Gaung Informatika, Volume 13, Nomor 1, Januari-Juni 2020, ISSN 2086-4221 EISSN 2721-3900

PENGUKURAN PERFORMA SUPPORT VECTOR MACHINE DAN NEURAL NETWOK DALAM MERAMALKAN TINGKAT CURAH HUJAN

Diyah Ruswanti
Program Studi Teknik Informatika, Universitas Sahid Surakarta
Jl. Adi Sucipto No.154, Jajar, Surakarta
Email: diyahruswa@gmail.com

Abstract

Prediction models using Neural Network or Support Vector Machine have been developed in many areas of rainfall. In this studi we have compared the performance of NN and SVM models, in predition of monthly rainfall at R-17 station Kecepit Pemalang. The analyzed using NN and SVM in which testing with Root Mean Square Error (RMSE) for get the performance is done. The data obtained for 2009 to 2018 monthly rainfall were used as modelling and forecasting sample. The results showed that NN obtained smallest error rate compared to SVM. the recognized value of RMSE for SVM is 176,374, Neural Network is 22,289. In RMSE the smallest error rate showed the best performance of algorithm.
Key word : *neural network, support vector machine, performance, rainfall, prediction*

Pendahuluan

Latar Belakang

Sebagai Negara yang berada di garis khatulistiwa, Indonesia memiliki dua musim yaitu musim hujan dan musim kemarau. Adanya musim hujan dan musim kemarau ini mempengaruhi kehidupan mahluk hidup di wilayan Negara Indonesia, sehingga masyarakat harus mengenali faktor yang menyebahkan musim tersebut sehingga dapat melakukan perkiraan kegiatan yang akan dilakukan pada kedua musim tersebut. Salah satu faktor penyebab kedua musim tersebut adalah curah hujan.

Curah hujan merupakan jumlah air hujan yang jatuh selama periode waktu tertentu yang pengukurannya menggunakan satuan tinggi di atas permukaan tanah horizontal yang diasumsikan tidak terjadi infiltrasi, run off, maupun evaporasi. Definisi curah hujan atau yang sering disebut presipitasi dapat diartikan jumlah air hujan yang turun di daerah tertentu dalam satuan waktu tertentu. Jumlah curah hujan merupakan volume air yang terkumpul di permukaan bidang datar dalam suatu periode tertentu (harian, mingguan, bulanan, atau tahunan). Curah hujan merupakan jumlah air yang jatuh di permukaan tanah datar selama periode tertentu yang diukur dengan satuan tinggi milimeter (mm) di atas permukaan horizontal. Hujan juga dapat diartikan sebagai ketinggian air hujan yang terkumpul dalam tempat yang datar, tidak menguap, tidak meresap dan tidak mengalir.

Pengertian curah hujan dapat juga dikatakan sebagai air hujan yang memiliki ketinggian tertentu yang terkumpul dalam suatu penakar hujan, tidak meresap, tidak mengalir, dan tidak menyerap (tidak terjadi kebocoran). Tinggi air yang jatuh ini

biasanya dinyatakan dengan satuan milimeter. Curah hujan dalam 1 (satu) millimeter artinya dalam luasan satu meter persegi, tempat yang datar dapat menampung air hujan setinggi satu mm atau sebanyak satu liter.

Kesepakatan internasional di seluruh dunia menyatakan bahwa curah hujan mempunyai peran yang sangat penting baik dalam dunia penerbangan, meteorologi dan yang lainnya. Selain itu curah hujan yang tingi dapat menyebabkan banjir. Banyak kejadian banjir yang melanda Indonesia saat ini, yang kadang kadang datang dengan tiba tiba. Salah satu pertanda datangnya banjir adalah dengan melihat tingkat curah hujan (Sari, 2016). Curah hujan dengan rata-rata 200mm sampai dengan 400mm masuk dalam pengelompokkan sifat perkiraan surah hujan yang normal. Dias angka tersebut menjadikan sifat perkiraan surah hujan yang tinggi dan dibawah 200mm merupakan perkiraan curah hujan yang rendah. Dalam suatu kelompom data curah hujan, sifat curah hujan dibagi menjadi tiga, yaitu Aas Normal jika nilai curah hujan lebih dari 115% terhadap rata-ratanya, Nomral jika nilai curah hujan berkisar antara 85-115% terhadap rata-ratanya, dan Bawah Normal jika nilai curah hujan kurang dari 85% terhadap rata-ratanya. Di Kabupaten Pemalang, tingkat curah hujan dicatat oleh stasiun R-71 Kecepit Kabupaten Pemalang.

Data mining adalah ilmu yang mempelajari bagaimana cara mengolah data sehingga menjadi sesuatu hal yang lebih bermakna. Terdapat lima metode dalam data mining yaitu Estimasi, Prediksi, Klasifikasi, Klastering dan Asosiasi. Data mining telah digunakan dalam berbagai hal (Han et al., 2012a). Sering disebut Knowledge Discovery in Database (KDD) yang mempunyai kegiatan melakukan pengumpulan data, pemakaian data masa lalu untuk menemukan suatu pola keteraturan, menemukan pola atau hubungan dalam dataset berukuran besar. Output data mining dapat digunakan sebagai landasan untuk pengambilan keputusan menjadi lebih baik.

Salat satu metode dalam data mining adalah prediksi, yaitu melakukan proses pengestimasian nilai prediksi berdasarkan pola-pola di dalam sekumpulan data. Prediksi menggunakan beberapa variabel atatu field-field basis data untuk memprediksi nilai-nilai variabel masa mendatang yang diperlukan, yang belum diketahui saat ini. Algoritma prediksi biasanya digunakan untuk memperkirakan atau *forecasting* suatu kejadian sebelum kejadian atau peristiwa tertentu terjadi. Contohnya pada bidang Klimatologi dan Geofisika, yaitu bagaimana Badan Meterologi Dan Geofisika (BMKG) memperkirakan tanggal tertentu bagaimana Cuacanya, apakah Hujan, Panas dan lain sebagainya.

Mengukur kinerja algoritma peramalan menjadi hal yang penting, dikarenakan dapat dilihat seberapa baik hasil yang didapatkan dari algoritma peramalan tersebut. Salah satu cara mengukur algoritma peramalan adalah dengan *Root Mean Square Error (RMSE). RMSE* adalah metode alternatif untuk mengevaluasi teknik peramalan yang digunakan untuk mengukur tingkat akurasi hasil prakiraan suatu model. *RMSE* merupakan nilai rata-rata dari jumlah kuadrat kesalahan, juga dapat menyatakan ukuran besarnya kesalahan yang dihasilkan oleh suatu model prakiraan. Nilai *RMSE* rendah menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan mendekati variasi nilai obeservasinya (Gosasang et al., 2011).

Neural Network (NN) dan Support Vector Machine (SVM) merupakan algoritma peramalan yang mempunyai kinerja yang berbeda. Support Vector Machine, adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik

untuk metode klasifikasi maupun regresi. Dalam hal fungsi dan kondisi permasalahan yang ada, SVM masih satu kelas dengan Neural Network, dimana keduanya merupakan supervised learning. Perbedaannya SVM menemukan solusi yang global optimal sedangkan neural network local optimal. Dalam SVM, yang dicari adalah sebuah fungsi pemisah (hyperplane) yang optimal yang bisa memisahkan dua kelompok set data dari dua kelas yang berbeda (Hoang & Pham, 2015).

Permasalahan

Bagaimana mengetahui kinerja algoritma data mining yaitu SVM dan NN yang mempunyai tingkat kesalahan terkecil dalam melakukan peramalan tingkat curah hujan di Kabupaten Pemalang.

Tujuan Penulisan

Tujuan dari penelitian ini adalah untuk mengetahui kinerja dua algoritma peramalan dengan proses penemuan solusi yang berbeda yaitu NN dan SVM dalam melakukan peramalan tingkat curah hujan di Kabupaten Pemalang.

Landasan Teori

Data Mining

Data mining digunakan untuk melakukan pemgolahan data, agar data yang ada dapat diketahui informasi tersembunyi di dalamnya. Berbagai macam metode yang ada di data mining dengan model pembelajaran supervised dan unsupervised learning. Ada lima metode pokok dalam data mining yaitu estimasi, prediksi, klasifikasi, estimasi, dan asosiasi (Berry & Linoff, 2004). Fungsi Prediksi (prediction) merupakan proses untuk menemukan pola dari data dengan menggunakan beberapa variabel untuk memprediksikan variabel lain yang tidak diketahui jenis atau nilainya. Fungsi Deskripsi (description) merupakan proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data. Fungsi Klasifikasi (classification) merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan class atau konsep dari suatu data. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan. Fungsi Asosiasi (association) merupakan proses yang digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data (Han et al., 2012).

Terdapat lima langkah dalam melakukan pemrosesan data menggunakan data mining. Yang pertama adalah Data selection, yaitu pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional. Yang kedua adalah proses Pre-processing / cleaning, dimana sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Yang ketiga adalah Transformation, coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data. Yang ke empat adalah Data mining itu sendiri, yaitu proses mencari pola atau informasi

menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Yang terakhir adalah Interpretation / evalution, dimana pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya(Han et al., 2012).

Pada aplikasi data mining, algoritma yang digunakan diukur untuk mengetahui kinerja dari algoritma data mining. Pengukuran dapat dilakukan dengan berbagai macam algoritma, seperti root mean square error, davies bouldin index, confussion matrix, kurva ROC dan lain sebagainya (Donohue et al., 2017).

***Support Vector Machine* (SVM)**

Metode data mining yang mampu menghasilkan nilai akurasi yang baik untuk data dan atribut yang banyak dan beragam adalah Support Vector Mchine (SVM). SVM dalam proses pembelajarannya menggunakan supervised learning. Fungsi linier dapat diefinisikan sebagai berikut:

$g(x) = sgn(f(x))$

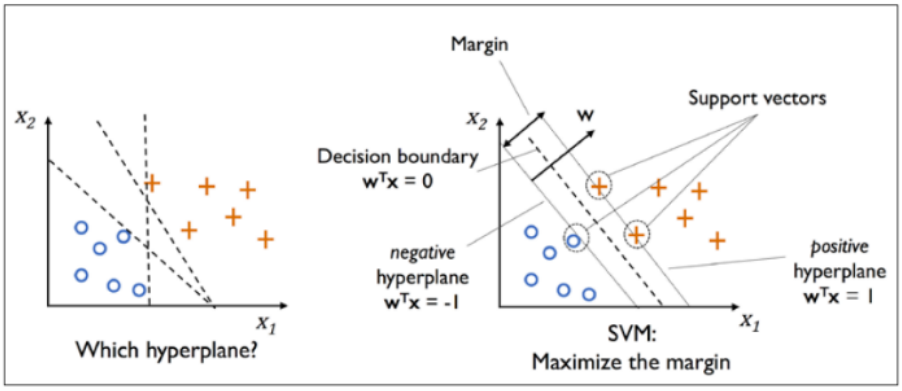
dengan $f(x)=w^Tx+b$

dimana $x, w \in Rn$ dan $b \in R$.

Set parameter (w,b), sehingga $f(xi) = <x,w>+b = yi$ untuk semua i

(Hoang & Pham, 2015).

SVM digunakan untuk mencari *hyperplane* terbaik dengan memaksimalkan jarak antar kelas. H*yperplane* adalah sebuah fungsi yang dapat digunakan untuk pemisah antar kelas. Dalam 2-D fungsi yang digunakan untuk klasifikasi antar kelas disebut sebagai *line whereas*, fungsi yang digunakan untuk klasifikasi antas kelas dalam 3-D disebut *plane similarly*, sedangan fungsi yang digunakan untuk klasifikasi di dalam ruang kelas dimensi yang lebih tinggi di sebut *hyperplane*. Mencari hyperplane terbaik ekuivalen dengan memaksimalkan margin atau jarak antara dua set obyek dari kelas yang berbeda (Deng et al., 2013). Gambar 1 menunjukkan hyperplane yang memisahkan dua kelas.



Gambar 1. Hyperplane yang memisahkan dua kelas positif dan kelas negatif

Jika wx1+b=+1 adalah hyperplane pendukung dari kelas +1 dan wx2+b=-1 adalah hyperplane pendukung dari kelas -1, maka margin dapat dihitung dengan mencari jarak kedua hyperplane pendukung. Sehingga (wx1+b=+1)-(wx2+b=-1) =>w(x1-x2)=2 (Hoang & Pham, 2015).

Neural Network

Ada tiga lapis proses pada metode neural network, yang disebut neural layers yaitu lapisan input (menerima pola inputan data dari luar yang menggambarkan permasalahan), lapisan tersembunyi(hiden layer) dan lapisan output (merupakan solusi terhadap suatu permasalahan). Pada masalah yang penelitian ini, dilihat dari data yang ada, merupakan masalah yang sederhana dengan 2 kelas sehingga dipilih arsitektur ANN yang single layers network. Untuk formulasi ANN adalah sebagai berikut: Jika net = $\sum x_i w_i$ maka fungsi aktivasinya adalah f(net)=f($\sum x_i w_i$). Beberapa fungsi aktivasi yang digunakan adalah Fungsi threshold, fungsi sigmoid dan fungsi identitas (Juahir et al., 2004).

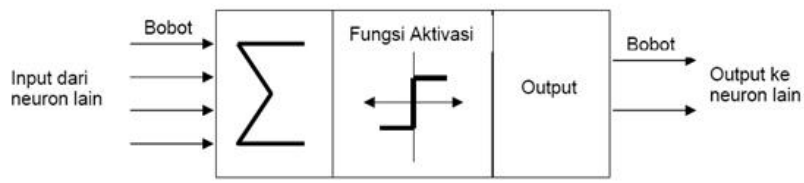
Karakteristik dari ANN dilihat dari pola hubungan antar neuron, metode penentuan bobot dari tiap koneksi, dan fungsi aktivasinya. Gambar 2 menjelaskan struktur ANN secara mendasar, yang dalam kenyataannya tidak hanya sederhana seperti itu. Bagian Input, berfungsi seperti dendrite pada otak manusia. Bagian Output berfungsi seperti akson dan Fungsi aktivasi, berfungsi seperti sinapsis. Beberapa fungsi aktivasi yang digunakan adalah :

- Fungsi threshold (batas ambang) $f(x)= \begin{cases} 1...x \geq a \\ 0...x \leq a \end{cases}$

  untuk kasus bilangan bipolar, 0 diganti dengan angka -1 sehingga persamaan diubah: $f(x)= \begin{cases} 1...x \geq a \\ -1...x \leq a \end{cases}$
- Fungsi sigmoid $f(x) = 1/(1+e^{-x})$

  Fungsi ini sering digunakan karena nilai fungsinya yang sangat mudah untuk didiferensiasikan, f(x)=f(x)(1-f(x))
- Fungsi identitas F(x)=x

  Digunakan jika keluaran yang dihasilkan oleh ANN merupakan sembarang bilangan real.



Gambar 2. Fungsi Aktivasi pada Neural Network

Root Mean Square Error (RMSE)

Root Mean Square Error merupakan suatu ukuran kesalahan yang didasarkan pada selisih antara dua buah nilai yang bersesuaian, yang didefinisikan sebagai berikut:

$$\text{RMSE} = \left| \sum_i \sum_d \frac{\left(\hat{T}_{id} - T_{id}\right)^2}{\text{N}.(\text{N}-1)} \right|^{\frac{1}{2}} \quad \text{untuk } i \neq d$$

Sedangkan deviasi standar dari selisih kedua nilai tersebut didefinisikan sebagai berikut:

$$\sigma = \left| \sum_i \sum_d \frac{\left(\hat{T}_{id} - T_{id}\right)^2}{\text{N}.(\text{N}-1)-1} \right|^{\frac{1}{2}} \quad \text{untuk } i \neq d$$

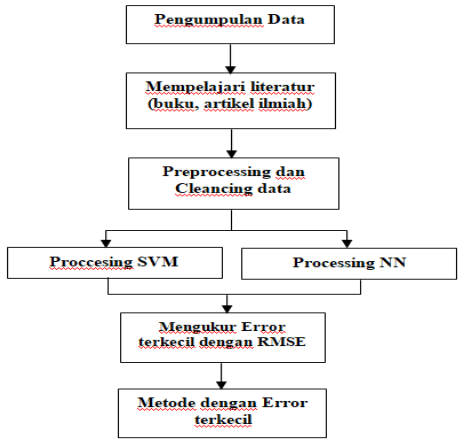
Dapat dilihat bahwa untuk nilai N yang besar, kedua persamaan diatas akan menghasilkan nilai yang bisa dikatakan sama. Dalam hal ini RMSE dapat disebut sebagai deviasi standar atau sebaliknya. Indikator RMSE dan Deviasi Standar tidak dapat membandingkan sebuah model yang sama jika diterapkan pada wilayah studi yang berbeda, karena nilai nilainya tergantung pada ukuran besarnya Matriks. Presentasi dari akar kesalahan kuadrat (prosentase RMSE) dapat mengatasi hal ini, dan didefinisikan %RMSE=(RMSE/T_i) x 100%. Semakin besar nilai RMSE, prosentase RMSE, deviasi standar dan koefisien variasinya mempaunyai arti bahawa semakin tidak tepat hasil dari algoritma peramalan yang digunakan (Gosasang et al., 2011).

Dataset

Dataset adalah suatu kumpulan data. Dalam kasus data tabular, satu set data sesuai dengan satu atau lebih tabel database, di mana setiap kolom tabel mewakili variabel tertentu, dan setiap baris sesuai dengan catatan tertentu dari set data yang dimaksud. Dalam penelitian ini dataset yang digunakan adalah data curah hujan di Stasiun.. Kabupaten Pemalang dalam kurun waku bulan Januari tahun 2009 sampai dengan bulan Desember tahun 2018. Tabel 1 menunjukkan dataset yang digunakan dalam penelitian ini yang berasal dari stasiun R-71 Kecepit Randudongkal Kab Pemalang.

Metodologi Penelitian

Metodologi pada penelitian ini terbagi menjadi beberapa tahapan yaitu:



Gambar 3. Tahapan Penelitian

Pada setiap tahapan yang dilakukan adalah :

1. Pengumpulan dataset
   Data set yang digunakan merupakan data curah hujan dari tahun 2009 sampai dengan tahun 2018 (120 bulan) pada stasiun R-71 Kabupaten Pemalang. Dataset ini didapatkan dari situs data.go.id.
2. Studi literatur
   Meliputi studi literatur untuk referensi dalam penelitian berupa jurnal dan karya ilmiah yang relevan dengan penelitian.
3. Preprocessing dan Cleansing data, meliputi data cleansing: membersihkan noise dan data yang tidak konsisten sehingga diperoleh data training dan data testing; data reduction: untuk menghapus atribut-atribut yang tidak diperlukan seperti pada row yang atributnya tidak lengkap atau tidak terisi dilakukan handling missing value.
4. Pemodelan, setelah data benar-benar bersih dari *empty field*, *garbage,*dan *redundancy* baru bisa diolah untuk kemudian pemilihan tipe validasi dimana menggunakan validasi tipe *Split Validation* baru kemudian dilakukan ditentukan metode untuk pemodelan. Pemodelan pertama dilakukan dengan menggunakan metode Support Vector Machine baru kemudian membuat pemodelan lain dengan menggunakan metode Neural Network.
5. Pengukuran Kinerja, untuk selanjutnya dilakukan pengukuran performansi dari masing-masing metode menggunakan metode RMSE.

Hasil dan Pembahasan

Data yang digunakan sebanyak 120, dengan data missing sebanyak 8 buah data. Data merupakan data bulanan curah hujan di stasiun pemantauan hujan R-71 Randudongkal Kabupaten Pemalang. Tabel 1 menunjukkan data yang digunakan.

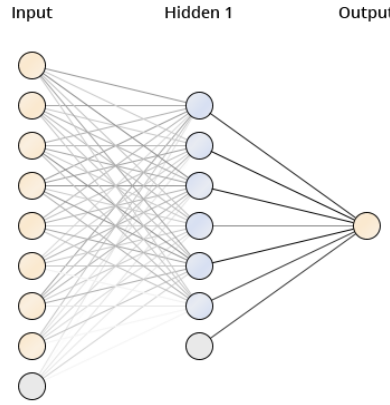
Tabel 1. Data curah hujan Stasiun R-71 Kecepit Randudongkal Kab. Pemalang

Tahun /Bulan	Jan	Peb	Mar	Apr	Mei	Jun	Jul	Ags	Sep	Okt	Nop	Des
2009	807	753	470	773	595	122	0	0	4	120	366	343
2010	802	722	504	438	337	261	245	105	304	273	183	429
2011	1088	785	405	310	309	96	127	0	7	248	301	836
2012	850	354	266	196	104	51	3	0	0	122	308	386
2013	756	532	220	469	193	320	217	39	79	182	209	463
2014	511	749	300	440	198	143	220	50	0	67	291	269
2015	754	467	421	380	145	76	11	5	0	11	272	452
2016	337	500	326	323	244	316	165	100	411	343	402	520
2017	592	517	335	318	142	156	43	5	70	154	340	417
2018	386	991	274	362	184	87	45	0	24	74	294	362

Dataset awal dilakukan proses normalisasi, yang bertujuan untuk menskala nilai sehingga cocok dalam rentang tertentu dan menyesuaikan rentang nilai sangat penting ketika berhadapan dengan atribut dan skala yang berbeda. Beirikut adalah hasil dari normalisasi dataset curah hujan:

```
Normalize 8 attributes to mean 0 and variance 1.
Using
2010.0 --> mean: 383.5833333333333, variance: 43811.7196969697
2009.0 --> mean: 362.75, variance: 99324.20454545454
2016.0 --> mean: 249.5, variance: 60599.90909090909
2015.0 --> mean: 269.8333333333333, variance: 45825.969696969696
2017.0 --> mean: 332.25, variance: 14971.295454545454
... 3 more attributes ...
```

Setelah proses pengolahan data awal selesai, maka data set untuk penelitian siap untuk digunakan. Validasi menggunakan Split Validation. Pemilihan Split validation sendiri agar lebih memudahkan penelitian menentukan pembagian data traning dan testing dengan ratio yg bisa kita tentukan sendiri. Setelah pemilihan jenis validasi, kemudian memilih metode mana yang akan dipakai untuk pemodelan, dalam penelitian ini dipilih *Neural Network* dasar untuk penelitian pertama, kemudian pemilihan penerapan model dan pengujian performance. Begitupun untuk penelitian metode *Support Vector Machine*. Metode yang diusulkan diawali dengan membagi dataset menjadi data training dan data testing dengan menggunakan 10-fold cross validation, yaitu dengan membagi data 90% untuk proses training dan 10% untuk proses testing.



Gambar 4. Improved Neural Network

Hidden 1
========
Node 1 (Sigmoid)

2009.0: -0.271
2010.0: -0.224
2012.0: -0.315
2013.0: -0.277
2014.0: -0.268
2015.0: -0.158
2016.0: -0.264
2017.0: -0.175
Bias: -0.069

Node 2 (Sigmoid)

2009.0: 0.171
2010.0: 0.092
2012.0: 0.194
2013.0: 0.209
2014.0: 0.175
2015.0: 0.054
2016.0: 0.109
2017.0: 0.065
Bias: 0.037

Node 3 (Sigmoid)

2009.0: -0.252
2010.0: -0.254
2012.0: -0.283
2013.0: -0.261
2014.0: -0.210
2015.0: -0.204
2016.0: -0.227
2017.0: -0.101
Bias: -0.118

Node 4 (Sigmoid)

2009.0: -0.165
2010.0: -0.130
2012.0: -0.108
2013.0: -0.127
2014.0: -0.120
2015.0: -0.023
2016.0: -0.117
2017.0: -0.080
Bias: -0.052

Node 5 (Sigmoid)

2009.0: 0.145
2010.0: 0.055
2012.0: 0.131
2013.0: 0.156
2014.0: 0.130
2015.0: 0.019
2016.0: 0.132
2017.0: 0.025
Bias: 0.016

Node 6 (Sigmoid)

2009.0: -0.217
2010.0: -0.169
2012.0: -0.151
2013.0: -0.197
2014.0: -0.127
2015.0: -0.093
2016.0: -0.145
2017.0: -0.094
Bias: -0.068

Output
======
Regression (Linear)

Node 1: -0.946
Node 2: 0.608
Node 3: -0.775
Node 4: -0.209
Node 5: 0.461
Node 6: -0.356
Threshold: 0.631

PerformanceVector:
root_mean_squared_error: 22.289 +/- 16.915 (micro average: 26.541 +/- 0.000)

Support Vector Machine
Total number of Support Vectors: 12
Bias (offset): 311.898
w[2009.0] = 0.887
w[2010.0] = 1.067
w[2012.0] = 1.057
w[2013.0] = 0.971
w[2014.0] = 1.113

w[2015.0] = 1.083
w[2016.0] = 1.313
w[2017.0] = 0.974
PerformanceVector:
root_mean_squared_error: 176.374 +/- 136.991 (micro average: 219.214 +/- 0.000)

Simpulan

Dari hasil pengukuran, diketahui nilai RMSE untuk SVM sebesar 176.374, Neural Network sebesar 22.289. Dari hasil keluaran RMSE tersebut maka algoritma yang mempunyai tingkat kesalahan terkecil dalam melakukan peramalan tingkat curah hujan di Kabupaten Pemalang adalah Algoritma Neural Network mempunyai tingkat kesalahan yang paling kecil dibandingan dengan SVM, sehingga NN lebih baik kinerjanya dalam melakukan peramalan tungkat surah hujan di Kab Pemalang dibandingkan dengan SVM.

Daftar Pustaka

Berry, M. J. a., & Linoff, G. S. (2004). Data mining techniques: for marketing, sales, and customer relationship management. In *Portal.Acm.Org*. http://portal.acm.org/citation.cfm?id=983642

Deng, N., Tian, Y., & Zhang, C. (2013). *Support Vector Machines*.

Donohue, M. C., Sun, C. K., Raman, R., Insel, P. S., & Aisen, P. S. (2017). Cross-validation of optimized composites for preclinical Alzheimer's disease. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, *3*(1), 123–129. https://doi.org/10.1016/j.trci.2016.12.001

Gosasang, V., Chandraprakaikul, W., & Kiattisin, S. (2011). *A Comparison of Statistical Technique and Neural Networks Forecasting Techniques for Container Throughput in Thailand*. *27*(3), 463–482.

Han, J., Kamber, M., & Pei, J. (2012a). Data Mining: Concepts and Techniques. In *San Francisco, CA, itd: Morgan Kaufmann*. https://doi.org/10.1016/B978-0-12-381479-1.00001-0

Han, J., Kamber, M., & Pei, J. (2012b). Data Mining: Concepts and Techniques. In *San Francisco, CA, itd: Morgan Kaufmann*. https://doi.org/10.1016/B978-0-12-381479-1.00001-0

Hoang, N.-D., & Pham, A.-D. (2015). Hybrid Artificial Intelligence Approach Based on Metaheuristic and Machine Learning for Slope Stability Assessment: A Multinational Data Analysis. *Expert Systems with Applications*, *46*, 60–68. https://doi.org/10.1016/j.eswa.2015.10.020

Juahir, H., Zain, S. M., Toriman, M. E., Mokhtar, M., & Man, H. C. (2004). *A PPLICATION OF A RTIFICIAL N EURAL N ETWORK M ODELS transportation . Langat River Basins is the most rapid urban area in Malaysia . The Langat River catchment straddles the main urban conurbation in the Klang Valley forming parts of the growing urban com*. *16*(2), 42–55.

Sari, H. L. (2016). FUZZY CLUSTERING DALAM PENGCLUSTERAN DATA CURAH HUJAN KOTA BENGKULU DENGAN ALGORITMA C-MEANS. *Jurnal Ilmiah Matrik*, *Vol 16 No*, 115–124.