

PREDIKSI CURAH HUJAN DI KAB.BANDUNG DENGAN ANALISIS TIME SERIES, MENGGUNAKAN MODEL SARIMA (*Seasonal Autoregressive Integrated Moving Average*)

Cesar Aini Soekendro

Sistem Informasi Telkom University Bandung

cesarkendro@student.telkomuniversity.ac.id

Abstrak

Prakiraan cuaca adalah salah satu kebutuhan utama untuk menunjang kegiatan diberbagai bidang , Informasi yang dibutuhkan adalah berupa prediksi cuaca atau iklim.Untuk mendapatkan hasil yang tepat dan akurat, banyak dikembangkan metode prakiraan cuaca yaitu contoh nya ARIMA dan SARIMA yang didasarkan beberapa aspek antara lain aspek waktu, aspek tingkat keakuratan prediksi yang di inginkan,dan aspek model yang ingin di amati. Berdasarkan fungsinya prakiraan cuaca untuk saat ini sangat dibutuhkan, seiring dengan semakin banyaknya pihak yang menggunakan dan memanfaatkan informasi tersebut. Oleh karena itu BMKG dituntut untuk dapat memberikan informasi yang cepat, akurat dan tepat. Dengan demikian, perkembangan ilmu pengetahuan dalam upaya menjawab tantangan tersebut menjadi semakin penting. Dibutuhkan metode-metode baru yang dapat mengakomodasi kebutuhan akan informasi cuaca yang semakin kompleks. Berdasarkan permasalahan yang ada dalam penelitian ini penulis akan meneliti dan mencari dan mencoba mencari model yang baik dengan menggunakan metode SARIMA karena model ini lebih mudah diuji ketika data memiliki *Seasonal*. Tujuan yang ingin dicapai dari penelitian ini adalah menentukan tingkat akurasi dari metode SARIMA untuk prediksi curah hujan dan mencoba hal baru dimana apakah pengaruh differencing untuk model SARIMA dapat mempengaruhi tingkat akurasi dari metode SARIMA tersebut dengan data yang digunakan adalah data curah hujan rata-rata bulan Januari 2016- Desember 2019 dan didapatkan model terbaik yaitu SARIMA $\{(2,1,1)(1,1,0)12\}$ dengan nilai MAPE 22,6641, MAD 7,2049 dan nilai MSD 74,1644.

Kata kunci : SARIMA,Curah hujan Rata-rata

PENDAHULUAN

LATAR BELAKANG

Informasi cuaca dan iklim adalah salah satu kebutuhan utama untuk menunjang kegiatan diberbagai bidang, Informasi yang dibutuhkan adalah berupa prakiraan cuaca atau iklim. Untuk mendapatkan hasil yang tepat dan akurat, banyak yang mengembangkan metode prakiraan cuaca. Selama ini banyak yang

menggunakan forecasting menggunakan metode statistik antara lain metode Smoothing,Box Jenkins, ekonometri dan regresi.Pemilihan metode bias didasarkan beberapa aspek,antara lain aspek pola data,aspek waktu,aspek tingkat keakuratan prediksi yang diinginkan ,dan aspek tipe model yang ingin di amati.Dan salah satu metode untuk peramalan cuaca dapat menggunakan Analisis Time Series,Analisis time series sendiri adalah mempelajari

pola Gerakan nilai nilai variabel pada suatu interval waktu(misalnya:hari,minggu,bulan,tahun).Dari Analisis Time series didapatkan data ukuran nilai nilai yang dapat digunakan untuk permalan masa depan.

Landasan Teori

Metode

- *Forecasting* atau peramalan adalah sebuah metode sebagai alat bantu dalam melakukan suatu perencanaan yang efisien dan efektif. Contohnya seperti peramalan tingkat permintaan suatu produk atau beberapa produk serta peramalan terhadap harga sembako di dalam kurun waktu tertentu di masa yang akan datang. (Anitasari, 2017).Metode yang sering digunakan dalam forecasting adalah metode stokastik dan deterministik.
 - Model stokastik adalah model matematik yang gejalanya dapat diukur dengan derajat kepastian yang tidak stabil. Pada model stokastik peluang dari masing-masing kejadian dapat dihitung. Pada umumnya model stokastik umumnya dapat dijumpai pada teori antrian dan teori permainan.
- Perbedaan anatara deterministik dan stokastik

adalah kalua deterministic nilai parameternya yang pasti dan time-invariant , sedangkan kalua stokatik nilai nya tidak pasti dan time-varian (Wibisono, 2014)

- Model Deterministik yaitu model simulasi yang tidak mengandung komponen yang sifatnya probabilistik (random) dan output yang telah dapat ditentukan begitu sejumlah input dan hubungan tertentu dimasukkan. Output yang diperoleh akan tetap sama jika inputnya sama walaupun diproses ulang. Model deterministik memusatkan penelaahannya pada faktor-faktor kritis yang diasumsikan mempunyai nilai eksak tertentu pada waktu yang spesifik (diri, 2015)

Faktor-faktor yang mempengaruhi Faktor-faktor yang mempengaruhi prediksi cuaca perkiraan cuaca dipengaruhi oleh beberapa faktor antara lain:

- Suhu adalah derajat panas dan dingin udara di atmosfer. Yang disebarkan di muka bumi, ada dua suhu udara, yaitu disebarkan secara horisontal dan vertikal. Terdapat banyak faktor yang dapat mempengaruhi suhu udara dipermukaan bumi, misalnya lamanya penyinaran matahari, Dapat menjadi dampak secara langsung yang mempengaruhi perubahan suhu udara, yang dapat diartikan suhu di bumi adalah relative (Dadan, 2018).
- Curah hujan yaitu jumlah air hujan yang turun pada suatu daerah dalam waktu tertentu. Alat untuk mengukur banyaknya curah hujan disebut Rain gauge. Curah hujan diukur dalam harian, bulanan, dan tahunan (Pambudi, 2021).

Time series

Analisis runtun waktu (time series) adalah sebuah metode statistik yang dipakai untuk menemukan pola data dimasa lalu yang dapat digunakan untuk meramalkan pola data dimasa yang akan datang. Salah satu asumsi penting yang harus dipenuhi dalam pemodelan data time series adalah stasioneritas. Suatu data time series dikatakan bersifat stasioner apabila data tersebut memiliki nilai rata-rata yang

konstan dan tidak tergantung pada unsur waktu t , serta memiliki nilai fungsi autokovarian pada waktu t dan s yang hanya bergantung melalui nilai absolut jarak antar kedua waktu tersebut, Secara umum, hal ini menunjukkan bahwa struktur data time series pada proses yang stasioner tidak dipengaruhi oleh waktu. Apabila asumsi stasioneritas belum dipenuhi, maka data time series tersebut perlu untuk distasionerkan dengan cara dilakukan differencing pada datanya. (Anwar, 2017). Analisis yang sering digunakan adalah metode ARIMA dan SARIMA.

SARIMA

SARIMA adalah Suatu *time series* $\{Z_t | t=1, 2, \dots, k\}$ dibangkitkan dari proses SARIMA $(p, d, q) (P, D, Q)_s$ dengan rata-rata μ dari model *time series* Box dan Jenkins jika:

dimana p, d, q, P, D, Q merupakan bilangan bulat dan s merupakan jangka waktu musiman (*periodicity*).

Model SARIMA melibatkan langkah-langkah berikut:

$$\varphi(B)\Phi(B^s)(1-B)^d(1-B^s)^D(Z_t - \mu) = \theta(B)\Theta(B^s)a_t$$

1. Mengidentifikasi struktur SARIMA (p, d, q) (P, D, Q)_s;
2. Mengestimasi parameter yang tidak diketahui;
3. Melakukan uji *Goodness of Fit* terhadap estimasi residual;
4. Melakukan peramalan dari data yang diketahui.

Untuk merubah data menjadi stasioner dalam varian dilakukan transformasi. Metode yang digunakan adalah Box-Cox, dimana data ditransformasi menjadi $Y^* = Y^\lambda$. (Budi, 2010)

Minitab

Minitab adalah salah satu aplikasi yang digunakan untuk mengolah data statistik. Selain minitab aplikasi lainnya yang digunakan dalam mengolah data statistik adalah: SPSS. SAS. StatGraph. Eviews. Statistica. Gambar atau grafik lebih menarik daripada SAS Kelemahan Perbandingan kelebihan dan kelemahan program aplikasi statistik Minitab SPSS SAS Eviews Kelebihan Analisis statistik Mengolah Analisis Analisis dalam bidang data statistik statistik statistik teknik dalam bidang- dalam

dalam bidang sosial bidang bidang teknik dan ekonomi sosial Analisis statistik dalam bidang teknik (Setiabudi, 2017).

Jupyter Notebook

Jupyter organisasi non-profit untuk mengembangkan software interaktif dalam berbagai bahasa

pemrograman. *Notebook* adalah satu software buatan Jupyter, adalah aplikasi web open-source yang memungkinkan Anda membuat dan berbagi dokumen interaktif yang berisi kode *live*, persamaan, visualisasi, dan teks naratif yang kaya (Priyono, 2019).

Differencing

Differencing adalah menghitung perubahan atau selisih nilai observasi. Nilai selisih yang diperoleh dicek lagi apakah *stasioner* atau tidak. Jika belum *stasioner* maka dilakukan *differencing* lagi. Jika varians tidak *stasioner*, maka dilakukan transformasi logaritma (Deden, 2004)

ACF

ACF digunakan untuk melihat apakah ada data yang

bersifat Moving Average (MA) dari suatu deret waktu, yang dalam persamaan ARIMA direpresentasikan oleh besaran q . Jika terdapat sifat MA, q pada umumnya bernilai 1 atau 2 (Dinamika, 2009).

PACF

Autokorelasi parsial (PACF) digunakan untuk mengukur derajat asosiasi antara Y_t dan Y_{t-k} , ketika efek dari rentang atau jangka waktu (time lag) dihilangkan. Seperti ACF, nilai PACF juga berkisar antara +1 dan -1. PACF umumnya digunakan untuk mengidentifikasi ada atau tidaknya sifat Autoregressive (AR), yang dinotasikan dengan besaran p . Jika terdapat sifat AR, pada umumnya PACF bernilai 1 atau 2, jarang ditemukan sifat AR dengan nilai p lebih besar dari 2 (Dinamika, 2009).

MAPE

Mean Absolute Percent Error (MAPE). Besaran ini bermaksud mengukur keakuratan hasil peramalan. Sesuai dengan namanya, MAPE memiliki satuan persentase (pecahan). Kata “mean” sendiri digunakan ketika ada banyak data yang akan dibandingkan dengan data realnya.

Rumusnya adalah sebagai berikut: (Rahmadya, 2020).

$$\left(\frac{1}{n} \sum \frac{|Actual - Forecast|}{|Actual|} \right) * 100$$

MAD

Pengertian Mean Absolute Deviation (MAD) adalah deviasi (simpangan) mutlak secara rata-rata pada sebuah data pusat. Mean Absolute Deviation (MAD) selalu digunakan dalam metode peramalan untuk menghitung tracking signal. sebaran data pada tracking signal digunakan untuk memastikan apakah metode peramalan bisa digunakan atau tidak. Pada pembahasan ini, saya akan menggunakan Mean Absolute Deviation (MAD) pada metode peramalan Moving Averages (KHOIRI, 2020).

MSD

MSD merupakan ukuran penyimpangan ramalan dengan merata-ratakan kuadrat error (penyimpangan semua ramalan). Persamaannya adalah sebagai berikut:

$$MSD = (1/n) \sum (Y(t) - Y'(t))^2$$

Tujuan optimalisasi statistik seringkali dilakukan untuk memilih suatu model agar nilai MSD minimal, tetapi ukuran ini mempunyai dua kelemahan. Pertama ukuran ini menunjukkan pencocokan (*fitting*) suatu model terhadap data historis. Pencocokan seperti ini tidak selalumengimplikasikan peramalan yang baik. Suatu model yang terlalu cocok (*over fitting*) dengan deret data berarti sama dengan memasukkan unsur random sebagai bagian proses bangkitan, adalah sama buruknya dengan dengan tidak berhasil mengenai pola non acak dalam data. Kekurangan kedua dalam MSD sebagai ukuran ketepatan model adalah berhubungan dengan kenyataan bahwa metode berbeda akan menggunakan prosedur yang berbeda pula dalam fase pencocokan (Sungkawa, 2011).

Metodologi Penelitian

Konseptual Model

Data yang digunakan dalam penelitian ini didapatkan dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) Stasiun Geofisika Bandung, dengan periode

Januari 2016 Desember 2019. Variabel cuaca yang dikaji adalah curah hujan rata-rata

Metode dan Analisis Data

Sistem yang dibentuk berupa sistem yang dapat meramalkan nilai curah hujan pada waktu tertentu. Selain sistem melakukan proses peramalan, sistem juga melakukan training dan testing. Berikut langkah-langkah yang digunakan pada tugas akhir peramalan curah hujan menggunakan analisis *timeseries* dengan algoritma SARIMA (*Seasonal Autoregressive Integrated Moving Average*) di Kab.Bandung.

Sumber Data

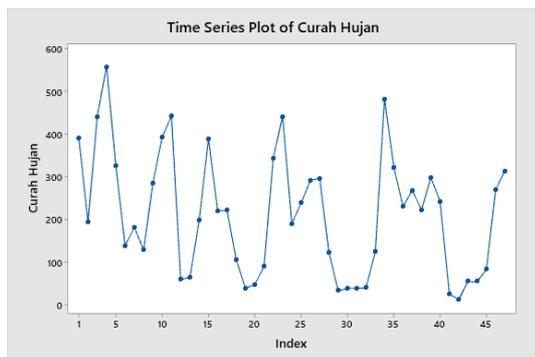
Data sinoptik berasal dari pengamatan stasiun meteorology Geofisika Bandung tahun 2016 sampai dengan 2019 yang berjenis raw data . Data Sinoptik adalah data pengamatan cuaca permukaan yang dikirim dari stasiun-stasiun pengamatan cuaca di seluruh Indonesia setiap tiga jam. Data tersebut meliputi curah hujan rata rata bulanan.

Pemilihan Perangkat Lunak dan Teknik Data Mining

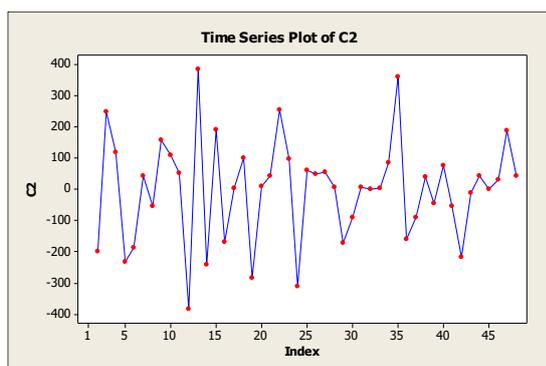
Perangkat lunak yang dipilih adalah Minitab, Perangkat lunak ini dipilih dengan alasan selain gratis, penulis ingin menguji kehandalan perangkat lunak ini dalam melakukan data mining.

Implementasi

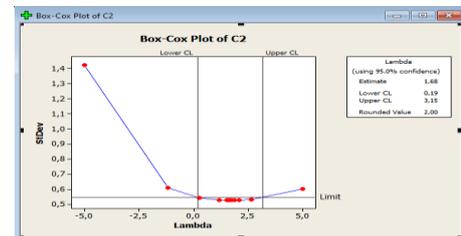
Data yang digunakan diperoleh dari data curah hujan dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) meliputi data rata-rata curah hujan bulanan di Kab. Bandung berikut gambar grafiknya.



merupakan grafik hasil nilai rata-rata Curah Hujan tahun 2016-2019 di Kab. Bandung yang ditampilkan melalui software Minitab dan gambar grafik diatas memiliki pola trend random dan tidak bisa ditentukan.



Gambar diatas adalah data hasil differencing yang dimana data hasil differencing tersebut akan

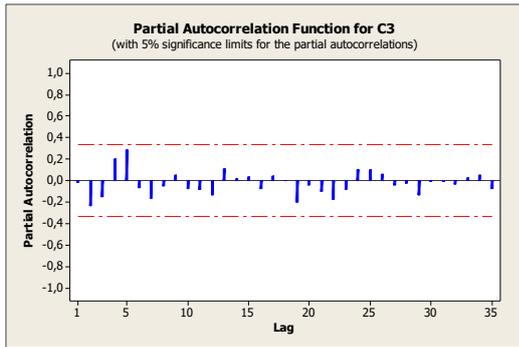


digunakan untuk melakukan analisis ACF dan PACF untuk mencari nilai ACF yaitu AR yang disimbolkan nilai p, dan PACF yaitu MA dan identifikasi apakah data tersebut memiliki Seasonal atau tidak.

Identifikasi Algoritma SARIMA

Berikut adalah Identifikasi dari dimulai *Preprocessing* data dengan melakukan analisis data apakah data sudah stasioner atau belum agar data dapat di olah.

Untuk mengetahui nilai orde model yang digunakan pada tugas akhir ini, yaitu memerlukan ACF dan PACF, untuk nilai periode *seasonal* bisa didapatkan dengan pemberian asumsi. Berikut gambar korelasi ACF dan PACF dapat dilihat dibawah ini:



Sedangkan untuk SARIMA tinggal menambahkan asumsi periode *seasonal* untuk SARIMA misalkan periodenya 6 bulan. Apabila positif maka nilai curah hujan sama nilai prediksi sama-sama baik, sedangkan kalau negatif maka curah hujan baik, dan nilai prediksinya buruk.

Lag	ACF	T
1	-0,288568	-1,98
2	-0,074017	-0,47
3	-0,148580	-0,94
4	-0,002172	-0,01
5	0,098326	0,61
6	-0,073830	-0,45

Analisis Model SARIMA

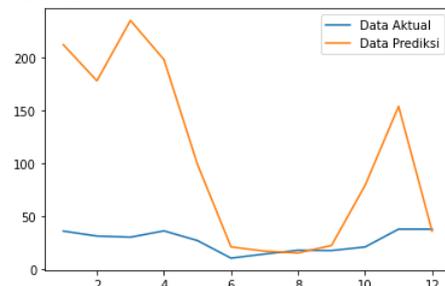
Estimasi parameter pada model SARIMA :

1. Model Algoritma SARIMA $\{(2,1,1)(1,1,0)12\}$ Berdasarkan rumus disamping, dapat diasumsikan dari rumus tersebut:

Dari tabel V.1 Estimasi parameter diatas dapat dilihat nilai P-value nya semuanya kurang dari 0,5 berarti P-value nya signifikan dan model dapat digunakan untuk forecasting.

Type	P
AR 1	0,329
SAR 12	0,000
SAR 24	0,000
MA 1	0,000
Constant	0,000

Pada algoritma SARIMA $\{(2,1,1)(1,1,0)12\}$ Parameter diatas dipergunakan untuk mencari nilai



prediksi Testing. Tahap Selanjutnya adalah Validasi menggunakan MAPE, MAD, dan MSD untuk melihat ke akuratan dari algoritma p,d, dan q

Selanjutnya untuk hasil training data peneliti menggunakan Jupyter Notebook untuk melakukan plotting dari model SARIMA

$\{(2,1,1)(1,1,0)12\}$ untuk Training, dan dapat dilihat dibawah mengenai grafik hasil Training.

menunjukkan perbandingan antara data Aktual Training dengan data prediksi Training menurut algoritma model SARIMA $\{(1,1,1)(2,1,0)12\}$ yang disebutkan garis berwarna biru adalah data actual, dan garis warna kuning adalah data hasil prediksi. Untuk data hasil prediksi bisa dilihat dari awal selalu meleset baik grafik Training maupun Testing karena data tidak memiliki informasi data sebelumnya

MAPE	MAD	MSD
22,6641	7,2049	74,1644

Diatas adalah hasil Validasi dari Aplikasi Mintiab yang dapat digunakan untuk menentukan apakah hasil dari prediksi ramalan cuaca bias dikatakan lebih akurat daripada model yang lain.

Kesimpulan dan Saran

Kesimpulan

- 1) Untuk model runtun waktu Ramalan yang terbaik dilihat dari nilai MSD yang diperoleh. Semakin kecil nilai MSD maka semakin baik ramalan tersebut.

Pada hasil analisis yaitu model algoritma SARIMA $\{(2,1,1)(1,1,0)12\}$.

- 2) Dari hasil Training dan Testing SARIMA diatas dapat saya simpulkan data tingkat keakuratan lebih baik jika data yang digunakan lebih banyak, karena dapat mempengaruhi nilai MAPE, MAD, dan MSD
- 3) Dan dari data BMKG yang saya dapatkan dari tahun 2016-2019 karena dari hasil nilai banyak yang T (periode) negatif yang menandakan hujan nya baik, dan prediksi nya buruk.

Saran

- 1) Untuk melakukan penilitan ini hanya menggunakan algoritma time series bulanan dan kalau bisa untuk kedepan banyak gunakan beberapa data bukan hanya dari BMKG dan di analisis apakah data tersebut dapat menggunakan model SARIMA atau ARIMA, karena jika data NonSeasonal maka lebih tepat untuk menggunakan ARIMA
- 2) Dan Untuk datanya karena hanya menggunakan data per bulan, harusnya agar lebih akurat di anjurkan menggunakan data per hari agar hasilnya lebih optimal.

DAFTAR PUSTAKA

- Anitasari, N. (2017, Desember 23). *Zahir blog*. Retrieved from MENGENAL METODE FORECASTING DALAM SEKTOR INDUSTRI: <https://zahiraccounting.com/id/blog/forecasting-dalam-sektor-industri/>
- Anwar, S. (2017, June). *Peramalan Suhu Udara Jangka Pendek di Kota Banda Aceh dengan Metode Autoregressive Integrated Moving Average (ARIMA)*. Retrieved from Researchgate: <https://www.researchgate.net/>
- Becker, J., Niehaves, B., & Janiesch, C. (2010). Socio-Technical Perspectives on Design Science in IS Research. *Information Systems and eBusiness Management, Vol.9, issue 1*, 109-131.
- Budi. (2010, April 20). *TIME SERIES FORECASTING DENGAN SEASONAL TIME SERIES ARIMA DAN FEED FORWARD NEURAL NETWORK*. Retrieved from Statistika Komputasi: <https://statistikakomputasi.wordpress.com/>
- Dadan. (2018, Februari 11). *Pengertian Suhu Udara (Temperatur Udara)*. Retrieved from Sridianti.com: <https://www.sridianti.com/>
- Deden. (2004). *ARIMA*. Retrieved from SCRIBD: <http://daps.bps.go.id/>
- Dinamika. (2009). *Pengertian ACF dan PACF*. Retrieved from repository dinamika: <http://repository.dinamika.ac.id/>
- diri, K. (2015, April 5). *Model Deterministik*. Retrieved from Teknologi Informasi Indonesia: <http://tiindonesia.blogspot.com>
- ekaningsari, R. (2016, September 12). *Metode ARIMA (Autoregressive Integrated Moving Average)*. Retrieved from ekaningsari.wordpress: <https://ekaningsari.wordpress.com/>
- haniif. (2007, Agustus 1). *Santri's Blog*. Retrieved from Association Rule: <https://haniif.wordpress.com/>
- Hevner, A. R., Ram, S., March, S. T., & Park, J. (2004). Design Science in Information Systems Research. *MIS Quarterly Vol. 28 No. 1*, 75-105.
- Hevner, A., & Chatterjee, S. (2010). *Design Research in Information System : Theory and Practice*. New York: Springer.
- Hildenbrand, T., Rothlauf, R., Geisser, M., Heinzl, A., & Kude, T. (2008). Approach to Collaborative Software Development. *International Conference on Complex, Intelligent and Software Intensive Systems* (pp. 523-528). Barcelona: IEEE.
- Istriana. (2009). *PEMODELAN CURAH HUJAN DENGAN PENDEKATAN ADAPTIVE SPLINE THRESHOLD AUTOREGRESSIVE ASTAR*. Retrieved from DIGITAL LIBRARY.
- KHOIRI. (2020, Oktober 05). *Pengertian dan Cara Menghitung Mean Absolute Deviation (MAD)*. Retrieved from KHOIRI.COM: <https://www.khoiri.com/2020/10/metode-perkiraan-mad-mean-absolute.html>
- Mubarok, M. I. (2018, Agustus 14). *Algoritma C4.5*. Retrieved from MIM: <https://muhammadilhammubarok.wordpress.com/>
- Muhshi, F. A. (2019, Juni 13). *Pengukuran, dan Alat Ukur*. Retrieved from Foresteract: <https://foresteract.com/>
- Mulyana, A. (2007, April 15). *Pengertian Cuaca Dan Iklim Menurut para Ahli*. Retrieved from <https://www.scribd.com/doc/87190140/Pengertia-Cuaca-Dan-Iklim-Menurut-Para-Ahli>

- Mushi, F. A. (2019, Juni 13). *foreverspin*. Retrieved from foresteract: foresteract.com
- nendrabretus. (2015, April 25). *softovator.com*. Retrieved from Eksplorasi Data Mining Menggunakan RapidMiner: <https://www.softovator.com/>
- Pambudi, A. (2021). *Pengertian Curah Hujan dan Klasifikasi Hujan*. Retrieved from Geografi.org: <https://www.geografi.org/>
- Rahmadya. (2020, Juni 28). *Menghitung Mean Absolute Percent Error (MAPE) dengan MATLAB*. Retrieved from Rahmadya Trias Handayanto: <https://rahmadya.com/>
- Setiabudi, V. (2017). *MINITAB. Perbandingan kelebihan dan kelemahan program aplikasi statistik Minitab SPSS SAS Eviews. Analisis statistik dalam*. Retrieved from docplayer: <https://docplayer.info/>
- Suhartono, D. (2012, July 28). *School of Computer science*. Retrieved from Dasar Pemahaman Neural Network: <https://socs.binus.ac.id/2012/07/26/konsep-neural-network/>
- Sungkawa, I. (2011, Desember 02). *PENERAPAN UKURAN KETEPATAN*. Retrieved from media.neliti.com: <https://media.neliti.com>
- Wibisono, T. (2014, 12 22). *Metode Stokastik*. Retrieved from <http://ti3605telkomuniversity.blogspot.com/2014/12/model-stokastik-probabilistik.html>