kinerja performa yang bagus. Uji coba dilakukan dengan membandingkan model prediksi GRNN dengan model prediksi K-Means dan GRNN yang menggunakan data tingkat kemiskinan pada tahun 2016. Hasil uji coba menunjukkan bahwa besarnya nilai MAPE dari hasil uji coba prediksi tingkat kemiskinan pada tahun 2016 menghasilkan nilai MAPE sebesar 2.3%, sedangkan prediksi tingkat kemiskinan menggunakan metode GRNN saja memiliki nilai MAPE sebesar 2.9%.

**Kata Kunci** : prediksi, tingkat kemiskinan, GRNN, K-Means

### ABSTRACT

Decreasing the poverty rate is important and is the main focus of the Gorontalo provincial government, because it is one of the poorest areas in Indonesia. Gorontalo Province has a percentage of poor population data that is non-linear. One important aspect to support poverty reduction strategies is the availability of accurate and targeted poverty data. The measurement system proposed in this research uses a method that is very suitable for non-linear data. Generalized Regression Neural Network (GRNN) is a stable probabilistic neural network prediction method that is suitable to be applied to population poverty data that has diverse characteristics and is non-linear. Increasing the use of the amount of training data on GRNN can result in overfitting problems that have an impact on decreasing the performance of prediction results. In overcoming this, this research uses the K-Means method in the process of selecting appropriate training data so that it will produce a more efficient network structure or GRNN model with good performance. The result was conducted by comparing the GRNN prediction model with the K-Means model and GRNN which used poverty rate data in 2016. The results of the trial showed that the MAPE value from the results of the 2016 poverty rate prediction resulted in a MAPE value of 2.3%, while the prediction of poverty rate using the GRNN method alone has a MAPE value of 2.9%.

**Keywords** : prediction, poverty rate, GRNN, K-Means.

### PENDAHULUAN

Negara Indonesia merupakan negara berkembang yang terdiri dari 34 Provinsi dengan

tingkat kemiskinan masing-masing provinsi yang berbeda-beda. Provinsi Gorontalo merupakan provinsi yang ke 32 di Indonesia. Provinsi Gorontalo terdiri dari 6 kabupaten/kota dengan

dengan jumlah penduduk dan persentase tingkat kemiskinan yang berbeda-beda di setiap wilayah kabupaten/kota. Berdasarkan Survei Badan Pusat Statistik (BPS) Provinsi Gorontalo pada tahun 2009 hingga tahun 2016, menunjukkan naik turunnya persentase tingkat kemiskinan yang dinyatakan dengan penduduk miskin di (kabupaten/kota) yang ada provinsi tersebut. Data BPS sejak tahun 2009 hingga 2017 menunjukkan bahwa persentase penduduk miskin di kabupaten/kota Provinsi Gorontalo mengalami perubahan yang fluktuatif. Pencapaian tingkat kemiskinan ini menjadi hal penting dan menjadi fokus utama pemerintah, karena Provinsi Gorontalo merupakan salah satu daerah termiskin di Indonesia. Data yang fluktuatif ini cukup menyiratkan adanya upaya pemerintah Provinsi Gorontalo untuk menurunkan tingkat kemiskinan tersebut.

Dalam hal upaya mengatasi masalah kemiskinan, diperlukan data kemiskinan yang akurat sehingga dapat melaksanakan program bantuan sosial secara tepat sasaran dan merata. Oleh karena itu, masalah pengukuran tingkat kemiskinan ini telah mengundang banyak minat dari para peneliti untuk mengkaji pengukuran yang lebih akurat dan dapat diandalkan. Amina dan Irwan (2014) menggunakan Backpropagation Neural Network dengan metode PCA dan regresi berganda dalam memprediksikan jumlah penduduk miskin di Kalimantan Selatan. PCA dan regresi berganda digunakan untuk memilih indikator-indikator yang sangat berpengaruh dengan penentuan jumlah penduduk miskin[1].

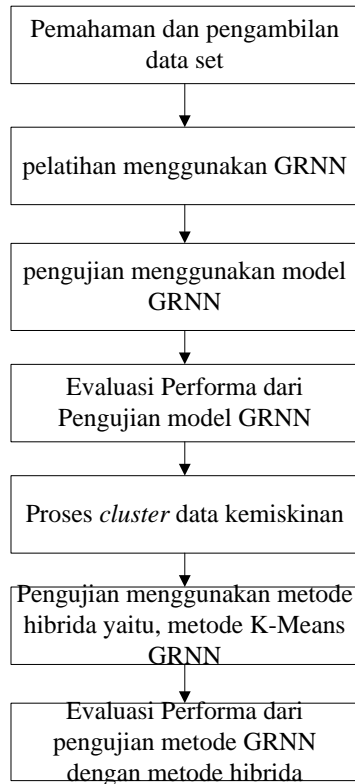
Prediksi kemiskinan menggunakan metode yang sesuai dengan karakteristik data merupakan salah satu upaya untuk dapat meningkatkan performa model prediksi yang digunakan. Prediksi tingkat kemiskinan menggunakan metode Generalized Regression Neural Network (GRNN) memberikan hasil prediksi tingkat kemiskinan yang cukup akurat yaitu dengan nilai MAPE sebesar 3 %[2]. Lain halnya dengan Nugroho dkk (2012) yang melakukan *clustering* kelompok swadaya masyarakat dengan menggunakan metode Self-Organizing Feature Map (SOM) yang diintegrasikan dengan metode K-Means dalam menentukan kebijakan bantuan Badan Pemberdayaan Masyarakat di Kota Surabaya. Terdapat 6 golongan *cluster* yang memiliki

karakteristik berbeda-beda yang diperoleh dari perhitungan SOM dan K-Means[3]. Selanjutnya, penelitian Iskandar dan Suprpto (2015) membandingkan Algoritma C4.5 dan Naive Bayes Classifier (NBC) untuk klasifikasi tingkat kemiskinan. Data yang digunakan adalah BDT (Basis Data Terpadu) yang dikeluarkan oleh TNP2K yang menggunakan 14 atribut. Hasil dari proses klasifikasi diperoleh bahwa metode C4.5 memiliki tingkat akurasi lebih baik jika dibandingkan dengan metode Naive Bayes[4].

Penelitian ini bertujuan untuk memprediksi tingkat kemiskinan di wilayah kabupaten/kota di Provinsi Gorontalo dengan menggunakan metode hibrida K-Means dan GRNN. Prediksi dilakukan menggunakan indikator kependudukan dan tenaga kerja, pendidikan, dan Indikator pada bidang pertanian, industri, dan perdagangan. Indikator-indikator yang digunakan dalam penelitian ini yang diperoleh dari kantor BPS kabupaten/kota yang ada di Provinsi Gorontalo. Indikator yang diperoleh tersebut kemudian diujicobakan pada metode hibrida K-Means dan GRNN. Metode ini sangat tepat digunakan pada data yang bersifat *non linear*. Sebagaimana penelitian [5] yang mengusulkan penggabungan metode *cluster* K-Means dan GRNN untuk memprediksi data yang bersifat *non linear*. Usulan tersebut diterapkan pada data arus lalu lintas. Metode *cluster* digunakan untuk menentukan data latih yang sesuai sehingga dapat mengatasi masalah peningkatan ukuran jaringan GRNN seiring dengan bertambahnya data latih. Penggabungan metode tersebut bertujuan untuk mendapatkan model prediksi dengan struktur jaringan yang lebih efisien, dan diperoleh hasil akurasi prediksi dengan performa yang lebih baik apabila dibandingkan dengan prediksi hanya menggunakan metode GRNN. Berangkat dari hal tersebut, penggunaan metode ini sangat sesuai digunakan untuk memprediksi tingkat kemiskinan yang bersifat *non linear* (berdasarkan fluktuatif data). Sistem prediksi dengan hasil akurasi yang lebih baik, dapat dijadikan sebagai solusi yang tepat dalam penanganan masalah kemiskinan khususnya dalam melaksanakan program bantuan sosial secara merata dan tepat sasaran di Provinsi Gorontalo.

## METODE

Tahapan pada penelitian ini dapat dilihat pada Gambar 1. Proses pada setiap tahapan akan dijelaskan pada sub bab berikut ini.



Gambar 1 Metodologi Penelitian

**a. Pemahaman dan Pengambilan data set**

Data set yang digunakan adalah data-data yang menjadi indikator tingkat kemiskinan penduduk (kabupaten/kota) yang digunakan oleh BPS Provinsi Gorontalo, yang dalam penelitian ini disebut dengan indikator-indikator yang mempengaruhi tingkat kemiskinan. Data yang digunakan adalah data tahunan perwilayah (kabupaten/kota) selama 8 tahun yaitu mulai dari tahun 2009-2016. Data tersebut terdiri dari beberapa indikator, yaitu kependudukan dan tenaga kerja, pendidikan, dan Indikator pada bidang pertanian, industri, dan perdagangan.

**b. Pelatihan Menggunakan GRNN**

Metode yang digunakan untuk melakukan prediksi tingkat kemiskinan adalah GRNN. Proses pelatihan GRNN menggunakan data-data indikator kemiskinan dari tahun 2009 hingga tahun 2015 untuk prediksi tingkat kemiskinan pada tahun 2016. Proses pelatihan

dilakukan untuk mendapatkan Model GRNN terbaik. Data latih akan dibagi menjadi data pembelajaran dan data validasi menggunakan *Leave One Out*. *Leave One Out* merupakan bentuk khusus dari metode *K-Fold Cross Validation*, dimana K diset menjadi k=N, dengan N adalah jumlah data dalam set data. Selanjutnya dilakukan proses perhitungan *Gaussian* dengan menggunakan *smoothing factor* yang bernilai [0.05, 1] dengan setiap kenaikan 0.05. Pemilihan parameter *smoothing factor* yang terbaik berdasarkan Mean Absolute Percentage Error (MAPE) yang paling minimal. MAPE dinyatakan dengan Persamaan (1).

$$MAPE = \frac{1}{N} \left( \sum_{i=1}^N \frac{|Y_i - Y'_i|}{Y_i} \right) \tag{1}$$

N adalah banyaknya data validasi yang diprediksi,  $Y_i$  adalah output nilai prediksi dan  $Y'_i$  adalah output nilai aktual

**c. Pengujian Menggunakan Model GRNN**

Model GRNN dengan parameter yang terbaik digunakan sebagai proses pengujian. Pada proses ini dilakukan perhitungan *Gaussian* antara data uji dan data latih menggunakan Persamaan (2), Persamaan (3), Persamaan (4) dan Persamaan (5). Pada Persamaan (5) akan diperoleh hasil prediksi tingkat kemiskinan suatu wilayah (kabupaten/kota) untuk tahun berikutnya. Dari hasil proses pengujian ini akan dihitung akurasi dan kesalahannya menggunakan MAPE yang secara matematis dinyatakan pada Persamaan (1).

$$p_i = \exp \left[ \frac{-(X-X_i)^T(X-X_i)}{2\sigma_i^2} \right], i = 1, 2, \dots, n \tag{2}$$

X adalah vektor input dari variabel prediktor untuk GRNN,  $X_i$  adalah vektor pelatihan yang diwakili neuron pola ke-i,  $\sigma_i$  adalah parameter *smoothing factor* yang merepresentasikan lebar atau rentang fungsi. Jumlah neuron pada *paternn layer* sama dengan jumlah data latih yang digunakan.

$$S_N = \sum_{i=1}^n y_i \exp \left[ \frac{-(X-X_i)^T(X-X_i)}{2\sigma_i^2} \right] \tag{3}$$

$$S_D = \sum_{i=1}^n \exp \left[ \frac{-(x-x_i)^T(x-x_i)}{2\sigma_i^2} \right] \quad (4)$$

$y_i$  adalah bobot penghubung antara neuron ke- $i$  pada *patern layer* dan *summation neurons* pada penjumlahan numerator.

$$Y'_i = \frac{\sum_{i=1}^n y_i \exp \left[ \frac{-(x-x_i)^T(x-x_i)}{2\sigma_i^2} \right]}{\sum_{i=1}^n \exp \left[ \frac{-(x-x_i)^T(x-x_i)}{2\sigma_i^2} \right]} \quad (5)$$

$Y'_i$  adalah output prediksi sampel pembelajaran ke- $i$ .

#### d. Evaluasi Performa dari Pengujian Model GRNN

Untuk mengevaluasi kinerja model prediksi presentase tingkat kemiskinan pada penelitian ini digunakan suatu metode evaluasi yaitu Mean Absolute Percentage Error (MAPE). MAPE digunakan untuk mengukur seberapa selisih antara persentase tingkat kemiskinan yang diprediksi dengan persentase tingkat kemiskinan yang sebenarnya. Semakin kecil nilainya, berarti selisihnya semakin kecil juga. Perhitungan MAPE dinyatakan pada Persamaan (1).

#### e. Proses cluster data kemiskinan

Data yang di-*cluster* adalah data set yang telah dipisahkan sebagai data latih sedangkan data uji tidak perlu dilakukan proses *cluster*. Data latih di-*cluster* menggunakan metode K-means. Proses *cluster* menggunakan variabel-variabel yang mempengaruhi data kemiskinan yaitu konsumsi dan pengeluaran, kependudukan dan tenaga kerja, pendidikan, kesehatan, pertanian, industri, dan perdagangan.

#### f. Pengujian menggunakan metode Hibrida yaitu, Metode K-Means GRNN

Model GRNN dengan parameter yang terbaik digunakan sebagai proses pengujian. Tahap awal yang akan dilakukan sebagai persiapan proses pengujian adalah menentukan model prediksi yang akan digunakan oleh data uji. Pada tahapan awal ini, setiap nilai-nilai variabel prediktor pada data uji akan memilih sebuah model GRNN yang sesuai dengan karakteristiknya dengan cara melakukan perhitungan jarak *Euclidean* antara nilai-nilai

variabel prediktor dengan nilai *centroid* dari masing-masing data latih pada suatu model GRNN yang telah terbentuk. Dari hasil perhitungan tersebut, pilih suatu nilai *centroid* dari kelompok data latih pada suatu model GRNN yang terdekat dengan nilai-nilai variabel prediktor tersebut. Jadi model GRNN yang dipilih adalah model GRNN yang data latihnya memiliki nilai *centroid* yang dekat dengan nilai-nilai variabel prediktor tersebut.

Proses berikutnya adalah pengujian pada model GRNN yang telah dipilih. Pada proses ini dilakukan perhitungan *Gaussian* antara data uji dan data latih yang terpilih menggunakan Persamaan (2), Persamaan (3), Persamaan (4) dan Persamaan (5). Pada Persamaan (5) akan diperoleh hasil prediksi tingkat kemiskinan suatu wilayah (kabupaten/kota) untuk tahun berikutnya. Dari hasil proses pengujian ini akan dihitung akurasi dan kesalahannya menggunakan MAPE yang secara matematis dinyatakan pada Persamaan (1).

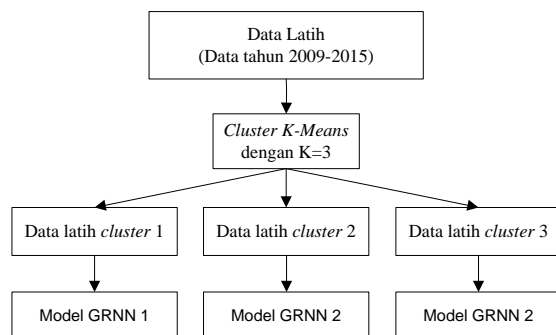
#### g. Evaluasi Performa dari Pengujian Metode GRNN dengan Metode Hibrida

Evaluasi kinerja ini dilakukan Membandingkan Performa dari model prediksi tingkat kemiskinan menggunakan metode GRNN dengan model prediksi menggunakan metode K-Means GRNN

## HASIL DAN PEMBAHASAN

Uji coba ini dilakukan untuk mengetahui performa model prediksi tingkat kemiskinan menggunakan metode hibrida yang diusulkan yaitu K-Means-GRNN dengan data uji pada tahun 2016. Data dibagi menjadi data latih dan data uji. Data tahun 2009-2015 digunakan sebagai data latih. Data latih tersebut telah dilakukan proses cluster menggunakan K-Means yang menghasilkan kelompok-kelompok data kemiskinan sebanyak jumlah *cluster* yang diberikan. Uji coba running K-means dilakukan sebanyak 5 kali dan memilih hasil *cluster* yang memiliki total jarak kuadrat di antara setiap titik data dengan representasi *cluster* terdekat yang nilainya paling kecil. Kelompok-kelompok data kemiskinan ini masing-masing akan membentuk suatu model prediksi dengan metode GRNN. Berdasarkan hal tersebut, banyaknya model prediksi yang dihasilkan pada skenario ini adalah sebanyak jumlah *cluster* pada data latih. Pada uji

coba ini, *cluster* yang dilakukan adalah sebanyak  $K=3$  sehingga model GRNN yang terbentuk juga sebanyak 3 buah yang dapat dilihat pada Gambar 2. Tiap model prediksi yang terbentuk masing-masing memiliki estimasi parameter *smoothing factor* terbaik. Sedangkan pada proses pengujian, setiap variabel data uji (nilai-nilai prediktor atau satu set atribut prediktor) hanya menggunakan sebuah model GRNN untuk proses pengujian. Pemilihan model GRNN yang akan digunakan pada proses pengujian yaitu dengan cara menghitung jarak *Euclidean* antara nilai-nilai prediktor dengan nilai *centroid* dari masing-masing data latih pada suatu model GRNN yang telah terbentuk. Dari hasil perhitungan tersebut, pilih suatu nilai *centroid* dari kelompok data latih pada suatu model GRNN yang terdekat dengan nilai-nilai prediktor tersebut. Jadi model GRNN yang dipilih adalah model GRNN yang data latihnya memiliki nilai *centroid* yang dekat dengan satu set atribut prediktor.



Gambar 2. Blok Diagram Pembentukan Model K-Means GRNN dengan  $K=3$  Untuk Pengujian Data Tahun 2016

Selanjutnya, hasil uji coba penelitian ini ditampilkan pada Tabel 1 yang merupakan perbandingan hasil prediksi tingkat kemiskinan menggunakan metode gabungan K-Means-GRNN dengan metode GRNN saja. Hasil prediksi tingkat kemiskinan pada tahun 2016 menggunakan metode K-Means-GRNN menghasilkan nilai MAPE sebesar 2.3%. Sedangkan hasil prediksi menggunakan metode GRNN pada data tahun 2016 menghasilkan MAPE dengan nilai 2.9%.

Tabel 1. Perbandingan Hasil Prediksi Tingkat Kemiskinan Menggunakan Metode Gabungan dengan Metode GRNN

| Kab/Kota               | Data Aktual (%) | Data Prediksi (%) |        |
|------------------------|-----------------|-------------------|--------|
|                        |                 | Kmeans GRNN       | GRNN   |
| <b>Bone Bolango</b>    | 17.970          | 18.150            | 18.492 |
| <b>Kab Gorontalo</b>   | 21.031          | 21.687            | 21.786 |
| <b>Pohuwato</b>        | 21.175          | 22.167            | 22.430 |
| <b>Gorontalo utara</b> | 18.514          | 18.877            | 18.927 |
| <b>Kota Gorontalo</b>  | 6.049           | 6.006             | 6.051  |
| <b>Boalemo</b>         | 21.110          | 21.650            | 21.665 |
| <b>MAPE (%)</b>        |                 | 2.3               | 2.9    |

## SIMPULAN

Berdasarkan hasil uji coba yang telah dilakukan, maka dapat disimpulkan bahwa prediksi tingkat kemiskinan menggunakan metode gabungan K-Means GRNN memberikan performa yang lebih baik dari pada dengan menggunakan metode GRNN saja. Hal tersebut dapat ditunjukkan dengan hasil prediksi kemiskinan pada data tahun 2016 menggunakan metode Gabungan dan GRNN saja masing-masing memiliki nilai MAPE sebesar 2.3 % dan 2.9 %.

## DAFTAR PUSTAKA

- [1] F. Amina and M. I. Irawan, "Prediksi Jumlah Penduduk Miskin di Provinsi Kalimantan Selatan Menggunakan Jaringan Syaraf Tiruan Backpropagation," *Pasca Sarj. Jur. Mat. FMIPA, Inst. Teknol. Sepuluh Nopember.*, 2014.
- [2] S. Mamase and R. S. Sinukun, "Prediksi Tingkat Kemiskinan Provinsi Gorontalo Dengan Metode GRNN," in *Seminar Nasional Humaniora & Aplikasi Teknologi Informasi 2018*, 2018, no. Sehati.
- [3] C. A. Nugroho, R. A. Hendrawan, and I. Hafidz, "Clustering Kelompok Swadaya Masyarakat ( KSM ) dalam Menentukan Kebijakan Bantuan Badan Pemberdayaan Masyarakat di Kota Surabaya dengan Menggunakan Metode Self-Organizing

- Map (SOM) dan K-Means,” *J. Teknk ITS*, vol. 1, no. 1, pp. 368–373, 2012.
- [4] D. Iskandar and Y. K. Suprpto, “Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan Antara Algoritma C.45 dan Naive Bayes,” *J. Ilm. NERO*, vol. 2, no. 1, pp. 37–43, 2015.
- [5] S. Mamase and J. L. Buliali, “Metode Hibrida K-Means dan Generalized Regression Neural Network Untuk Prediksi Arus Lalu Lintas,” *J. Buana Inform.*, vol. 7, no. 3, pp. 159–168, 2016.