

Metode Hibrida K-Means dan Generalized Regression Neural Network Untuk Prediksi Arus Lalu Lintas

Saprina Mamase¹, Joko Lianto Buliali²

Program Studi Pasca Sarjana, Jurusan Teknik Informatika, Institut Teknologi Sepuluh Nopember
Jl. Raya ITS, Kampus ITS, Sukolilo, Surabaya 60111
Email: ¹saprina14@mhs.if.its.ac.id, ²joko@cs.its.ac.id

Masuk: 14 Desember 2015; Direvisi: 28 Desember 2015; Diterima: 4 Januari 2016

Abstract. *Traffic flow forecasting is a popular research topic in the development of Intelligent Transportation System. There have been many forecasting methods used for traffic flow forecasting, such as Generalized Regression Neural Network (GRNN) which has a fairly good accuracy. One of the GRNN's characteristics is that the number of neurons in pattern layer increases as the number of training samples raise and this can cause overfitting problem. In this research, a hybrid method to predict traffic flow is proposed, that is K-means and GRNN algorithm. K-means method aims to solve overfitting problem in GRNN model by choosing training samples based on their similar characteristics. Leave One Out Cross Validation (LOOCV) is used to select an appropriate smoothing factor parameter at each GRNN's model. Mean Absolute Percentage Error (MAPE) is used as the evaluation criterion in the testing process. The results show that the proposed method could improve the accuracy of predictions by reducing the value of MAPE by 0.82-3.81%.*

Keywords: *Traffic flow forecasting, K-means, Generalized Regression Neural Network, Leave One Out Cross Validation*

Abstrak. *Prediksi arus lalu lintas telah menjadi tren topik penelitian untuk pengembangan sistem transportasi cerdas. Telah banyak metode yang digunakan terkait prediksi arus lalu lintas, diantaranya yaitu Generalized Regression Neural Network (GRNN) yang memiliki akurasi yang cukup baik. Salah satu karakteristik GRNN adalah jumlah neuron pada pattern layer akan bertambah seiring meningkatnya jumlah data latih yang akan mengakibatkan masalah overfitting. Dalam penelitian ini diusulkan metode hibrida K-means dan GRNN untuk prediksi arus lalu lintas. Metode K-means bertujuan untuk mengatasi masalah overfitting pada model GRNN dengan memilih data latih berdasarkan kemiripan karakteristiknya. Algoritma Leave One Out Cross Validation (LOOCV) digunakan untuk memilih parameter smoothing factor terbaik pada setiap model GRNN. Mean Absolute Percentage Error (MAPE) digunakan sebagai kriteria evaluasi model prediksi. Hasil menunjukkan bahwa metode yang diusulkan dapat meningkatkan akurasi prediksi dengan penurunan nilai MAPE sebesar 0,82-3,81%.*

Kata Kunci: *Prediksi arus lalu lintas, K-means, Generalized Regression Neural Network, Leave One Out Cross Validation*

1. Pendahuluan

Data arus lalu lintas (*traffic flow*) menjadi faktor utama untuk menganalisis kondisi lalu lintas jalan. Arus lalu lintas jangka pendek memiliki karakteristik perubahan secara dinamis, dan bersifat *non linear*. Prediksi arus lalu lintas jangka pendek secara akurat dan *real-time* akan memberikan informasi arus lalu lintas yang sesuai kepada pengguna jalan dan efektif memecahkan kemacetan lalu lintas (Kuang, dkk., 2010). Berdasarkan hal tersebut, prediksi arus lalu lintas memiliki pengaruh yang besar pada efisiensi pelaksanaan sistem lalu lintas.

Prediksi arus lalu lintas melalui pendekatan statistika dan komputasi telah banyak diusulkan. Vlahogianni, dkk. (2006) mengusulkan suatu metode statistika untuk mendeteksi karakteristik arus lalu lintas jangka pendek yang bersifat *non linear* dan *non stasioner*. Han, dkk. (2004) mengusulkan metode ARIMA adaptif, Enrique (2008) mengusulkan model *Bayesian* untuk memprediksi arus lalu lintas. Leng, dkk. (2013) membandingkan metode GRNN, *Radial Basis Function Neural Network* (RBFNN) dan *Back Propagation Neural Network* (BPNN) untuk memprediksi arus lalu lintas. Dari ketiga metode tersebut, metode GRNN yang memiliki hasil akurasi yang lebih tinggi.

GRNN diusulkan oleh Specht (1991) dan masuk dalam kategori *probabilistic neural network*. GRNN merupakan model sederhana ANN yang bersifat stabil dan cocok diterapkan pada data arus lalu lintas yang memiliki karakteristik perubahan secara dinamis dan bersifat *non linear* (Kuang, dkk. 2010). Salah satu karakteristik GRNN adalah jumlah *neuron* pada *pattern layer* akan bertambah seiring meningkatnya jumlah data latih (Specht, 1991). Akan tetapi, penggunaan jumlah *neuron* yang terlalu banyak akan mengakibatkan masalah *overfitting* (Jeong, dkk., 2012). Untuk mengatasi masalah tersebut, diperlukan suatu proses reduksi atau pemilihan data latih yang sesuai yang juga akan menghasilkan struktur jaringan yang lebih efisien atau model GRNN dengan kinerja performa yang bagus.

Penelitian ini mengusulkan penggabungan metode *K-means* dan GRNN. Metode *K-means* pada penelitian ini bertujuan untuk mendapatkan berbagai kelompok data latih yang dikelompokkan berdasarkan karakteristik arus lalu lintas yang mirip pada waktu-waktu tertentu. Berdasarkan hal tersebut, pemilihan data latih pada model GRNN dapat dilakukan dengan memilih data latih yang berada pada kelompok arus lalu lintas yang sama. Penggabungan kedua metode tersebut diharapkan dapat meningkatkan kinerja performa model GRNN dan memberikan hasil akurasi prediksi yang lebih baik apabila dibandingkan dengan hasil prediksi yang hanya menggunakan metode prediksi GRNN.

2. Pendekatan dan Tinjauan Pustaka

2.1. Prediksi Arus Lalu Lintas

Arus lalu lintas adalah jumlah kendaraan yang melewati suatu titik pada penggalan jalan tertentu dalam periode tertentu, diukur dengan satuan kendaraan persatuan waktu tertentu (Julianto, 2010). Prediksi arus lalu lintas merupakan bagian penting dari perencanaan transportasi, kontrol lalu lintas, dan sistem cerdas transportasi (Tan, dkk., 2007). Berbagai metode dan model telah diterapkan dalam konteks prediksi arus lalu lintas jangka pendek yang berdasarkan pada jenis data yang tersedia, diantaranya adalah analisis *time series* (Williams, 2001), model Bayesian (Sun, dkk., 2006), metode *Neural Network* (Dia, 2001), dan metode statistik (Vlahogianni, dkk., 2006). Hampir secara universal pada literatur prediksi tidak ada metode tunggal yang terbaik dalam setiap situasi (Tan, dkk., 2009). Banyak peneliti telah menunjukkan bahwa menggabungkan metode prediksi dengan beberapa model menghasilkan akurasi prediksi yang lebih tinggi dibandingkan dengan model tunggal (Makridakis, 1989). Penggunaan model hibrida telah menjadi suatu pekerjaan umum untuk meningkatkan akurasi prediksi dan juga menjadi kombinasi dari beberapa model yang digunakan dalam prediksi arus lalu lintas (Zhang, 2003).

2.2. Analisis Cluster K-Means

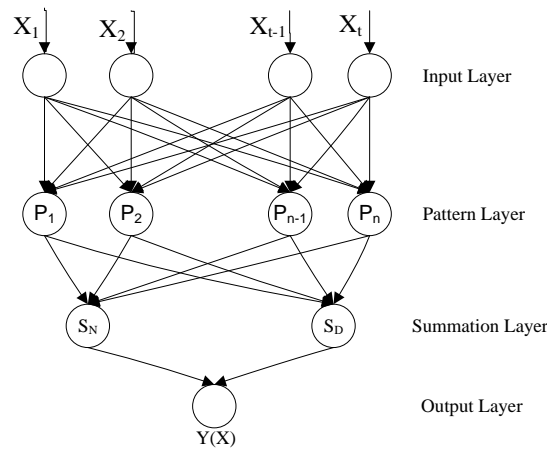
Analisis *cluster K-Means* merupakan salah satu metode *cluster analysis* non hirarki yang melakukan partisi data yang ada ke dalam K jumlah *cluster* atau kelompok. Data tersebut dikelompokkan berdasarkan kemiripan karakteristiknya, sehingga data yang dikelompokkan dalam satu *cluster* yang sama mempunyai karakteristik yang mirip. Tujuan pengelompokan adalah untuk meminimalkan *objective function* yang di set dalam proses *clustering*, yang pada dasarnya berusaha untuk meminimalkan variasi dalam satu *cluster* dan memaksimalkan variasi antar *cluster*. Metrik ukuran kemiripan yang digunakan pada penelitian ini adalah *Euclidean*.

Penggunaan metode *cluster* untuk menganalisis data *time series*, seperti data arus lalu lintas tidak hanya dapat menemukan beberapa pola tertentu dari arus lalu lintas tetapi juga dapat

mengelompokkan penggalan atau bagian dari jalan raya berdasarkan perbedaan karakteristik arus lalu lintas (Yin, dkk., 2006). Model arus lalu lintas di waktu yang berbeda memiliki pola yang berbeda (Zhanquan, dkk., 2008), sehingga peningkatan akurasi prediksi dapat dilakukan dengan membangun berbagai model prediksi pada waktu yang berbeda juga.

2.3. Generalized Regression Neural Network

Prediksi arus lalu lintas menggunakan metode GRNN memiliki hasil akurasi yang lebih baik dari pada menggunakan metode RBFNN dan BPNN (Leng, dkk., 2013). Berbeda dengan BPNN, GRNN tidak memerlukan prosedur pelatihan iteratif atau penyesuaian bobot koneksi antara *neuron* untuk menjaga stabilitas jaringan, sehingga dapat meningkatkan kecepatan konvergensi dan waktu pelatihan. Selain itu hanya terdapat satu parameter yang perlu disesuaikan, yang dikenal sebagai *smoothing factor*. Penyesuaian parameter ini untuk meningkatkan kinerja jaringan, agar memperoleh hasil akurasi prediksi yang lebih baik (Wang & Lin, 2010). Topologi GRNN seperti yang dijelaskan pada Gambar 1, pada prinsipnya GRNN terdiri dari empat *layer* unit pemrosesan (*neuron*) yaitu *input layer*, *pattern layer* yang memiliki jumlah *neuron* yang sama dengan jumlah data latih, *summation layer* dan *output layer*.



Gambar 1. Arsitektur GRNN

Secara umum model GRNN menggunakan algoritma sebagai berikut (Specht, 1991): (1) *Neuron-neuron* pada *input layer* berfungsi untuk mengambil informasi. *Neuron input* ini bersifat unik untuk tiap variabel prediktor pada vektor *input* X. (2) Menghitung nilai aktivasi setiap unit *pattern neuron* menggunakan fungsi Gaussian seperti pada persamaan (1), dimana X adalah vektor *input* dari variabel prediktor untuk GRNN, X_i adalah vektor pelatihan yang diwakili *neuron* pola ke-i, σ_i adalah parameter *smoothing factor* yang merepresentasikan lebar atau rentang fungsi. (3) Menghitung semua sinyal masukan pada unit *summation*. Secara teknis ada dua tipe penjumlahan yang dibentuk dalam *summation neurons* seperti pada persamaan (2) dan persamaan (3) yaitu S_N adalah penjumlahan terboboti dan S_D merupakan penjumlahan aritmatik sederhana, dimana y_i adalah bobot penghubung antara *neuron* ke-i pada *paterrn layer* dan *summation neurons* pada penjumlahan numerator. (4) Menghitung semua sinyal masukan pada unit *output*. *Neuron output* kemudian membentuk pembagian seperti pada persamaan (4) untuk mendapatkan *output* regresi GRNN, dimana Y_i' adalah *output* prediksi sampel pembelajaran ke-i.

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

$$S_N = \sum_{i=1}^n y_i \exp \left[\frac{-(\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i)}{2\sigma_i^2} \right] \quad (2)$$

$$S_D = \sum_{i=1}^n \exp \left[\frac{-(\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i)}{2\sigma_i^2} \right] \quad (3)$$

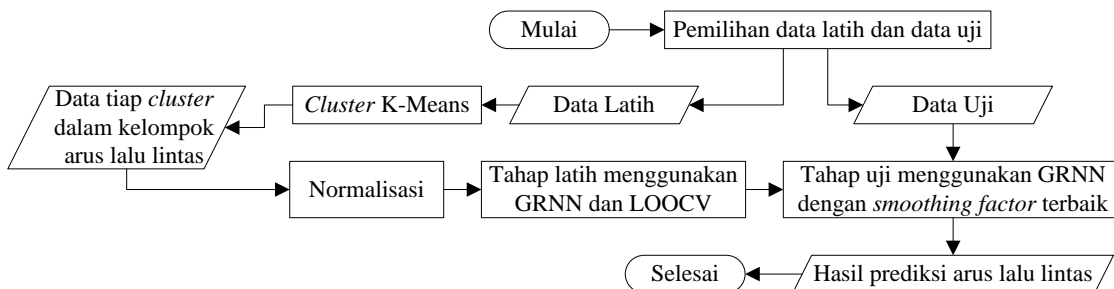
$$Y_i' = \frac{S_N}{S_D} \quad (4)$$

2.4. Leave One Out Cross Validation

Penggunaan algoritma *cross validation* untuk menentukan *smoothing factor* yang sesuai pada GRNN dilakukan untuk menghindari terjadinya *overfitting* (Leng, dkk., 2013). Bentuk umum algoritma ini disebut dengan *k-fold-cross validation*, yang memecahkan set data menjadi *k* bagian set data dengan ukuran yang sama. Setiap kali berjalan, satu pecahan berperan sebagai set data validasi sedangkan pecahan lainnya menjadi set data latih. Prosedur tersebut dilakukan sebanyak *k* kali sehingga setiap data berkesempatan menjadi data validasi tepat satu kali dan menjadi data latih sebanyak *k-1* kali. Bentuk khusus dari metode ini adalah ketika *k* di set $k=N$, jumlah data dalam set data. Metode ini disebut dengan *leave-one-out*, yaitu set data uji hanya berisi satu data saja, sedangkan proses pelatihan (dan validasi) dilakukan sebanyak *N* kali.

3. Metodologi

Metodologi penelitian ini dapat dilihat pada Gambar 2. Proses pada setiap tahapan akan dijelaskan pada sub bab berikut ini.



Gambar 2. Metodologi Penelitian

3.1. Pembagian Data Set

Dataset yang digunakan dalam penelitian ini adalah data arus lalu lintas selama satu bulan. Pembagian dataset menjadi data latih dan data uji dilakukan pada tahap awal pada metodologi penelitian ini. Data di minggu pertama, kedua dan ketiga dijadikan sebagai data latih sedangkan data pada minggu keempat dijadikan sebagai data uji. Pada penelitian ini hanya menggunakan data arus lalu lintas di hari kerja (Senin-Jumat).

3.2. Proses Pelatihan Menggunakan Metode K-Means dan GRNN

Data yang di-*cluster* adalah data latih. Data latih selama tiga minggu di-*cluster* menggunakan metode *K-means*. Proses *cluster* tersebut akan menghasilkan berbagai kelompok data arus lalu lintas berdasarkan kesamaan karakteristiknya. Setiap kelompok data latih akan membentuk sebuah model GRNN sebagai model prediksi. Hasil akhir dari proses pelatihan ini akan menghasilkan model GRNN sebanyak *K cluster* yang dilakukan yang masing-masing memiliki estimasi parameter terbaik. Sebagai contoh, jika terdapat tiga kelompok data arus lalu lintas dari hasil *cluster*, maka akan ada tiga model GRNN yang terbentuk.

Penggunaan fitur data pada proses *cluster* direpresentasikan dengan penggunaan variabel rata-rata arus lalu lintas setiap 15 menit pada periode waktu sebelumnya yang akan digunakan

pada model prediksi. Sebagai contoh, ada tiga variabel rata-rata arus lalu lintas yang disebut dengan nilai-nilai variabel prediktor yang akan digunakan pada proses *cluster*, yaitu x_{t-2} , x_{t-1} , x_t . Variabel tersebut merupakan data-data arus lalu lintas pada waktu t-2, t-1 dan t juga digunakan sebagai parameter *neuron* pada *input layer* GRNN, untuk memprediksi arus lalu lintas pada waktu t+1. Tabel 1 merupakan contoh hasil *cluster* K-means dengan K=3 yang menggunakan tiga variabel data arus lalu lintas.

Tabel 1. Contoh Proses Cluster Data Arus

Variabel Rata-Rata Arus Lalu Lintas (unit/15 menit)			Waktu Prediksi (t+1)	Kelompok Data
X_{t-2}	X_{t-1}	X_t	(Jam:Menit)	Arus Lalu Lintas
45.5	43	57	5:15	2
43	57	80	5:30	2
57	80	117	5:45	2
80	117	112	6:00	2
117	112	139.5	6:30	3
112	139.5	154.5	6:45	3
139.5	154.5	217.5	7:00	3
154.5	217.5	238	7:15	3
217.5	238	253.5	7:30	1
238	253.5	276.5	7:45	1
253.5	276.5	315.5	8:00	1

Cluster data arus lalu lintas mendeskripsikan pembentukan kelompok-kelompok data arus lalu lintas. Pada Tabel 1 terdapat tiga kelompok data arus lalu lintas. Tujuan dari proses *cluster* pada penelitian ini adalah untuk mendapatkan berbagai kelompok data arus lalu lintas yang telah dikelompokkan berdasarkan kemiripan karakteristik arus lalu lintasnya. Selanjutnya, data latih pada setiap kelompok data tersebut akan dilakukan proses normalisasi yang akan menghasilkan data yang mempunyai sifat *zero-mean* dan *unit-variance* sehingga semua variabel *input* berada pada kisaran rentang dan variasi yang sama. Hal tersebut dilakukan karena sesuai dengan penggunaan fungsi *Gaussian* sebagai fungsi aktivasi pada *pattern layer* dengan *kernel* yang memiliki rentang yang sama di masing-masing dimensi. Rumus normalisasi yang digunakan dinyatakan dalam persamaan (5), dimana $mean(x_k)$ dan $stdev(x_k)$ adalah nilai rata-rata dan standar deviasi dari sampel data, x_{ik} adalah nilai normalisasi.

Proses pelatihan bertujuan untuk mendapatkan model prediksi dengan estimasi parameter terbaik. Data latih yang digunakan untuk pembentukan model GRNN adalah data yang sesuai dengan kelompok data arus lalu lintas yang diperoleh dari hasil *cluster*. Nilai *centroid* yang diperoleh dari hasil *cluster* tidak hanya digunakan sebagai kriteria pengelompokkan data akan tetapi digunakan juga sebagai kriteria dalam pemilihan model GRNN yang akan digunakan pada proses pengujian. Penjelasan tentang pemilihan model GRNN mana yang akan digunakan pada proses pengujian dipaparkan pada sub bab 3.3.

Pada GRNN terdapat sebuah parameter yang disebut *smoothing factor*. Pemilihan parameter *smoothing factor* yang terbaik dilakukan menggunakan algoritma LOOCV dengan membagi data latih menjadi data pembelajaran dan data validasi. Pemilihan parameter *smoothing factor* yang terbaik berdasarkan *Mean Absolute Percentage Error* (MAPE) yang paling minimal. MAPE dinyatakan dalam persamaan (6), dimana N adalah banyaknya data validasi yang diprediksi, Y_i adalah *output* nilai prediksi dan Y_i' adalah *output* nilai aktual.

$$x_{ik} = \frac{x_{ik} - mean(x_k)}{stdev(x_k)} \tag{5}$$

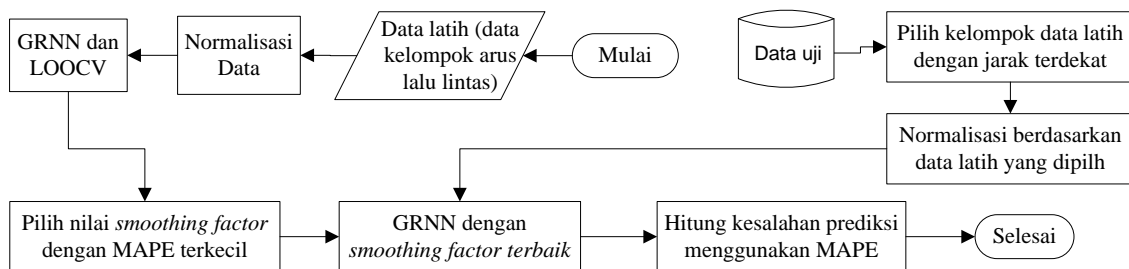
$$MAPE = \frac{1}{N} \left(\sum \frac{|Y_i - Y_i'|}{Y_i} \right) \tag{6}$$

Selain *parameter smoothing factor*, parameter *input* GRNN juga dapat mempengaruhi performa model (Specht, 1991). Parameter jumlah *neuron* pada *input layer* GRNN ini

direpresentasikan sebagai variable arus lalu lintas pada waktu-waktu sebelumnya secara *time series* yang digunakan untuk memprediksi arus lalu lintas pada waktu berikutnya. Sebagai contoh, untuk memprediksi arus lalu lintas pada periode $(t+1)$, maka parameter jumlah *neuron* pada *input layer* yang digunakan adalah arus lalu lintas pada periode $(t-m+1)$, $(t-m+2)$, ..., (t) .

3.3. Proses Pengujian Menggunakan Metode GRNN

Model GRNN dengan parameter *smoothing factor* 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 sub bab 3.1 telah dijelaskan bahwa data uji adalah data pada minggu keempat di hari kerja. 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. Selanjutnya, lakukan proses normalisasi seperti proses normalisasi pada data latih. Proses berikutnya adalah pengujian pada model GRNN yang telah dipilih. Dari hasil proses pengujian ini akan dihitung akurasi dan kesalahannya menggunakan MAPE yang secara matematis dinyatakan pada persamaan (6). Proses pelatihan dan pengujian sistem dapat dilihat pada Gambar 3.



Gambar 3. Flowchart Proses Pelatihan dan Pengujian Sistem

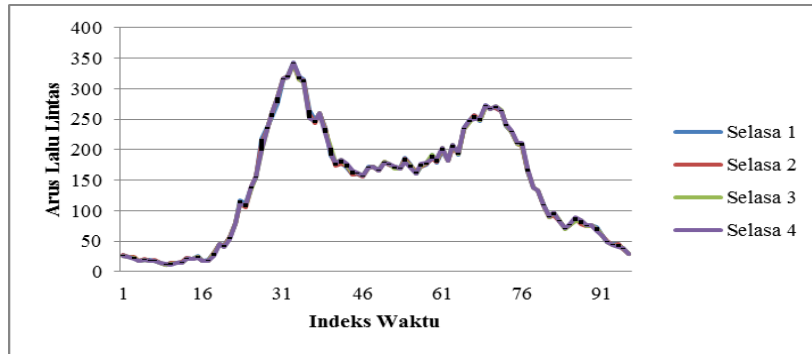
4. Hasil dan Pembahasan

Dataset yang digunakan adalah data arus lalu lintas di Inggris yang dihimpun oleh *Traffic High Way Agency* dan dapat diunduh secara bebas pada laman (<https://data.gov.uk/dataset/dft-eng-srn-routes-journey-times>). Pada dataset tersebut, data arus lalu lintas ditampilkan pada setiap periode waktu 15 menit dan dinyatakan dalam indeks waktu dalam sehari (1-96), dimana 1 menunjukkan pukul 00:00 sedangkan 96 menunjukkan pukul 23:45 waktu setempat.

Penelitian ini menggunakan data arus lalu lintas pada suatu ruas jalan dengan kode linkref AL100. Data yang digunakan adalah data di hari kerja (Senin-Jumat) pada bulan November 2014. Data arus lalu di hari kerja memiliki pola yang sama di hari yang sama seperti pada Gambar 4 dan tidak dipengaruhi oleh keadaan-keadaan yang tidak normal seperti kemacetan, kecelakaan lalu lintas ataupun cuaca yang ekstrim. Pada Gambar 4, Selasa 1 sampai 4 menyatakan arus lalu lintas pada hari Selasa di minggu pertama hingga Selasa di minggu ke-4.

Hasil percobaan dilakukan untuk mengetahui performa model prediksi arus lalu lintas. Performa model prediksi menggunakan metode yang diusulkan yaitu metode hibrida *K-means* dan GRNN dibandingkan dengan model prediksi arus lalu lintas menggunakan metode GRNN. Pada penelitian ini, proses *cluster* data arus lalu lintas menggunakan *K-means* dengan $K=3$. Metode Hibrida yang diusulkan, diterapkan dengan melakukan *cluster* data seluruh hari dan *cluster* data di hari yang sama. Sehingga pada penelitian ini terdapat dua sistem prediksi berdasarkan metode hibrida yang diusulkan, yaitu sistem prediksi dengan *cluster* data seluruh hari dan sistem prediksi dengan *cluster* data di hari yang sama. Yang dimaksud dengan *cluster*

data seluruh hari adalah proses *cluster* dengan menggunakan data di seluruh hari kerja (Senin-Jumat), sedangkan *cluster* data di hari yang sama adalah proses *cluster* dengan menggunakan data di hari yang sama pada hari kerja. Data yang di-*cluster* adalah hanya data latih yaitu data pada minggu ke-1, ke-2, dan ke-3 (hanya hari kerja). Data latih tersebut di-*cluster* dan setiap *cluster* akan membentuk model GRNN. Sedangkan data pada minggu ke-4 digunakan sebagai data uji dan tidak perlu di-*cluster*. MAPE digunakan untuk mengevaluasi kinerja model prediksi arus lalu lintas pada penelitian ini.



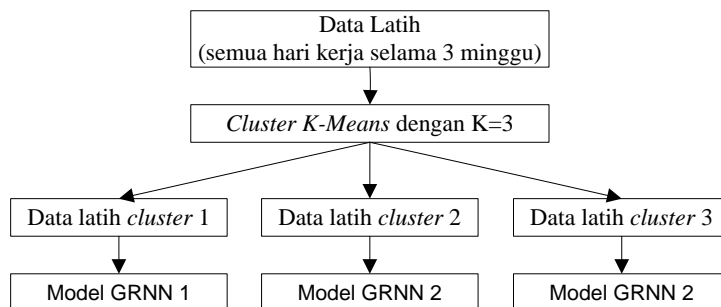
Gambar 4. Grafik Arus Lalu Lintas di Hari Selasa Bulan November 2014

Model prediksi arus lalu lintas dengan GRNN, hanya membentuk sebuah model prediksi dengan menggunakan data latih selama tiga minggu (di hari kerja) dan data uji selama seminggu (di hari kerja). Performa model prediksi arus lalu lintas menggunakan metode GRNN dapat dilihat pada Tabel 2. Kolom jumlah *neuron* mendeskripsikan banyaknya nilai-nilai variabel prediktor yang digunakan dengan periode setiap 15 menit. Sebagai contoh, jumlah *neuron* sebanyak 2 menyatakan data arus lalu lintas pada $t-1$ dan t yang digunakan untuk memprediksi arus lalu lintas pada waktu berikutnya yaitu $t+1$.

Tabel 2. Hasil Prediksi Arus Lalu Lintas dengan Metode GRNN

Jumlah Neuron	MAPE (%)					Rata-rata MAPE (%) Setiap Jumlah Neuron
	Senin	Selasa	Rabu	Kamis	Jumat	
2	10.48	7.32	8.54	8.26	7	8.32
3	8.17	5.58	6.57	6.55	5.01	6.376
4	6.88	4.85	5.27	4.84	4.01	5.17
5	5.89	4.19	4.62	4.16	3.23	4.418
6	5.16	3.63	3.92	3.82	2.85	3.876

Model prediksi arus lalu lintas dengan metode yang diusulkan pada sistem prediksi dengan meng-*cluster* data seluruh hari, membentuk model prediksi sebanyak jumlah *cluster* pada data latih yaitu tiga buah model GRNN, seperti yang dideskripsikan pada Gambar 5. Setiap variabel data uji (nilai-nilai prediktor atau satu set atribut prediktor) hanya menggunakan sebuah model GRNN untuk proses pengujian.



Gambar 5. Block Diagram Pembentukan Model GRNN Berdasarkan Cluster K-Means pada Data Latih di Seluruh Hari Kerja

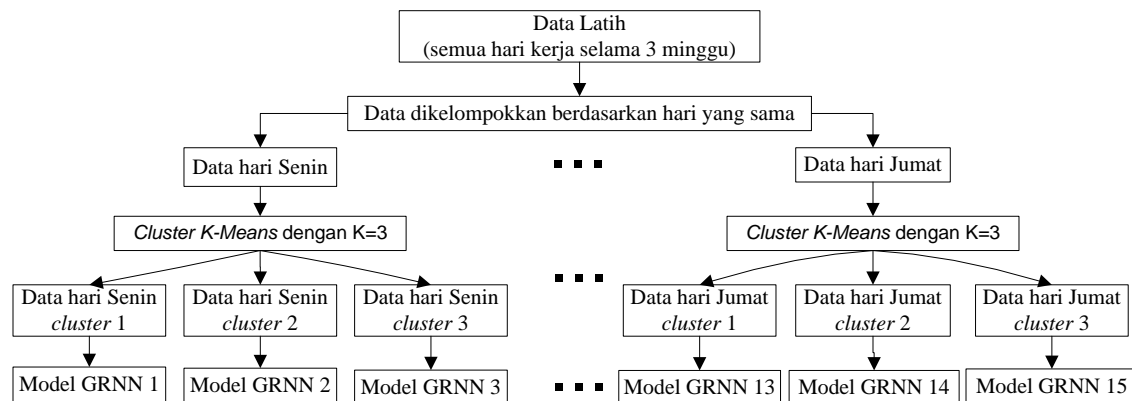
Pada sub bab 3.3 telah dijelaskan tentang pemilihan model GRNN yang akan digunakan pada proses pengujian yaitu dengan cara menghitung jarak *Euclidean* antara nilai-nilai prediktor dengan nilai *centroid* dari masing-masing data latih pada suatu model GRNN yang telah terbentuk. Dari perhitungan tersebut, 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. Tiap model prediksi yang terbentuk masing-masing memiliki estimasi parameter *smoothing factor* terbaik. Pada penelitian ini *smoothing factor* yang digunakan berada pada rentang [0,05, 1] dengan interval 0,05. Parameter tersebut dipilih berdasarkan kriteria nilai MAPE yang paling minimal pada proses validasi menggunakan LOOCV. Performa model prediksi dengan *cluster* data seluruh hari dapat dilihat pada Tabel 3.

Tabel 3. Hasil Prediksi Arus Lalu Lintas dengan Metode K-Means-GRNN dengan Cluster Data Seluruh hari

Jumlah Neuron	Mape (%)					Rata-rata MAPE (%) Setiap Jumlah Neuron
	Senin	Selasa	Rabu	Kamis	Jumat	
2	6.67	6.16	8.27	8.12	5.23	6.89
3	5.17	4.45	5.43	5.96	3.61	4.924
4	4.05	3.17	3.58	3.55	3.09	3.488
5	3.01	2.25	2.8	3.06	2.31	2.686
6	2.4	1.98	2.28	2.84	2.03	2.306

Nilai rata-rata MAPE pada Tabel 2 dan Tabel 3 menunjukkan bahwa performa model prediksi menggunakan metode hibrida yang diusulkan dengan *cluster* data seluruh hari memiliki performa yang lebih baik dari pada model prediksi menggunakan metode GRNN.

Gambar 6 mendeskripsikan sistem prediksi metode K-Means-GRNN dengan *cluster* data di hari yang sama. Banyaknya model prediksi yang terbentuk adalah $K \times H$. Dimana K adalah jumlah *cluster*, yaitu tiga dan H adalah jumlah hari yang digunakan, yaitu lima hari (Senin-Jumat). Dengan kata lain banyaknya model GRNN yang terbentuk pada sistem prediksi ini adalah 15 GRNN dengan rincian yaitu data latih di hari Senin sampai Jumat yang masing-masing membentuk tiga model GRNN berdasarkan jumlah *cluster* yang dilakukan.



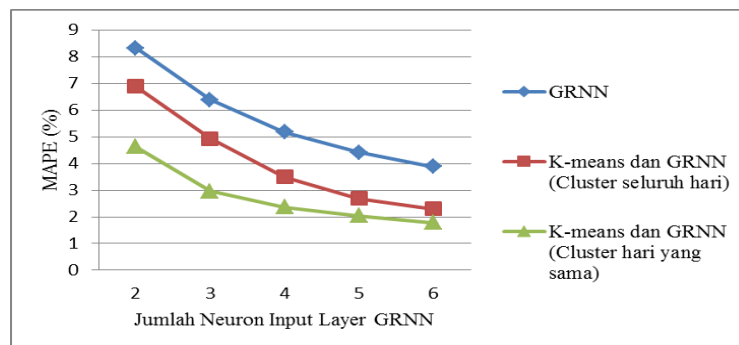
Gambar 6. Block Diagram Pembentukan Model GRNN Berdasarkan Cluster K-Means pada Data Latih di Hari yang Sama

Proses pengujian pada sistem prediksi ini, terlebih dahulu harus mengetahui informasi hari dari satu set atribut prediktor tersebut. Misalnya satu set atribut prediktor tersebut merupakan hari Selasa, selanjutnya akan memilih salah satu dari tiga model GRNN yang telah dibentuk oleh kumpulan data latih hari Selasa. Selanjutnya pilih Model GRNN yang data latihnya memiliki nilai *centroid* yang dekat dengan satu set atribut prediktor. Performa model prediksi arus lalu lintas menggunakan metode *K-means* dan GRNN berdasarkan *cluster* data di hari yang sama dapat dilihat pada Tabel 4.

Tabel 4. Hasil Prediksi Arus Lalu Lintas dengan Metode K-Means-GRNN dengan Cluster Data di Hari yang Sama

Jumlah Neuron	Mape (%)					Rata-rata MAPE(%) Setiap Jumlah Neuron
	Senin	Selasa	Rabu	Kamis	Jumat	
2	4.9	3.95	5.22	4.86	4.16	4.618
3	3.61	2.09	3.51	3.32	2.31	2.968
4	2.87	1.93	2.72	2.34	1.93	2.358
5	2.47	1.51	2.49	2	1.72	2.038
6	1.89	1.63	1.87	1.96	1.53	1.776

Gambar 7 merupakan hasil perbandingan performa model prediksi arus lalu lintas berdasarkan rata-rata MAPE pada setiap jumlah *neuron*, yang menunjukkan bahwa model prediksi menggunakan metode hibrida *K-means-GRNN* memiliki performa yang lebih baik dibandingkan menggunakan metode GRNN. Hal ini dikarenakan pada metode *K-means-GRNN*, data latih yang membentuk suatu model GRNN adalah data latih yang telah dipilih berdasarkan kemiripan karakteristiknya. Dengan adanya proses *cluster* pada data latih, maka tingkat variasi data latih pada model GRNN akan berkurang. Model prediksi arus lalu lintas menggunakan metode *K-means-GRNN* berdasarkan *cluster* data di hari yang sama memiliki performa yang lebih baik dibandingkan dengan *cluster* di seluruh hari, karena dataset arus lalu lintas di hari yang sama selama sebulan di hari kerja memiliki karakteristik yang lebih mirip (Gambar 4). Oleh karena itu *cluster* data latih di hari yang sama pada model GRNN dapat mengurangi tingkat variasi data apabila dibandingkan dengan data latih yang di-*cluster* di seluruh hari.

**Gambar 7. Grafik Perbandingan Performa Berbagai Model Prediksi**

Tidak hanya penggunaan data latih yang sesuai pada model GRNN, akan tetapi penggunaan banyaknya data pada periode waktu sebelumnya untuk memprediksi arus lalu lintas pada waktu berikutnya juga mempengaruhi kinerja performa model prediksi. Banyaknya data pada periode waktu sebelumnya direpresentasikan dengan jumlah *neuron* pada *input layer* GRNN. Berdasarkan grafik pada Gambar 7, model prediksi arus lalu lintas yang menggunakan data pada periode $t-5$, $t-4$, $t-3$, $t-2$, dan t untuk memprediksi data pada $t+1$ memiliki performa yang paling baik untuk semua metode prediksi yang dilakukan. Dari hasil percobaan dilakukan, model prediksi arus lalu lintas dengan performa terbaik diukur berdasarkan nilai MAPE adalah model prediksi menggunakan metode K-Means-GRNN dengan *cluster* data di hari yang sama dan menggunakan enam data pada periode sebelumnya.

5. Kesimpulan

Prediksi arus lalu lintas menggunakan metode hibrida *K-means* dan GRNN pada penelitian ini mampu memberikan performa model prediksi yang lebih baik dari pada model prediksi yang hanya menggunakan metode GRNN berdasarkan evaluasi kriteria MAPE. Hasil tersebut dapat ditunjukkan dengan menurunnya nilai MAPE sebesar 0,82-3,81% pada hasil percobaan dengan *cluster* data seluruh hari. Sedangkan untuk hasil percobaan dengan *cluster* data di hari yang sama memiliki nilai MAPE yang cukup kecil yaitu nilai MAPE di bawah nilai 6%. Hasil yang diperoleh tersebut terbukti dapat meningkatkan hasil akurasi prediksi arus lalu lintas yang signifikan.

Referensi

- Dia, H., 2001. An Object Oriented Neural Network Approach to Short Term Traffic Forecasting. *European Journal of Operational Research*, 131(2), pp.253–261.
- Enrique, C., 2008. Predicting Traffic Flow Using Bayesian Networks. *Transportation Research Part B*, 42, pp.482–509.
- Han, C., Song, S. & Wang, C.H., 2004. A Real Time Short Term Traffic Flow Adaptive Forecasting Method Based on ARIMA Model. *Journal of System Simulation*, 16(7), pp.1530–1535.
- Wang, H.W. & Lin, J.L. 2010. Sales Prediction based on Improved GRNN. In *Computer Engineering and Science*. pp. 153–155.
- Jeong, C., Min, J.H. & Kim, M.S., 2012. A Tuning Method For The Architecture of Neural Network Models Incorporating GAM and GA As Applied To Bankruptcy Prediction. *Expert Systems with Applications*, 39(3), pp.3650–3658.
- Julianto, E.N., 2010. Hubungan Antara Kecepatan, Volume dan Kepadatan Lalu Lintas Ruas Jalan Siliwangi Semarang. *Jurnal Teknik Sipil & Perencanaan*, 12(2), pp.151–160.
- Kuang, X., Wu, C., Huang, Y., & Xu, L., 2010. Traffic Flow Combination Forecasting Based on Grey Model and GRNN. In *International Conference on Intelligent Computation Technology and Automation*. pp. 1072–1075.
- Leng, Z., Gao, J., Qin, Y., Liu, X., & Yin, J., 2013. Short-term Forecasting Model of Traffic Flow Based on GRNN. In *Chinese Control and Decision Conference*. pp. 3816–3820.
- Makridakis, S., 1989. Why Combining works? *International Journal of Forecasting*, 5(4), pp.601–603.
- Specht, D.F., 1991. A General Regression Neural Network. *IEEE Transactions on Neural Network*, 2(6), pp.568–576.
- Sun, S., Zhang, C. & Yu, G., 2006. A Bayesian Network Approach to Traffic Flow Forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 7(1), pp.124–131.
- Tan, M.C., Feng, L.B. & Xu, J.M., 2007. Traffic Flow Prediction based on Hybrid ARIMA and ANN Model. In *Power Electronics and Intelligent Transportation System*. pp. 118–121.
- Tan, M., Wong, S. C., Xu, J., Guan, Z., & Zhang, P., 2009. An Aggregation Approach to Short-Term Traffic Flow Prediction. *IEE Transactions on Intelligent Transportation Systems*, 10(1), pp.60–69.
- Vlahogianni, I.E., Karlaftis, M.G. & Golias, J.C., 2006. Statistical Methods for Detecting Non Linearity and Non Stationer in Univariate Short Term Time Series of Traffic Volume. *Transportation Research Part C: Emerging Technologies*, 14, pp.351–367.
- Williams, B.M., 2001. Multivariate Vehicular Traffic Flow Prediction: Evaluation of ARIMAX modeling. In *Transportation Research Record*. pp. 194–200.
- Yin, J., Zhou, D. & Xie, Q., 2006. A Clustering Algorithm For Time Series Data. In *Proceedings of the Seventh International Conference on Parallel and Distributed Computing, Application and Technologies*. pp. 1–4.
- Zhang, G.P., 2003. Time Series Forecasting Using A Hybrid ARIMA and Neural Network Model. *Neurocomputing*, 50, pp.159–175.
- Zhanquan, S., Yinglong, W. & Jingshan, P., 2008. Short Term Traffic Flow Forecasting Based on Clustering and Feature Selection. In *International Joint Conference on Neural Networks*. pp. 577–583.