

# Prediksi Intensitas Hujan Kota Surabaya dengan Matlab menggunakan Teknik Random Forest dan CART (Studi Kasus Kota Surabaya)

Maulana Dhawangkara dan Edwin Riksakomara

Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

*e-mail*: iwangkara95@gmail.com, riksakomara@gmail.com

**Abstrak**— Keakuratan prediksi potensi curah hujan di Kota Surabaya dibutuhkan untukantisipasi bencana akibat hujan seperti banjir bandang, membantu memprediksi kondisi penerbangan dan membantu mahaemen saluran sanitasi di Surabaya. Prediksi dilakukan dengan data hari sebelumnya menggunakan perbandingan teknik Classification and Regression Trees (CART) dan Random Forest (RF) pada data cuaca selama 17 tahun (2000-2016) berasal dari stasiun cuaca Juanda, Surabaya melalui website NCDC (National Climate Data Center) yang terdiri dari data suhu udara, titik embun, kecepatan angin, tekanan udara, visibilitas dan curah hujan. Evaluasi pembuatan model dengan pengukuran akurasi, precision dan recall menunjukkan bahwa baik metode CART maupun Random Forest mampu mengklasifikasi dengan akurasi baik sebesar 78% untuk 4 dari 5 kelas intensitas hujan, dengan kelas terakhir belum mampu diklasifikasi oleh kedua metode. Metode Random forest memiliki nilai performa sedikit lebih baik dibandingkan dengan CART sebesar 6%. Eksperimen tuning parameter untuk kedua metode membuktikan performa lebih baik dibandingkan parameter *default* metode dan mampu memberikan kestabilan hasil performa dari segi uji coba proporsi data training dan testing. Variabel yang berpengaruh besar dalam model CART dan random forest dengan nilai uji performa yang baik antara lain adalah suhu udara, titik embun, suhu udara maksimum dan suhu udara minimum beserta variabel turunannya (selisih suhu udara maksimum dan minimum, selisih suhu udara dan titik embun dan kelembapan relatif). Penelitian ini menghasilkan aplikasi pengklasifikasi intensitas hujan yang memiliki akurasi baik atas kelas intensitas hujan (tidak hujan, ringan, sedang, deras, sangat deras).

**Kata Kunci**— CART, Random Forest, RF, Matlab, Klasifikasi, Curah Hujan, Intensitas Hujan

## I. PENDAHULUAN

**K**ETERSEDIAAN data cuaca dan iklim selama beberapa dekade terakhir yang dikumpulkan melalui observasi, radar maupun satelit membuka peluang besar untuk membuat suatu alat yang akurat dan efektif untuk menganalisis pengetahuan tersebut dalam data tersebut. Penggalan data cuaca dan iklim memiliki tantangan besar dalam metode penggalan pengetahuan dari jumlah data yang besar [1].

Jumlah data yang sangat besar tidak lagi dapat ditangani dengan metodologi tradisional. Para peneliti melihat peluang untuk melahirkan teknologi baru yang menjawab kebutuhan ini, yaitu data mining. Teknologi ini diaplikasikan oleh berbagai perusahaan untuk memecahkan permasalahan bisnis

dengan mengacu pada analisis data, dimana secara garis besar data mining berfungsi mencari pengetahuan dari data dengan jumlah yang sangat besar. Pengetahuan yang didapatkan dari hasil analisis data dan peramalan cuaca dapat berperan besar dalam berbagai sektor yang dipengaruhi cuaca dan iklim secara langsung, seperti agrikultur, pariwisata dan pengelolaan air [2]. Pada sektor transportasi terutama transportasi laut dan udara membutuhkan prakiraan cuaca yang akurat secara periodik dalam hitungan menit untuk memutuskan persetujuan penerbangan dan pelayaran, hal ini dikarenakan kondisi cuaca berkontribusi besar pada keselamatan jiwa penumpang dan awak [3].

Pendekatan data mining dengan menggunakan kombinasi multidimensional association rule mining yaitu algoritma Random Forest dan CART (Classification and Regression Trees) untuk menentukan pola keterkaitan antara curah hujan dengan faktor lainnya. Dalam penelitian [5] dilakukan prediksi kejadian hujan dengan metode klasifikasi menggunakan Decision Trees dan Random Forest yang menunjukkan bahwa performa Random Forest memiliki keunggulan dari segi kemampuan prediktif dibandingkan dengan pohon keputusan tunggal. Sementara itu CART yang merupakan pohon keputusan tunggal cenderung memiliki model kurang stabil, karena perubahan sedikit saja pada data akan berpengaruh pada hasil prediksi dan cenderung overfitting [6]. Sehingga, untuk meningkatkan kestabilan dan kemampuan prediktif, diaplikasikan metode ensemble, salah satu metode ensemble yaitu Random Forest yang memiliki keunggulan untuk menghindari overfitting [7].

Implementasi metode CART dan Random Forest pada data cuaca diharapkan dapat menjadi salah satu metode prediktif untuk meramalkan kejadian hujan dan intensitas hujan. Data yang digunakan berasal dari stasiun pemantauan cuaca dalam jangka waktu 17 tahun (2000-2016) diambil dari stasiun meteorologi di Juanda, Surabaya melalui website NCDC (National Climate Data Center) yang merupakan repositori data cuaca untuk seluruh stasiun meteorologi seluruh negara. Data cuaca yang digunakan meliputi data cuaca yang berkaitan dengan curah hujan seperti suhu udara rata-rata, suhu udara maksimum, titik embun, kecepatan angin rata-rata, kecepatan angin maksimum, kecepatan angin minimum, tekanan permukaan laut, dan visibilitas.

Dengan adanya penelitian studi ini diharapkan mampu

membantu adanya upaya pendeteksian potensi banjir bandang dan membantu manajemen saluran limbah, manajemen pintu air, membantu proyek pembangunan dengan memberikan prediksi potensi curah hujan pada kota Surabaya. Pengembangan perangkat lunak mengimplementasikan teknik random forest dan CART untuk memprediksi kejadian hujan dan intensitas curah hujan diharapkan mampu membantu pemerintah kota Surabaya dalam mempercepat ketanggapan proses antisipasi banjir dan manajemen saluran limbah.

II. TINJAUAN PUSTAKA

A. Data Cuaca Surabaya

Pada penelitian ini, dataset yang digunakan adalah data cuaca kota Surabaya dengan 9 atribut seperti pada Tabel 1. Dataset dipilih dari repositori data cuaca NCDC (National Climatic Data Center) yang merupakan organisasi yang menyediakan rekam data cuaca secara global untuk seluruh stasiun meteorologi. Dataset sendiri merupakan data harian ringkasan cuaca pada kota Surabaya yang direkam oleh stasiun meteorologi Juanda, Surabaya.

Berikut ini adalah klasifikasi intensitas hujan menurut Badan Meteorologi, Klimatologi dan Geofisika [2]:

1. Hujan ringan dengan intensitas : 0,1 - 5,0 mm/jam atau 5 - 20 mm/hari
2. Hujan sedang dengan intensitas : 5,0 - 10,0 mm/jam atau 20 - 50 mm/hari
3. Hujan lebat dengan intensitas : 10,0 - 20 mm/jam atau 50 - 100 mm/hari
4. Hujan sangat lebat dengan intensitas : >20 mm/jam atau >100 mm/hari

Tabel 1. Variabel Dataset

| No | Variabel           | Deskripsi  |
|----|--------------------|--|
| 1  | Temperature        | Rata-rata suhu udara dalam derajat Celcius   |
| 2  | Dew Point          | Suhu titik embun dalam derajat Celcius, dimana uap air mulai mengembun dari udara                          |
| 3  | Sea Level Pressure | Tekanan permukaan laut dalam milibar   |
| 4  | Visibility         | Ukuran tingkat jarak pandang dalam km  |
| 5  | Wind Speed         | Rata-rata kecepatan angin dalam knots  |
| 6  | Max Wind Speed     | Kecepatan angin maksimum yang dicatat dalam satu hari dalam knots  |
| 7  | Max Temperature    | Suhu udara maksimum dalam satu hari dalam derajat Celcius  |
| 8  | Min Temperature    | Suhu udara minimum dalam satu hari dalam derajat Celcius   |
| 9  | Precipitation      | Total curah hujan dalam mm yang akan didiskritisasi sesuai klasifikasi intensitas hujan per hari dari BMKG |

B. Persiapan Data

Tahap persiapan data merupakan tahap penting dalam proses data mining. Frase “Garbage In, Garbage Out” dapat diaplikasikan pada proses data mining, dengan kata lain apabila data buruk masuk diolah maka akan menghasilkan

hasil yang buruk juga. Sehingga, untuk meningkatkan kualitas data dan hasil data mining, tahap pemrosesan data sangat penting. Tahapan persiapan data, antara lain:

1) Penghapusan variabel dependen (complete case)

Metode *complete case* secara implisit mengasumsikan bahwa data hilang secara acak yaitu, tidak terkait dengan data yang hilang dan diamati dari dataset dan bila asumsi ini tidak terpenuhi, metode ini akan menghasilkan estimasi parameter bias [3].

2) Penggantian nilai missing variabel independen menggunakan Multiple Imputation

Metode Imputasi adalah pengisian nilai hilang pada suatu dataset dengan menempatkan kembali nilai hilang dengan nilai tunggal dan kemudian memperlakukannya sebagaimana nilai tersebut merupakan nilai sebenarnya.

3) Transformasi Data

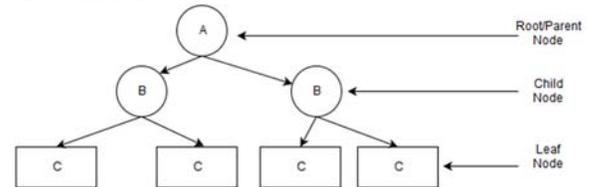
Diskritisasi curah hujan menjadi kelas intensitas hujan menurut BMKG serta konversi dari satuan USA menjadi Indonesia.

C. Metode Data Mining

1) CART

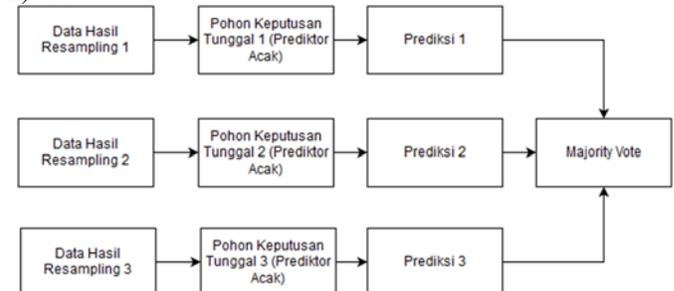
CART (*Classification and Regression Trees*) merupakan algoritma dari teknik eksplorasi data yaitu teknik *Decision Tree* (pohon keputusan). CART menghasilkan *Classification Tree* (pohon klasifikasi) untuk nilai respon kategorik, dan menghasilkan *Regression Tree* (pohon regresi) untuk nilai responnya kontinyu.

Struktur pohon keputusan pada CART adalah seperti Gambar 1. Simpul paling atas disebut simpul akar (root node), kemudian root node dibagi lagi menjadi parent node (simpul induk) yang terbagi menjadi beberapa child nodes (simpul anak) dan terakhir adalah Leaf node (simpul daun) dimana nilai kelas melekat.



Gambar 1. Struktur Simpul Pohon Keputusan

2) Random Forest



Gambar 2. Algoritma sederhana Random Forest

Random Forest (RF) adalah metode klasifikasi dan regresi berdasarkan agregasi sejumlah pohon keputusan, seperti pada Gambar 2. Secara khusus, merupakan sebuah *ensemble* (kumpulan) dari pohon yang dibangun dari kumpulan data pelatihan dan yang divalidasi secara internal untuk menghasilkan nilai respon yang dari nilai prediktor [4].

**D. Teknologi Perangkat Lunak**

Aplikasi matlab memfasilitasi konstruksi GUI menggunakan GUIDE serta tools yang mengakomodasi metode teknik data mining yang penulis gunakan pada studi ini. *Tools* Fitcree digunakan untuk pengolahan data dengan teknik CART [5], sedangkan *tools* Treebagger untuk teknik Random Forest [6].

**E. Pengukuran Performa**

Precision berfungsi untuk mengukur ketepatan pengklasifikasi, ditunjukkan pada persamaan (1) sedangkan recall untuk mengukur kelengkapan pengklasifikasi, ditunjukkan pada persamaan (2). Akurasi berfungsi untuk mengukur keseluruhan performa pengklasifikasi, ditunjukkan pada persamaan (3).

Pada Tabel 2, misalnya untuk klasifikasi kelas 0, maka TP adalah pada pemetaan kelas aktual 0 ke kelas prediksi 0. FP adalah pemetaan kelas aktual non 0 ke kelas prediksi 0, TN adalah pemetaan kelas aktual non 0 yang benar diprediksikan sedangkan FN adalah pemetaan kelas aktual non 0 ke kelas prediksi non 0 yang salah.

Tabel 2. Confusion matrix multi kelas

| Confusion Matrix | Prediksi |    |    |     |    |  |
|------------------|----------|----|----|-----|----|--|
|                  | 0        | 1  | 2  | ... | j  |  |
| 0                | TP       | FN | FN | FN  | FN |  |
| 1                | FP       | TN | FN | FN  | FN |  |
| Aktual 2         | FP       | FN | TN | FN  | FN |  |
| :                | FP       | FN | FN | TN  | FN |  |
| j                | FP       | FN | FN | FN  | TN |  |

$$Recall (TP Rate) = \frac{TP}{TP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{3}$$

**III. METODOLOGI**

Pada bagian ini akan dijelaskan tahapan dilakukan dalam pengerjaan studi pada Tabel 3.

**IV. ANALISIS DAN PEMBAHASAN**

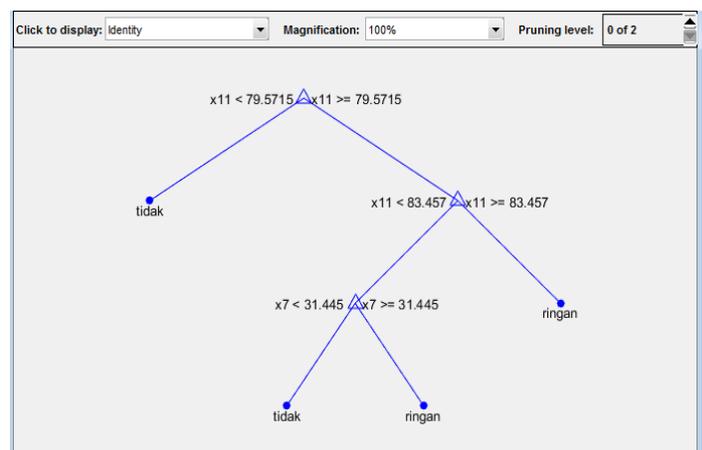
**A. Metode CART**

1. Variabel yang dipilih untuk dilibatkan dalam pembuatan model adalah 12 variabel (8 variabel dataser, 4 variabel turunan) karena memiliki nilai uji performa yang baik dari segi akurasi dan dari segi presisi dan recall. Model dengan 12 variabel ini menggunakan parameter default minimal leaf sejumlah 10 dan minimal parent size 1 dan proporsi training/testing 70%/30%, dengan akurasi training yaitu 0.9321, akurasi testing yaitu 0.7286, nilai weighted accuracy 0.3216 serta mampu mengklasifikasi kelas 0,1,2,3 jika dilihat dari precision dan recall, namun menghasilkan decision tree 31 level.
2. Pemangkasan CART maksimal dengan Model 12 variabel menghasilkan 2 level decision tree, menyisakan 2 variabel yaitu x11-RELH (estimasi kelembapan relatif)

, x7 (Suhu udara maksimum) dengan nilai akurasi training yaitu 0.8067, untuk nilai akurasi testing yaitu 0.7877, nilai weighted accuracy yaitu 0.2937 serta hanya mampu memprediksi kelas 0,1.

Tabel 3. Metodologi Penelitian

| INPUT                                    | AKTIVITAS                               | OUTPUT  |
|--|---|---|
| -Jurnal Ilmiah<br>-Tugas Akhir/<br>Tesis | <b>Studi Literatur</b>                  | -Metode Data Mining yang cocok untuk jenis data<br>-Metode pengukuran performa                        |
| -Dataset cuaca kota Surabaya             | <b>Persiapan Data</b>                   | -Dataset cuaca ditransformasikan sesuai kebutuhan penelitian  |
| -Dataset siap diolah                     | <b>Klasifikasi dengan Teknik CART</b>   | -Hasil klasifikasi dengan CART : Pohon klasifikasi maksimal dan optimal<br>-Hasil pengukuran performa |
| -Dataset siap diolah                     | <b>Klasifikasi dengan Random Forest</b> | -Hasil klasifikasi dengan RF, hasil tuning parameter<br>-Hasil pengukuran performa                    |
| Model klasifikasi Random Forest, CART    | <b>Konstruksi Aplikasi</b>              | Konstruksi Aplikasi dengan Matlab dengan tools fitcree dan treebagger                                 |
| Produk tugas akhir                       | <b>Pembahasan dan Dokumentasi</b>       | Laporan tugas akhir   |



Gambar 3. Pohon keputusan hasil CART dengan pemangkasan

Dari Gambar menunjukkan hasil pohon keputusan setelah dipangkas menjadi 2 level, dapat diketahui beberapa aturan untuk klasifikasi kelas hujan, antara lain: x11-RELH (estimasi kelembapan relatif), x7 (Suhu udara maksimum)

1. Jika RELH kurang dari 79.5715 maka tidak akan terjadi hujan
2. Jika RELH lebih dari sama dengan 83.457 maka akan terjadi hujan ringan
3. Jika RELH berada di antara 79.5715 dan 83.457, serta Suhu udara maksimum kurang dari 31.445, maka tidak akan terjadi hujan
4. Jika RELH berada di antara 79.5715 dan 83.457, serta Suhu udara maksimum lebih dari sama dengan 31.445, maka akan terjadi hujan ringan

Pengaturan kedalaman pohon, pada Tab adalah hasil sesuai hasil masing-masing uji performa, dibagi menjadi 2 bagian yaitu hasil yang stabil dan yang tinggi:

Tabel 4. Nilai parameter berdasarkan uji performa pengendalian kedalaman pohon model CART

|               | Train  | Test  | WaccTest | Presisi (memprediksi 0,1,2,3) | Recall (memprediksi 0,1,2,3) |
|---------------|--------|-------|----------|-------------------------------|------------------------------|
| <b>Stabil</b> |        |       |          |                               |                              |
| <b>Leaf</b>   | 20-100 | 30-60 | 50-90    |                               |                              |
| <b>Parent</b> | 50-90  | 50-90 | All      |                               |                              |
| <b>Tinggi</b> |        |       |          |                               |                              |
| <b>Leaf</b>   | 1-5    | 30-60 | 1-5      | 3-5                           | 1-5                          |
| <b>Parent</b> | 1-10   | 50-90 | 10       | 1-50                          | 1-50                         |

Ditarik kesimpulan, model yang paling baik yang dapat memprediksi 0,1,2,3 adalah dengan leaf (1-5) dan parent (30).

1. Proporsi training dan testing untuk model CART dengan leaf 3 dan parent 30 untuk proporsi 60:40, 70:30, 75:25, 80:20, 85:15 dan 90:10 cenderung stabil, untuk data training berkisar antara 0,85-0,86 dan untuk data testing berkisar antara 0,76-0,775 serta nilai Weighted Accuracy, cenderung stabil berkisar antara 0,30-0,32.
2. Melakukan pengujian kestabilan proporsi data training/testing untuk model CART dengan leaf 3 dan parent 30 menunjukkan proporsi 60% data training dan 40% data testing dengan nilai akurasi training 0,858, akurasi testing 0,775 dan weighted accuracy testing 0,318.

**B. Metode Random Forest**

1. Variabel yang dipilih untuk dilibatkan dalam pembuatan model adalah model II dengan pemilihan atribut (5 variabel: suhu udara, suhu udara maksimum, selisih suhu udara dan titik embun, selisih suhu udara maksimum dan minimum serta estimasi kelembapan relatif) karena memiliki nilai uji performa yang baik dari segi akurasi dan dari segi presisi dan recall. Model ini menggunakan parameter default jumlah tree berkisar 50-250 dan jumlah variabel acak sejumlah 2 dan proporsi training/testing 70%/30%.
2. Model II dengan pemilihan atribut (5 variabel) memiliki akurasi training yaitu 0,99, untuk nilai akurasi testing paling tinggi yaitu 0,77, nilai weighted accuracy paling tinggi yaitu 0,31-0,32 serta mampu memprediksi kelas 0,1,2 dan 3 jika dilihat dari precision dan recall.
3. Pengendalian jumlah pohon dan jumlah variabel acak: berikut ini adalah hasil akurasi sesuai hasil masing-

masing uji performa, dibagi menjadi 2 bagian yaitu hasil yang stabil dan yang tinggi, seperti dapat dilihat pada Tab:

Tabel 5. Nilai parameter berdasarkan uji performa pengendalian jumlah pohon dan jumlah variabel model random forest

|                            | Train  | Test   | WaccTest | Presisi (memprediksi 0,1,2,3) | Recall (memprediksi 0,1,2,3) |
|----------------------------|--------|--------|----------|-------------------------------|------------------------------|
| <b>Stabil</b>              |        |        |          |                               |                              |
| <b>Variabel Acak Pohon</b> | All    | 10-12  | 7-10     | All                           | All                          |
|                            | 50-250 | 50-250 | 20-250   | 10-20                         | 10-20                        |
| <b>Tinggi</b>              |        |        |          |                               |                              |
| <b>Variabel Acak Pohon</b> | All    | 2      | 6        |                               |                              |
|                            | 50-250 | 50-250 | 10       |                               |                              |

Ditarik kesimpulan, model yang paling baik yang dapat memprediksi 0,1,2,3 adalah dengan jumlah variabel acak 6 dan jumlah pohon (10).

1. Proporsi training dan testing untuk model Random Forest dengan jumlah variabel acak 6 dan jumlah pohon 10 untuk proporsi 60:40, 70:30, 75:25, 80:20, 85:15 dan 90:10 cenderung stabil, untuk data training berkisar antara 0,98-0,99 dan untuk data testing berkisar antara 0,76-0,78 serta nilai Weighted Accuracy, cenderung stabil berkisar antara 0,29-0,33.
  2. Melakukan pengujian kestabilan proporsi data training/testing untuk model Random Forest dengan jumlah variabel acak 6 dan jumlah pohon 10 menunjukkan proporsi 70% data training dan 30% data testing memiliki performa paling baik dengan nilai akurasi training 0,991, akurasi testing 0,788 dan weighted accuracy testing 0,330.
- C. Analisis Perbandingan CART dan Random Forest**
1. Menggunakan parameter default, terdapat perbedaan 6% akurasi antara CART dan Random Forest, untuk CART berkisar 0,72-0,73 sedangkan Random Forest adalah 0,78. Parameter default yang dimaksud adalah yang menurut Breiman [17], untuk CART dengan minimal leaf node 10 dan minimal parent 1, sedangkan untuk Random Forest jumlah variabel acak akar jumlah variabel.
  2. Pada CART, variabel yang memiliki kontribusi besar antara lain index 1,2,9 dan 11 yaitu suhu udara, titik embun, selisih suhu udara dan titik embun serta selisih suhu udara minimum dan maksimum. Pada Random Forest variabel yang berkontribusi besar antara lain index 1,2,7,9,10,11 yaitu suhu udara, titik embun, suhu udara maksimum, selisih suhu udara dan titik embun, kelembapan relatif dan selisih suhu udara maksimum dan minimum. Dari kedua metode, dapat ditarik kesimpulan bahwa suhu udara, titik embun, suhu udara maksimum dan suhu udara minimum beserta variabel turunannya (selisih suhu udara maksimum dan minimum, selisih suhu udara dan titik embun dan kelembapan relatif) yang berpengaruh besar terhadap pembangunan model, dengan tingkat akurasi mencapai 0,78.
  3. Baik CART, maupun Random forest memiliki kestabilan dari segi nilai akurasi untuk beberapa kombinasi proporsi

data training dan data testing, yaitu 60:40, 70:30, 75:25, 80:20, 85:15 dan 90:10

D. Konstruksi Aplikasi

Perancangan perangkat lunak untuk aplikasi prediksi intensitas curah hujan kota Surabaya, terdapat 7 use case utama, antara lain:

- a. Memasukkan Data
- b. Memrediksi Intensitas Hujan dengan Metode CART Maksimal
- c. Memrediksi Data dengan Metode CART Pruned
- d. Melihat Variable Importance CART
- e. Melihat Grafik Tree untuk CART
- f. Memrediksi Data dengan Metode Random Forest
- g. Melihat Variable Importance Random Forest

Pada Gambar 3, Gambar 4, Gambar 5 pada lampiran dapat dilihat tampilan aplikasi berupa halaman memasukkan data, halaman pengolahan data dengan CART dan halaman pengolahan data dengan random forest.

V. KESIMPULAN DAN SARAN

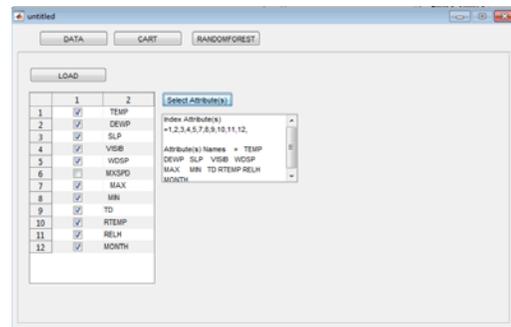
A. Kesimpulan

1. Model menggunakan parameter default Breiman, untuk metode CART dan Random Forest untuk dataset digunakan, berdasarkan hasil performa menunjukkan model Random Forest lebih baik di atas CART sebesar 6%.
2. Data dengan tingkat ketidakseimbangan tinggi antar kelas,, baik model CART maupun dan Random Forest dengan tuning parameter optimal memiliki kemampuan cukup baik yaitu mampu menghasilkan nilai akurasi 77%-78% dan mampu mengklasifikasi 4 kelas (0,1,2,3) dari 5 kelas (0,1,2,3,4). Kelas 4 (kelas hujan sangat lebat) tidak mampu diklasifikasi oleh kedua metode meskipun dilakukan tuning parameter untuk mencapai model terbaik. Performa model dari segi presisi dan recall menunjukkan untuk model dengan tuning parameter memiliki performa akurasi paling baik pada kelas 0 (80%-90%), kelas 1 (40%-50%), kelas 2 dan 3 (10%-20%) sedangkan kelas 4 sebesar 0%.
3. Model CART maupun dan Random Forest dengan tuning parameter yang optimal memiliki kestabilan nilai performa untuk proporsi data training/testing sebesar 60:40, 70:30, 75:25, 80:20, 85:15 dan 90:10.
4. Variabel berpengaruh dalam model CART dan random forest dengan nilai performa baik berdasarkan *variable importance* yaitu suhu udara, titik embun, suhu udara maksimum dan suhu udara minimum dan variabel turunannya (selisih suhu udara maksimum dan minimum, selisih suhu udara dan titik embun dan kelembapan relatif).

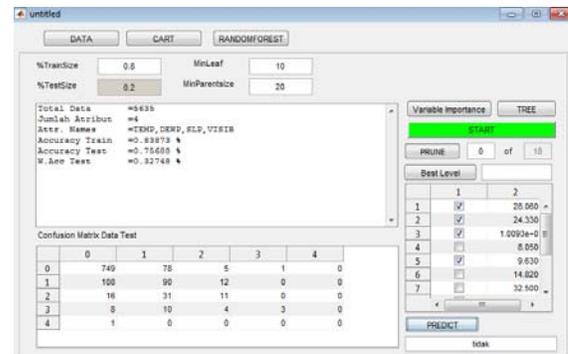
B. Saran

1. Analisis pencilaan dapat dilakukan untuk melihat karakteristik dan mengklasifikasi kelas 4 (hujan sangat lebat) karena metode CART dan random forest masih belum mampu mengklasifikasi kelas 4 tersebut.
2. Menggunakan data cuaca dari 2 stasiun cuaca lain di Surabaya, yaitu stasiun Perak dan stasiun Gedangan.
3. Menambahkan variabel baru seperti kelembapan rata-rata, kelembapan maksimum dan kelembapan minimum.

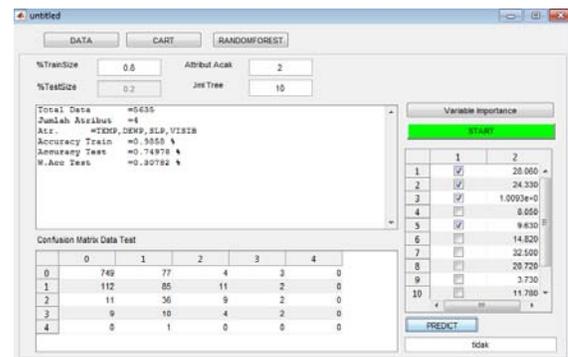
VI. LAMPIRAN



Gambar 3. Halaman fungsi memasukkan data



Gambar 4. Halaman fungsi metode CART



Gambar 5. Halaman fungsi metode Random Forest

DAFTAR PUSTAKA

- [1] & Shereef K Bab'oo S., "Applicability of Data Mining Techniques for Climate Prediction – A Survey Approach," *International Journal of Computer Science and Information Security*, vol. 8, April 2010.
- [2] Klimatologi dan Geofisika Badan Meteorologi. (2016, September) [Online]. <http://meteo.bmkg.go.id/prakiraan/mingguan>
- [3] Donald B. Rubin, "Inference and missing data," *Oxford Hournal Biometrika*, 1975.
- [4] Silke Janitzka, Jochen Kr uppa, Inke R. König Anne-Laure Boulesteix, "Over view of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics," *Technical Report Number 129, Department of Statistics, University of Munich*, 2012.
- [5] Mathworks. (2016) Mathworks Fitctree. [Online]. <http://www.mathworks.com/help/stats/fitctree.html>
- [6] Mathworks. (2016) Mathworks Treebagger. [Online]. <http://www.mathworks.com/help/stats/treebagger.html>
- [7] Friedman JH, Olshen RA, Stone CJ Breiman L, *Classification and Regression Trees*. New York: Chapman & Hall, 1993.
- [8] Wu W., Huang Y Zhang Z., "Mining dynamic interdimension association

- rules for local-scale weather prediction," *In the Proceedings of the 28th Annual International Computer Software and Applications Conference*, 2004.
- [9] Kementerian Perhubungan RI, "Peraturan Menteri Perhubungan Republik Indonesia no 9 tahun 2015 tentang Peraturan Keselamatan Penerbangan Sipil Bagian 174 Tentang Pelayanan Informasi Meteorologi Penerbangan (Aeronautical Meteorological Information Services," *Peraturan Menteri Perhubungan Republik Indonesia no 9 tahun 2015 tentang Peraturan Keselamatan Penerbangan Sipil Bagian 174 (ivil Aviation Safety Regulations Part 174) Tentanf*, 2015.
- [10] Tuan Zea Tan, "Rainfall intensity prediction by a spatial-temporal ensemble," *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008.
- [11] T R Prajwala, "A Comparative Study on Decision Tree and Random Forest Using R Tool," vol. 4, no. 1, 2015.
- [12] C. D. Sutton, "Classification and Regression Trees, Bagging, and Boosting," *Handbook of Statistics*, 2005.
- [13] D. A. Ahijevych, C. J. Kessinger, T. R. Saxen, M. Steiner and S. Dettling John K. Williams, "A Machine Learning Approach to Finding Weather Regimes and Skillful Predictor Combinations for Short-term Storm Forecasting," *National Center for Atmospheric Research*, 2008.
- [14] Elia Georgiana Petre, "A Decision Tree for Weather Prediction," *Buletinul UniversităŃii Petrol – Gaze din Ploiești*, 2009.
- [15] Institute Indian Agricultural Statistics Research, *Data Preprocessing Techniques for Data Mining*.: Institute Indian Agricultural Statistics Research, 2007.
- [16] Satoshi Usami, Ross Jacobucci, and John J. McArdle Timothy Hayes, "Using Classification and Regression Trees (CART) and Random Forests to Analyze Attrition: Results From Two Simulations," *Psychol Aging*, 2015.
- [17] R. Kahavi, and M. Sahami J. Dougherty, "Supervised and unsupervised discretization of continuous features," *In Machine Learning: Proceedings of the Twelfth International Conference*, 1995.
- [18] Amir Ahmad, "Data Transformation For Decision Tree Ensembles," *Thesis for Degree of Doctor of Phylosophy University Of Manchester*, 2009.
- [19] Michael J.A dan Linoff, Gordon S Berry, *Data Mining Techniques For Marketing, Sales, Customer Relationship Management Second Editon*. United States of America: Wiley Publishing, Inc , 2004.
- [20] Sneha Soni, "Implementation of Multivariate Data Set By Cart Algorithm," *International Journal of Information Technology and Knowledge Management*, 2010.
- [21] Friedman JH, Olshen RA, Stone CJ Brieman L, *Classification and Regression Trees*. New York: Chapman & Hall, 1984.
- [22] Mohammed Zakariah, "Classification of genome data using Random Forest Algorithm: Review," *Interational Journal Computer Technology and Application (IJCTA)*, 2014.
- [23] E. C. Polley, and F. B. S. Briggs. B. A. Goldstein, "Random forests for genetic association studies," *Statistical Applications in Genetics and Molecular Biology*, 2011.
- [24] L. C. Thomas, "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers.," *International Journal of Forecasting*, vol. 16(2), 2000.
- [25] O. A., and R. E. Eskridge Alduchov, "Improved Magnus' form approximation of saturation vapor pressure," *J. Appl. Meteor*, pp. 601-609, 1996.