

Prediksi Curah Hujan Dari Data Satelit Himawari-8 Menggunakan Metode *Random Forest*

1st M. Imam Whidyarto
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia

mimamw@student.telkomuniversity.ac.id

2nd Casi Setianingsih
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia

setiacasie@student.telkomuniversity.ac.id

3rd Wendi Harjupa
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia

wendihjr@telkomuniversity.ac.id

4th Risyanto
Research Center for Climate and
Atmosphere
National Research and Innovation
Agency
Bandung, Indonesia
risyanto@brin.go.id

Abstrak— Cuaca merupakan keadaan darurat udara saat tertentu dan wilayah tertentu relatif sempit dan jangka waktu yang singkat. Perkiraan cuaca pada daerah Indonesia tidak menentu, biasanya akan terjadi perubahan setiap sepekan. Untuk mengestimasi dan memprediksi cuaca dan ketinggian awan pada wilayah Indonesia dengan akurat diperlukan teknologi cuaca dan algoritma pemrograman. Maka digunakanlah Machine Learning (ML) dengan menggunakan metode *random forest* (RF) agar didapatkan hasil akurasi yang baik serta akurat, untuk ketinggian awan yang mempengaruhi cuaca dan curah hujan. Pada penelitian Tugas Akhir ini dibahas prediksi ketinggian awan dan hujan menggunakan data pada satelit Himawari-8 dengan menerapkan metode RF, setelah hasil prediksi yang diperoleh maka dilakukan perbandingan untuk mengetahui kesesuaian yang ada untuk mengetahui dan mendapatkan hasil yang optimal. Prediksi curah hujan menggunakan data netCDF untuk memprediksi ketinggian awan dan hujan menggunakan metode *Random Forest* (RF) dengan titik acuan, yaitu kilometer (km). Pada pengujian data sebanyak 120 data yang dilakukan didapatkan akurasi tertinggi ialah 100%.

Kata kunci— Kata kunci sedapat mungkin menjelaskan isi tulisan, dan ditulis dengan huruf kecil, kecuali akronim. Kata kunci tidak lebih dari 6 kata

I. PENDAHULUAN

Cuaca merupakan keadaan udara saat tertentu dan wilayah tertentu relative sempit dan jangka waktu singkat. Cuaca terbentuk dari gabungan unsur-unsur cuaca yang hanya beberapa jam saja. Misalnya keadaan udara pada pagi hari dapat berubah siang, sore hari dan malam hari. Iklim adalah keadaan cuaca dan iklim merupakan kombinasi dari

variabel-variabel atmosfer yang sama disebut unsur-unsur iklim [1]. Perkiraan cuaca pada daerah Indonesia tidak menentu, biasanya akan terjadi perubahan setiap sepekan. Perubahan iklim tidak menentu menyebabkan sulitnya memprediksi cuaca, cuaca ekstrem dengan kejadian luar biasa, berpotensi untuk menimbulkan bencana alam yang disebabkan oleh alam.

Untuk mengestimasi dan memprediksi cuaca, suhu dan ketinggian awan pada wilayah Indonesia dengan akurat diperlukan teknologi cuaca, suhu, dan algoritma pemrograman. Perkembangan yang menarik dan relative baru adalah penggunaan *Machine learning* (ML) pada ilmu alam, tujuan utamanya adalah untuk mendapatkan wawasan dan penemuan ilmiah baru dari data pengamatan. Keberhasilan ini telah menyebabkan ML cukup besar di banyak bidang ilmiah [2]. ML dipakai untuk memprediksi dan mengolah data, yang didapatkan pada masa lalu hingga saat ini, dengan data dan input yang diberikan pada computer akan diproses menggunakan metode *Random Forest* (RF), RF adalah kombinasi pohon director sedemikian rupa sehingga setiap pohon tergantung pada nilai vector acak yang disampel secara independent dan dengan distribusi yang sama untuk semua pohon [3].

Pada penelitian yang dilaksanakan untuk mengetahui curah hujan, ketinggian awan dari data satelit Himawari-8, satelit ini generasi baru satelit meteorologi geostasioner Jepang yang mempunyai sensor resolusi radiometric, dan spasial yang jauh lebih tinggi daripada yang sebelumnya tersedia di orbit geostasioner. Cara untuk mengetahui prediksi curah hujan ialah memanfaatkan data satelit Himawari-8 agar memprediksi curah hujan yang terjadi dengan algoritma RF.

Untuk performa dalam prediksi curah hujan dengan RF dengan menambah banyaknya *tree* agar menghasilkan akurasi terbaik.

II. KAJIAN TEORI

A. HIMAWARI-8/9

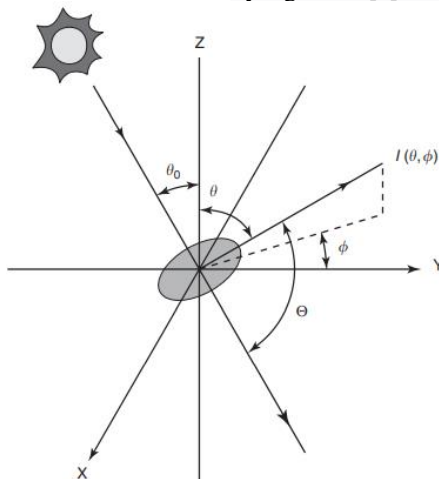
Himawari-8 adalah generasi terbaru satelit meteorologi stasioner Jepang, yang mempunyai optik canggih dengan sensor resolusi radiometric, spectral dan spasial yang jauh lebih tinggi daripada yang sebelumnya ada pada orbit geostasioner, yang memiliki 16 pita pengamatan dan resolusi spasialnya adalah 0,5 atau 1 km untuk terlihat sedangkan pita inframerah 2 km [4].

Selain itu data curah hujan didapatkan melalui satelit Himawari-8, yang dapat **mengurangi** hidrometeorologi bencana, satelit Himawari-8 dapat juga digunakan sebagai estimasi curah hujan untuk berbagai daerah [5].

1. Albedo

Albedo merupakan besaran yang menggambarkan perbandingan antara sinar Matahari yang tiba di permukaan bumi dan yang dipantulkan ke angkasa dengan terjadi perubahan panjang gelombang (*outgoing longwave radiation*). Permukaan albedo merupakan bagian dari sinar matahari yang datang. Radiasi yang tidak dipantulkan akan diserap oleh permukaan. Permukaan albedo adalah komponen kunci dalam penginderaan jarak jauh dan atmosfer di ruang angkasa.

Permukaan yang berbeda memiliki albedo yang berbeda juga, seperti lautan danau dan hutan yang memantulkan sinar matahari yang masuk relatif lebih sedikit dan memiliki albedo rendah. Salju, es dan gurun relatif memantulkan sebagian besar sinar matahari yang masuk dan memiliki albedo yang besar [6].



GAMBAR 1

(Orientasi datangnya sinar matahari dan sinar pantul, θ_0 adalah sudut zenith matahari, θ sudut pandang zenith ϕ relatif sudut azimuth dan θ sudut hamburan. Sinar matahari datang adalah berada di bidang YZ [6].)

2. Albedo 01 -03 (Visible)

Pada kanal 01 dan 02 memiliki resolusi spasial 1 km, dan kanal 03 dengan resolusi spasial 0,5 km. Kanal 01 memiliki Panjang gelombang 0,47 μm , kanal 02 dengan Panjang gelombang 0,51 μm dan pada kanal 03 memiliki Panjang gelombang 0,64 μm . Properti fisik yang dimiliki

kanal 01 dan 03 bisa melihat tumbuhan, aerosol, fase awan dan ukuran partikel [7].

3. Albedo 04 -06 (Near-Infrared)

Pada kanal 04 mempunyai resolusi spasial 1 km, dan pada kanal 05 dan 06 mempunyai ukuran resolusi spasial 2 km. kemudian untuk kanal 04 memiliki Panjang gelombang 0,86 μm , kanal 05 dengan panjang gelombang 1,6 μm dan pada kanal 06 memiliki panjang gelombang 2,3 μm . Properti fisik yang dimiliki pada kanal 04 sampai 06 bisa melihat tumbuhan, aerosol, fase awan dan ukuran partikel [7].

4. Brightness Temperature Of Band 07 – 16 (Infrared)

Brightness temperature (Suhu kecerahan) adalah suhu yang mempunyai kesetimbangan termal untuk menduplikasi intensitas terukur dari benda abu-abu yang dimatan pada frekuensi elektromagnetik tertentu.

Pada setiap kanal Infrared (IR) memiliki resolusi spasial, seperti pada kanal 07 memiliki ukuran 3,9 μm , kanal 08 memiliki ukuran 6,2 μm , kanal 09 memiliki ukuran 6,9 μm , kanal 10 memiliki ukuran 7,3 μm , kanal 11 memiliki ukuran 8,6 μm , kanal 12 memiliki ukuran 9,6 μm , pada kanal 13 memiliki ukuran 10,4 μm pada kanal 14 memiliki ukuran 11,2 μm , pada kanal 15 memiliki ukuran 12,4 μm dan kanal 16 memiliki 13,3 μm [8].

B. Machine Learning

Machine learning (ML) adalah cabang dari kecerdasan buatan (AI) dan ilmu komputer yang berfokus pada penggunaan data dan algoritma untuk meniru cara belajarnya manusia. ML merupakan komponen yang penting pada bidang ilmu data yang berkembang. Algoritma ML dilatih untuk membuat klasifikasi ataupun prediksi.

Saat ini ML digunakan secara luas pada aplikasi komersial. Penyerapan ML cukup besar pada bidang ilmiah, biasanya model yang dilatih itu untuk akurasi yang tinggi. Pengaplikasian ML agar lebih mengoptimalkan dan menghasilkan hasil ilmiah [2].

C. Random Forest

Random Forest (RF) merupakan salah satu metode *Decision Tree*, atau pohon mengambil keputusan sebuah diagram alir yang berbentuk seperti pohon yang memiliki sebuah *root node* yang digunakan untuk mengumpulkan data. *Decision Tree* mengklasifikasi suatu sampel data yang belum diketahui kelasnya pada kelas-kelas yang ada. Penggunaan *decision tree* dapat menghindari *overfitting* pada sebuah set data ketika mencapai akurasi yang maksimum.

RF adalah kombinasi dari masing-masing *tree* yang baik kemudian dikombinasikan kedalam satu model. RF bergantung pada nilai vector random dengan distribusi yang sama pada semua pohon yang setiap *decision tree* memiliki kedalaman maksimal. RF adalah *classifier* yang berbentuk pohon $\{h(x, \theta_k), k = 1, \dots\}$ dimana θ_k adalah random vector yang didistribusikan secara independent dan setiap *tree* pada sebuah unit akan memilih *class* yang paling populer pada input x [9].

Decision tree memang ideal untuk metode *ensemble* karena memiliki bias yang rendah dan varian yang tinggi membuatnya sangat mendapat dari proses rata-rata. Kesulitan pada metode RF adalah menginjeksi keacakan sambil meminimalkan (x) secara bersamaan dan memperthankan bias rendah di RF.

Berikut karakteristik akurasi akurasi pada RF:

1. Memusatkan RF

Terdapat *classifier* $h_1(x)$, $h_2(x)$, $h_k(x)$ dan dengan training set dari distribusi pada vector Y dan X fungsi yang terbentuk adalah :

$$Mg(X, Y) = \underset{j \neq Y}{\max} \underset{k}{\text{av}_k I(h_k(X) = Y)} \quad (1)$$

Fungsi error

$$PE^* = P_{X,Y}(mg(X, Y) < 0) \quad (2)$$

Hasil dari penggabungan fungsi

$$P_{X,Y} \left(\underset{j \neq Y}{\max} P_{\theta}(h(X, \theta) = Y) - \underset{j \neq Y}{\max} P_{\theta}(h(X, \theta) = j) < 0 \right) \quad (3)$$

Hasil penggabungan fungsi ini menjelaskan mengapa RF tidak *overfit* saat *tree* ditambahkan, akan tetapi menghasilkan nilai yang terbatas pada error [9].

D. Anaconda

Anaconda adalah perangkat lunak gratis yang menyediakan alat yang disesuaikan untuk penelitian dan sains. Anaconda memberi *environment* yang dapat di program dengan Python atau dapat disebut IDE, adalah platform atau aplikasi yang didukung [10].

E. Spyder

Spyder adalah IDE berkinerja tinggi yang dirancang khusus analisis data. Spyder merupakan IDE lengkap yang menyertakan editor untuk menulis skrip kode, penjelajah variabel untuk memudahkan melihat data dan konsol ipython. Mencakup pemeriksaan kode yang kuat dan alat debugging yang memungkinkan pengguna untuk menganalisis kode sepenuhnya atau baris demi baris untuk mengidentifikasi dan memperbaiki kesalahan [10].

F. GrADS

GrADS (**Grid Analysis and Display System**) adalah alat desktop interaktif yang digunakan untuk memudahkan akses, manipulasi, dan visualisasi data ilmu bumi. Format data dapat berupa biner, GRIB, NetCDF, atau HDF-SDS (Scientific Data Sets). GrADS menggunakan lingkungan data 4-Dimensi: garis bujur, garis lintang, tingkat vertikal dan waktu. Kumpulan data ditempatkan dalam ruang 4-D dengan menggunakan file descriptor data [11].

G. Panoply

Panoply adalah aplikasi lintas platform yang berjalan di Macintosh, Windows, Linux, dan komputer desktop lainnya. Panoply dapat membuat plot warna dari lintang-bujur, lintang-vertikal, bujur-vertikal atau vertikal-waktu yang direferensikan dari 2D atau variable multidimensi yang lebih besar [12].

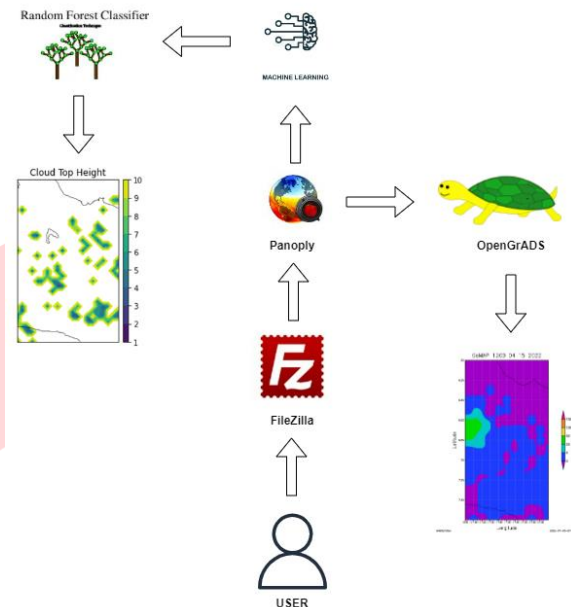
H. NetCDF

NetCDF (Network Common Data Form) adalah satu set pustakan perangkat lunak dan format data independen mesin yang mendukung pembuatan, akses dan berbagi data ilmiah berorientasi array. Proyek ini dimulai pada tahun 1988 dan masih didukung secara aktif oleh University Corporation for

Atmospheric Research (UCAR). Format biner netCDF asli (dirilis pada tahun 1990, sekarang dikenal sebagai "format klasik netCDF") masih banyak digunakan di seluruh dunia dan terus didukung penuh di semua rilis netCDF [13].

III. METODE

A. SINGKATAN DAN AKRONIM



GAMBAR 2
(DESAIN PERANCANGAN SISTEM)

Pada penelitian ini diaplikasikan secara langsung menggunakan perangkat sederhana, proses sistem ini dimulai dengan pengambilan data pada satelit Himawari-8 menggunakan FileZilla lalu membuka data dengan Panoply untuk melakukan proses cek pada data yang sudah di unduh melalui FileZilla, proses selanjutnya akan dilakukan *training* data pada ML dengan menerapkan metode algoritma RF dengan menggunakan Spyder IDE dan OpenGrADS, yang akan mengeluarkan hasil berupa plot map dan nilai akurasi, setelah mendapatkan plot dan nilai akurasi hasil akan dilakukan perbandingan hasil dari data awan dan data hujan untuk dilakukan prediksi hujan, seperti pada gambar 2.

Pada gambar desain sistem diatas merupakan gambaran yang akan diproses untuk mendapatkan nilai akurasi ketinggian untuk memprediksi curah hujan. Berikut adalah penjelasan urutan sistem pada diagram blok :

1. Untuk user terlebih dahulu melakukan registrasi untuk menapatkan akun agar dapat men-download data dari satelit Himawari-8.
2. Setelah mendapatkan akun untuk mengakses data satelit, selanjutnya akan menggunakan aplikasi Filezilla untuk login terlebih dahulu sebelum mengakses satelit Himawari-8.
3. Proses selanjutnya pengunduhan data melalui Filezilla, dan user telah memiliki akses pada satelit Himawari-8
4. Setelah User melakukan pengunduhan data, User bisa melihat terlebih dahulu data satelit pada aplikasi Panoply.
5. Selanjutnya akan diproses pengolahan data aplikasi ML yang akan deprogram memakai IDE Spyder dan memakai metode RF.

- Proses selanjutnya akan dilakukan klasifikasi data training dan data validasi untuk memprediksi ketinggian awan (Cloud Top Height).
- Setelah selesai membuat kode, serta tidak menampilkan error maka akan menampilkan hasil plot dan nilai akurasi.
- Setelah mendapatkan hasil plot dan nilai akurasi untuk mendapatkan prediksi hujan, dengan melihat ketinggian dari awan yang diperoleh dari implementasi metode RF.

B. Data Yang Digunakan

Data yang digunakan adalah data netCDF (.nc) yang di unduh melalui satelit Himawari-8, yang berupa data awan dan data hujan. Dijelaskan sebagai berikut:

1. Akuisisi Data

FileZilla digunakan sebagai pengamiran data satelit Himawari-8, dengan login terlebih dahulu menggunakan username dan password yang didapatkan melalui pendaftaran di web berikut :

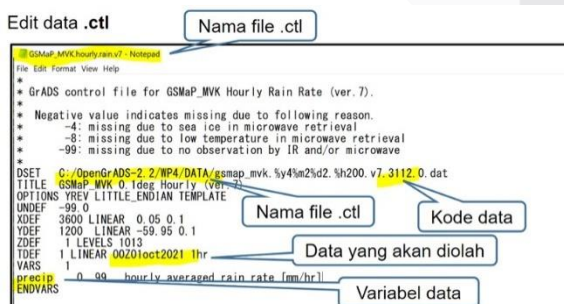
- <https://sharaku.eorc.jaxa.jp/GSMaP/registration.html>
- https://www.eorc.jaxa.jp/ptree/registration_top.htm

1

setelah login pada server Himawari-8 di link pertama untuk mengunduh data yang dibutuhkan sebanyak 120 data, yang berekstensi netCDF (.nc) yang ada pada server Himawari-8. Waktu data yang digunakan diambil dari tanggal 15 April 2022 dimulai dari pukul 08.00 UTC sampai 13.00 UTC atau pukul 15.00 sampai pukul 19.00 WIB. Setelah mengunduh data yang ditraining selanjutnya mengunduh data curah hujan dari GSMaP pada link kedua untuk mengambil data yang berekstensi (.ctl) dan (.dat) mulai dari tanggal 15 April 2022 dimulai pukul 08.00 UTC sampai 16.00 UTC atau pukul 15.00 sampai 22.00 WIB Untuk wilayah yang diprediksi adalah wilayah Jawa Barat.

2. Preprocessing Data

Setelah didapatkan data awan, selanjutnya melakukan training pada data awan menggunakan ML dengan algoritma RF untuk melihat perkembangan awan sebelum terjadi hujan. Untuk data hujan yang di-running pada GrADS dilakukan perubahan pada data ctl di notepad seperti berikut:



GAMBAR 3

(ISI DATA CTL PADA NOTEPAD)

Mengedit data (.ctl) untuk memasukkan data (.dat) kepada format (.ctl) untuk dimasukkan kedalam GrADS. Pengeditan data (.ctl) pada notepad hal yang dirubah ialah seperti direktori file yang dipakai

kemudian nama file (.dat) yang di unduh, memasukkan kode data dan memasukkan tanggal untuk data yang diolah.

```
*
* GRADS control file for GSMaP_MVK Hourly Rain Rate (ver.7).
*
* Negative value indicates missing due to following reason.
* -4: missing due to sea ice in microwave retrieval
* -8: missing due to low temperature in microwave retrieval
* -99: missing due to no observation by IR and/or microwave
*
DSET C:\OpenGrADS-2.2\WP4\DATA\gsmmap_mv_k.20220415.1100.v7.3112.0.dat
TITLE GSMaP_MVK 0.1deg Hourly (ver.7)
OPTIONS YREV LITTLE_ENDIAN TEMPLATE
UNDEF -99.0
XDEF 3600 LINEAR 0.05 0.1
YDEF 1200 LINEAR -59.95 0.1
ZDEF 1 LEVELS 1013
TDEF 87600 LINEAR 00Z15Apr2022 1hr
VARS 1
precip 0 99 hourly averaged rain rate [mm/hr]
ENDVARS
```

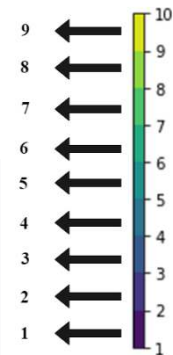
GAMBAR 3

(ISI DATA FILE CTL PADA NOTEPAD SETELAH DIEDIT)

Perubahan dilakukan agar dapat menampilkan plot map, perubahan pada file ctl di notepad dilakukan sesuai jam, tanggal dan bulan yang digunakan.

3. Class

Pada plot map yang dihasilkan dimunculkan legend dengan 9 warna yang digunakan untuk capaian label pada ketinggian awan dengan penjelasan sebagai berikut :



GAMBAR 4

(LEGEND 9 WARNA)

- Pada nomor 1 mempunyai warna *persian indigo* dengan jarak 1 sampai 2 km untuk mengetahui jika ada warna ini di prediksi hujan.
- Pada nomor 2 mempunyai warna *dark slate blue* dengan jarak 2 sampai 3 km untuk mengetahui jika ada warna ini di prediksi hujan.
- Pada nomor 3 mempunyai warna *tory blue* dengan jarak 3 sampai 4 km untuk mengetahui jika ada warna ini di prediksi hujan.
- Pada nomor 4 mempunyai warna *lochmara* dengan jarak 4 sampai 5 km untuk mengetahui jika ada warna ini di prediksi hujan ringan.
- Pada nomor 5 mempunyai warna *deep sea* dengan jarak 5 sampai 6 km dengan jarak 5 sampai 6 km untuk mengetahui jika ada warna ini di prediksi hujan ringan.
- Pada nomor 6 mempunyai warna *jungle green* dengan jarak 6 sampai 7 km untuk mengetahui jika ada warna ini di prediksi hujan ringan.
- Pada nomor 7 mempunyai warna *emerald* dengan jarak 7 sampai 8 km untuk mengetahui jika ada warna ini di prediksi tidak hujan.
- Pada nomor 8 mempunyai warna *atlantis* dengan jarak 8 sampai 9 km untuk mengetahui jika ada warna ini di prediksi tidak hujan.

- i. Pada nomor 9 mempunyai warna *pear* dengan jarak 9 sampai 10 km untuk mengetahui jika ada warna ini di prediksi tidak hujan.

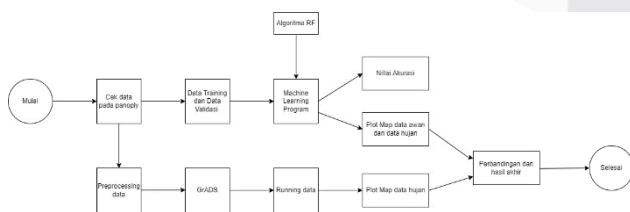
TABEL 1
(KASUS PREDIKSI KETINGGIAN AWAN)

No.	Jarak	Ketinggian Awan	Prediksi
1	1-2 km	Rendah	Hujan
2	2-3 km	Rendah	Hujan
3	3-4 km	Rendah	Hujan
4	4-5 km	Sedang	Hujan Ringan
5	5-6 km	Sedang	Hujan Ringan
6	6-7 km	Sedang	Hujan Ringan
7	7-8 km	Tinggi	Tidak Hujan
8	8-9 km	Tinggi	Tidak Hujan
9	9-10 km	Tinggi	Tidak Hujan

Pada tabel diatas adalah prediksi ketinggian awan, sebagai berikut:

- jika awan pada jarak 1 sampai 2 km dengan ketinggian awan rendah maka diprediksi akan hujan.
- Jika awan pada jarak 2 sampai 3 km dengan ketinggian awan rendah maka diprediksi akan hujan.
- Jika awan pada jarak 3 sampai 4 km dengan ketinggian awan rendah maka diprediksi akan hujan.
- Jika awan pada jarak 4 sampai 5 km dengan ketinggian awan sedang maka diprediksi akan hujan ringan.
- Jika awan pada jarak 5 sampai 6 km dengan ketinggian awan sedang maka diprediksi akan hujan ringan.
- Jika awan pada jarak 6 sampai 7 km dengan ketinggian awan sedang maka diprediksi akan hujan ringan.
- Jika awan pada jarak 7 sampai 8 km dengan ketinggian awan tinggi diprediksi tidak hujan.
- Jika awan pada jarak 8 sampai 9 km dengan ketinggian awan tinggi diprediksi tidak hujan.
- Jika awan pada jarak 9 sampai 10 km dengan ketinggian awan tinggi diprediksi tidak hujan.

C. Sistem Perangkat Lunak



Gambar 5

(Flowchart perangkat lunak)

Pada gambar diatas terdapat flowchart perangkat lunak, dengan penjelasan sebagai berikut :

- Sebelum melakukan klasifikasi data, dilakukan terlebih dahulu cek data di Panoply untuk dilihat kelengkapan data dari jenis kanal hingga plot data.

- Selanjutnya melakukan klasifikasi data menjadi dua bagian yaitu data training dan data validasi.
- Proses training dilakukan pada ML dengan menerapkan algoritma RF untuk mendapatkan hasil nilai akurasi serta plot map data awan dan data hujan.
- Setelah mendapatkan nilai akurasi dan plot map dari proses training data pada ML, selanjutnya melakukan preprocessing data (mengubah nama dataset, tanggal, jam dan kode data) untuk penyesuaian dengan data yang di-running pada GrADS agar dapat memunculkan plot map.
- Melakukan running data di GrADS.
- Mendapatkan hasil plot data hujan.
- Melakukan perbandingan letak titik hujan melalui perkembangan awan pada plot data awan dan letak titik hujan pada plot data hujan.
- Setelah perbandingan selesai, kemudian menemukan kecocokan antara data awan dan data hujan maka proses prediksi selesai.

D. Diagram Alir Sistem

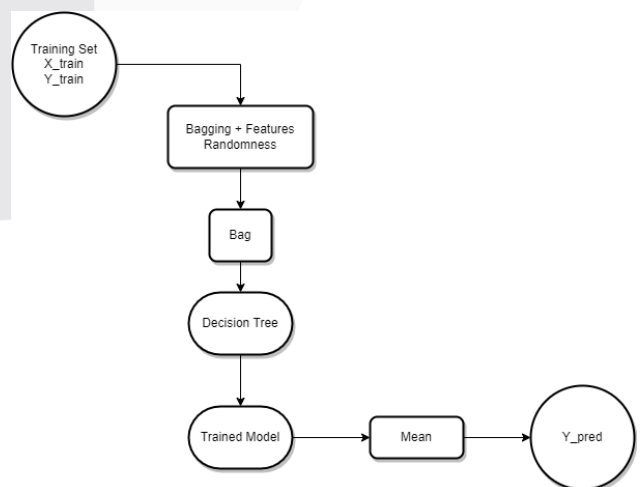
Sistem untuk mengklasifikasikan ketinggian awan berdasarkan tinggi yang didapatkan untuk mendapatkan curah hujan yang akan terjadi. Adapun diagram alir pada sistem sebagai berikut :



GAMBAR 6
(DIAGRAM ALIR SISTEM)

Seperti pada flowchart diatas dapat dilihat pada data Himawari-8 yang didapatkan ialah berbentuk netCDF (.nc), kemudian data netCDF diolah terlebih dahulu ke dalam Panoply untuk dilihat plot data yang dipakai ada atau tidak, kemudian data tersebut di masukkan kedalam ML dengan algoritma RF untuk mendapatkan hasil nilai prediksi pada ketinggian awan dan *output* plot ketinggian awan.

E. Cara Kerja Random Forest



GAMBAR 7
(FLOWCHART CARA KERJA RF)

Proses kerja RF pada contoh kasus di tugas akhir ini dengan titik acuan adalah ketinggian awan pada flowchart

cara kerja dari RF dimulai dari men-training data X_{train} dan Y_{train} kemudian akan dilanjutkan ke proses *Bootstrap Aggregating* (Bagging) dan fitur keacakan. Setelah melalui proses Bagging maka dilanjutkan ke proses Bag maka melakukan *Decision Tree* dan men-training model untuk mendapatkan nilai mean dan berhenti pada Y_{pred} untuk menampilkan akurasi yang diprediksikan parameter *tree* yang diuji mulai dari parameter $n_{estimators} = 2$ sampai 70, semakin tinggi parameter yang digunakan maka data titik awan pada plot akan semakin akurat.

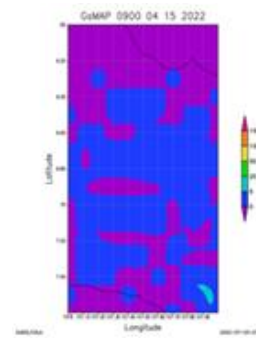
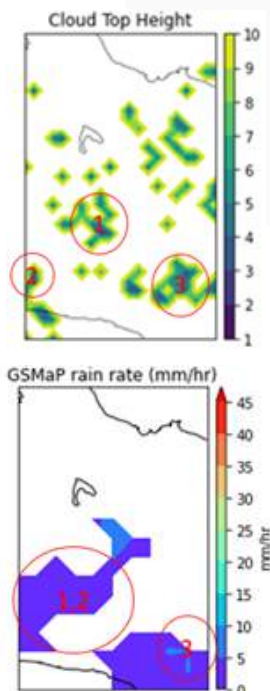
IV. HASIL DAN PEMBAHASAN

Pada proses pengujian dilakukan dengan membagi data menjadi data training dan data validasi. Tahap awal pengujian untuk mendapatkan output pada data awan yang berupa ketinggian awan dimulai dengan menggunakan parameter algoritma *Random Forest (RF)*, hasil yang diperoleh akan dibandingkan dengan memilih 5 akurasi terendah dan 5 akurasi tertinggi. Pada pengujian yang telah dilakukan proses training data sebanyak 120 data, setelah proses training 60 data didapatkan nilai akurasi tertinggi yaitu 0.9071202978212084, lalu menambahkan 60 data lagi untuk dilakukan training untuk mendapatkan nilai akurasi lebih besar dari 0.9071202978212084, maka hasil yang didapatkan mencapai 1.0.

CLTH (Cloud Top Height) diartikan sebagai ketinggian awan adalah parameter yang digunakan pada tugas akhir ini. UTC adalah (Coordinated Universal Time) waktu global dan lebih maju dari Indonesia. Plot berbentuk peta yang memperlihatkan wilayah Jawa Barat dan memperlihatkan ketinggian pada awan.

A. Perbandingan Data Awan dan Data Hujan Akurasi Terendah

Melakukan perbandingan secara kualitatif dilihat dari perkembangan awan dan titik hujan pada nilai akurasi terendah sebagai berikut:

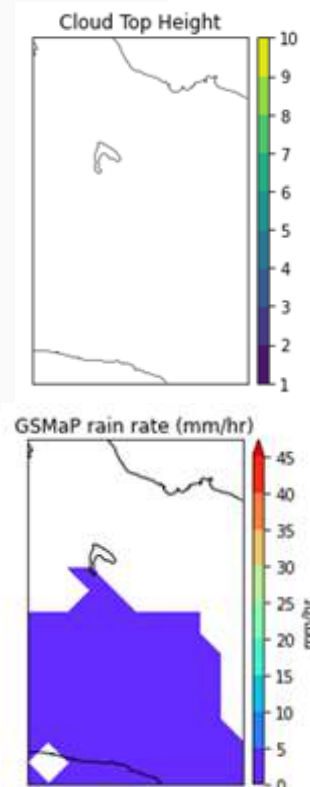


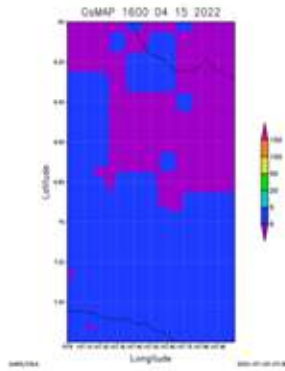
Gambar dengan akurasi yang dihasilkan ialah 76.67%, data yang digunakan bulan April 15 2022, pada jam 08.00 UTC atau jam 15.00 WIB, gambar di atas menunjukkan perbandingan antara hasil plot yang dihasilkan dari ML, GSMaP, dan GrADS. Pada citra ML dapat dilihat beberapa warna awan yang ditunjukkan pada Gambar 3.2 untuk warna dan tinggi yang dihasilkan serta hasil prediksi yang akan dihasilkan ada pada Tabel 3.1, sehingga dapat dibandingkan sebagai berikut:

- Pada lintang 7 – 7,4 derajat dan bujur 107 – 107,6 derajat telah terjadi hujan yang ditandai dengan lingkaran angka 1 dan 2.
- Pada lintang 7,4 - 7,5 derajat dan bujur 107,7 - 107,9 derajat telah terjadi hujan yang ditunjukkan dengan lingkaran nomor 3.

B. Perbandingan Data Awan dan Data Hujan Akurasi Tertinggi

Melakukan perbandingan secara kualitatif dilihat dari perkembangan awan dan titik hujan pada nilai akurasi terendah sebagai berikut:





C. Kesimpulan Penelitian

Tabel 2
(5 Akurasi Terendah)

No	Nilai N = 2	Nilai N = 10	Nilai N = 20	Nilai N = 30	Nilai N = 50	Nilai N = 70
1	69.84%	75.63%	76.18%	76.46%	76.56%	76.67%
2	72.41%	77.40%	78.15%	78.27%	78.54%	78.58%
3	77.63%	81.10%	81.36%	81.72%	81.72%	81.90%
4	81.14%	84.35%	84.75%	84.80%	85.08%	85.01%
5	83.81%	87.44%	87.70%	87.83%	87.91%	87.94%

Tabel 3
(5 Akurasi Tertinggi)

No	Nilai N = 2	Nilai N = 10	Nilai N = 20	Nilai N = 30	Nilai N = 50	Nilai N = 70
1	91.53%	93.87%	94.07%	94.16%	94.21%	94.23%
2	95.42%	95.48%	95.71%	97.48%	97.51%	97.52%
3	98.51%	98.77%	99.33%	99.35%	99.36%	99.36%
4	99.90%	99.86%	99.98%	99.98%	99.98%	99.98%
5	100 %	100 %	100 %	100 %	100 %	100 %

Pada hasil penilitan Tugas Akhir ini, proses training data dilakukan 6 kali dengan nilai N pada algoritma berbeda-beda, yaitu N = 2, N = 10, N = 20, N = 30, N = 50, dan N = 70. Dari semua hasil training data tersebut, disimpulkan bahwa hasil training data dari Nilai N = 70 adalah yang terbaik karena tidak mengalami penurunan pada nilai akurasi yang dihasilkan. Kemudian proses pembandingan plot data awan dan data hujan ditandai dengan angka untuk melihat kesesuaian dari letak perkembangan awan dan letak terjadinya hujan, jika disesuaikan dengan letak latitude dan longitude yang sudah sesuai. Namun pada akurasi tertinggi plot yang dihasilkan tidak dapat terlihat dikarenakan ketinggian awan sudah melebihi ketinggian yang ditentukan

V. KESIMPULAN

Penelitian prediksi hujan menggunakan hasil pelatihan data awan yang dapat menghasilkan peta plot dan tingkat akurasi 70% - 100%. Pada penelitian ini, kami berhasil melakukan perbandingan melalui pengujian 120 data yang terbagi menjadi 60 data latih dan 60 data validasi, yang mendapatkan 5 akurasi terendah pada setiap n_estimator dan 5 akurasi tertinggi pada setiap n_estimator. Namun setelah dilakukan pengujian sebanyak enam kali, dan merubah nilai n_estimator hasil akurasi yang didapat semakin meningkat namun semakin besar nilai n_estimator yang dijalankan pada program membutuhkan hasil yang sangat lama, dibutuhkan waktu hampir dua sampai tiga jam untuk menghasilkan plot dan nilai akurasi. Namun pada hasil nilai akurasi tertinggi, titik awan tidak ada, karena awan telah melewati batas lebih dari 10 kilometer.

REFERENSI

- [1] V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, Y. Chen, L. Goldfarb, M. I. Gomis, J. Matthews, S. Berger, M. Huang, O. Yelekçi, R. Yu, B. Zhou, E. Lonnoy Dan T. K. Maycock, *Climate Change 2021 The Physical Science Basis*, Switzerland: IPCC, 2021.
- [2] R. ROSCHER, B. BOHN, M. F. DUARTE Dan J. GARCKE, "Explainable Machine Learning For Scientific," *IEEE Access*, No. IEEE, P. 42200, 2020.
- [3] L. BREIMAN, "Random Forests," *Machine Learning*, No. Kluwer Academic Publishers. Manufactured In The Netherlands, P. 6, 2001.
- [4] K. BESSHO, K. DATE, M. HAYASHI, A. IKEDA, T. IMAI, H. INOUE, Y. KUMAGAI, T. MIYAKAWA, H. MURATA, T. OHNO, A. OKUYAMA, R. OYAMA, Y. SASAKI, Y. SHIMAZU, K. SHIMOJI, Y. SUMIDA Dan M. SUZUKI, "An Introduction To Himawari-8/9—Japan's New-Generation Geostationary Meteorological Satellites," *Journal Of The Meteorological Society Of Japan Ser II*, No. Artikel, P. 1, 2016.
- [5] N. Ayasha, "A COMPARISON OF RAINFALL ESTIMATION USING HIMAWARI-8 SATELLITE DATA IN DIFFERENT INDONESIAN TOPOGRAPHIES," *International Journal Of Remote Sensing*, No. International Journal, P. 1, 2020.
- [6] J. A. Coakley, "REFLECTANCE AND ALBEDO, SURFACE," *Encyclopedia Of The Atmosphere.*, No. Copyright 2003 Elsevier Science Ltd. All Rights Reserved., P. 1914, 2003.
- [7] X. Yan, Z. Li, N. Luo, W. Shi, W. Zhao, X. Yang Dan J. Jin, "A Minimum Albedo Aerosol Retrieval Method For The New-Generation Geostationary," No. Elsevier, P. 5, 2018.
- [8] V. Villani, "ATLASSIAN," ATLASSIAN CONFLUENCE, Rabu Desember 2021. [Online]. Available: <https://Opus.Nci.Org.Au/Display/NDP/Himawari->

AHI%2C+Full+Disk%2C+Brightness+Temperature.
[Diakses 29 Juli 2022].

- [9] G. Louppe, Understanding Random Forest, Belgium: University Of Liège, 2015.
- [10] M. R. T. R.-M. & K. C. Damien Rolon-M'Érette, "Introduction To Anaconda And Python: Installation And Setup," *Python For Reasearch In Psychology*, No. Crossmark, P. 55, 2020.
- [11] GRADS, "GRADS," Center For Ocean-Land-Atmosphere Studies, Institute Of Global Environment And Society, George Mason University, Virginia, 2021.
- [12] NASA, "Panoply Netcdf, HDF And GRIB Data Viewer," NASA, Rabu Juli 2022. [Online]. Available:
<https://www.giss.nasa.gov/tools/panoply/>.
[Diakses Rabu Juli 2022].
- [13] UNIDATA, "Network Common Data Form (Netcdf)," University Corporation For Atmospheric Research Community Programs, Sabtu Juli 2022. [Online]. Available:
<https://www.unidata.ucar.edu/software/netcdf/>.
[Diakses 31 Juli 2022].