

**PRAKIRAAN CUACA KOTA BANDAR LAMPUNG DENGAN
MENGUNAKAN *RANDOM FOREST* DAN *C4.5***

Tesis

Oleh

**RAHMA FERIKA SHAUMI
NPM 1925031007**



**FAKULTAS TEKNIK
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

**PRAKIRAAN CUACA KOTA BANDAR LAMPUNG DENGAN
MENGUNAKAN *RANDOM FOREST* DAN *C4.5***

Oleh

RAHMA FERIKA SHAUMI

Tesis

Sebagai Salah Satu Syarat untuk Mencapai Gelar Magister Teknik

Pada

**Prodi Magister Teknik Elektro
Fakultas Teknik Universitas Lampung**



**FAKULTAS TEKNIK
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRAK

PRAKIRAAN CUACA KOTA BANDAR LAMPUNG DENGAN MENGUNAKAN *RANDOM FOREST* DAN *C4.5*

Oleh

RAHMA FERIK SHAUMI

Prakiraan cuaca merupakan informasi yang penting bagi berbagai instansi maupun masyarakat luas. Prakiraan cuaca biasanya digunakan untuk kepentingan berbagai sektor seperti transportasi, pariwisata perkebunan dan lain-lain. Penelitian ini bertujuan untuk membuat model baru mengenai prakiraan cuaca dengan menggunakan algoritma random forest dan C4.5 dengan menggunakan aplikasi WEKA. Dataset menggunakan data Stasiun Meteorologi Maritim Panjang dengan banyaknya data 365 hari dengan 6 atribut.

Dataset yang digunakan terlebih dahulu melalui *preprocessing* data, *replace missing values*, *training* data dan penerapan kedua algoritma hingga menghasilkan prediksi yang kemudian dievaluasi dengan MAE dan RMSE. Hasil dari penelitian ini yaitu algoritma C4.5 lebih baik pada uji coba dengan *cross-validation folds* 5, 10 dan 15 dibandingkan dengan algoritma *random forest*, yaitu dengan nilai MAE 0.0284, 0.0277, dan 0.0285 serta nilai RMSE 0.1408, 0.1392 dan 0.1414. Sedangkan hasil dari *random forest* untuk nilai MAE adalah 0.1131, 0.112 dan 0.1119 serta nilai RMSE 0.216, 0.2146 dan 0.2146. Kemudian algoritma C4.5 juga lebih baik pada uji coba *percentage split* 25%, 66% dan 90% dibandingkan algoritma *random forest* yaitu dengan nilai MAE 0.0613, 0.0329 dan 0.0234 serta nilai RMSE 0.2287, 0.1463 dan 0.102. Sedangkan hasil dari *random forest* untuk nilai MAE adalah 0.111, 0.1202 dan 0.1295 serta nilai RMSE 0.2218, 0.2428 dan 0.2523.

Kata kunci: prakiraan cuaca, *random forest*, C4.5, WEKA.

ABSTRACT

WEATHER FORECAST FOR BANDAR LAMPUNG CITY USING RANDOM FOREST AND C4.5

Oleh

RAHMA FERIK SHAUMI

Weather forecasts are essential information for various agencies and the wider community. Weather forecasts are usually used to benefit multiple sectors such as transportation, tourism, plantations, and others. This study aims to create a new model regarding weather forecasting using the random forest and C4.5 algorithms using the WEKA application. The dataset uses data from the Long Maritime Meteorological Station with 365 days of data with six attributes.

The dataset used first goes through data preprocessing, replacing missing values, training data, and applying the two algorithms to produce predictions which are then evaluated with MAE and RMSE. The results of this study are that the C4.5 algorithm is better in trials with cross-validation folds 5, 10, and 15 compared to the random forest algorithm, with MAE values of 0.0284, 0.0277, and 0.0285 and RMSE values of 0.1408, 0.1392 and 0.1414. While the results of the random forest for the MAE values are 0.1131, 0.112, and 0.1119, and the RMSE values are 0.216, 0.2146, and 0.2146. Then the C4.5 algorithm is also better in the percentage split trials of 25%, 66%, and 90% compared to the random forest algorithm, with MAE values of 0.0613, 0.0329, and 0.0234 and RMSE values of 0.2287, 0.1463 and 0.102. While the results of the random forest for the MAE values are 0.111, 0.1202, and 0.1295, and the RMSE values are 0.2218, 0.2428, and 0.2523.

Keywords: weather forecast, random forest, C4.5, WEKA.

Judul Tesis : **PRAKIRAAN CUACA KOTA BANDAR
LAMPUNG DENGAN MENGGUNAKAN
RANDOM FOREST DAN C4.5**

Nama Mahasiswa : **Rahma Ferika Shaumi**

Nomor Pokok Mahasiswa : 1925031007

Program Studi : Magister Teknik Elektro

Fakultas : Teknik

MENYETUJUI

1. Komisi Pembimbing

Dr. Ir. Sri Ratna Sulistiyanti, M.T.
NIP. 196510211995122001

Dr. Eng. F. X. Arinto Setyawan, S.T., M.T.
NIP. 196912191999031002

2. Ketua Program Studi Magister Teknik Elektro

Misfa Susanto, S.T., M.Sc., Ph.D.
NIP. 197105251999031001

MENGESAHKAN

1. Komisi Penguji

Ketua : **Dr. Ir. Sri Ratna Sulistiyanti, M.T.**

Sekretaris : **Dr. Eng. F. X. Arinto Setyawan, S.T., M.T.**

Penguji : **Dr. Eng. Ir. Helmy Fitriawan, S.T., M.Sc.**

Anggota : **Misfa Susanto, S.T., M.Sc., Ph.D**

2. Dekan Fakultas Teknik

Dr. Eng. Ir. Helmy Fitriawan, S.T., M.Sc.
NIP. 19750928 200112 1 002

3. Direktur Program Pascasarjana

Prof. Dr. Ir. Murhadi, M.Si.
NIP. 19640326 198902 1 001

Tanggal Lulus Ujian Tesis : **17 Juni 2023**

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tesis yang saya buat tidak terdapat karya orang lain dan diterbitkan orang lain kecuali secara tertulis diacu dalam naskah sebagaimana telah dituliskan dalam daftar pustaka. Selain dari itu saya menyatakan bahwa skripsi ini saya buat sendiri.

Apabila pernyataan saya tidak benar maka, saya bersedia terkena sanksi sesuai dengan hukum yang berlaku.

Bandar Lampung, 19 Juni 2023

Rahma Ferika Shaumi
1925031007

RIWAYAT HIDUP



Penulis dilahirkan di Marga Agung pada tanggal 19 Februari 1996, anak dari bapak Haryono dan ibu Mariyaningsih, anak ke pertama dari tiga bersaudara. Pendidikan sekolah dasar, diselesaikan pada tahun 2008 di SDN 1 Marga Agung, sekolah menengah pertama diselesaikan di MTs. Al- Hidayah Jati Agung pada tahun 2011, kemudian penulis melanjutkan pendidikan sekolah menengah atas di SMA Gajah Mada Bandar Lampung.

Pada tahun 2014 penulis melanjutkan pendidikan di Universitas Lampung jurusan Teknik Elektro dengan jalur masuk SNMPTN (Seleksi Nasional Masuk Perguruan Tinggi Negeri). Penulis menyelesaikan studi S1 nya pada tanggal 12 November 2018. Pada tahun 2019 penulis melanjutkan pendidikan Magister di Universitas Lampung dengan Jurusan Teknik Elektro. Pada tahun 2021 penulis mulai bekerja sebagai PNS di Instansi Kemenkominfo pada Satuan Kerja LPP TVRI Stasiun Lampung. Program Studi Magister Teknik Elektro Penulis menyelesaikan studinya pada tanggal 17 Juni 2023 dengan judul tesis “Prakiraan Cuaca Kota Bandar Lampung Dengan Menggunakan *Random Forest* dan *C4.5*”.

Ku persembahkan tesis ini untuk:

Suamiku Ardhi Istiadi.

Kedua orang tuaku, Bapak Haryono dan Ibu Mariyaningsih.

Kedua mertuaku, Bapak Mislam Ismail (Alm) dan Ibu Nunsiyanti.

Adikku Ahmad Taufik Fadhani, Dicky Arya Novandi,

Artha Maulana Rahman,

serta untuk keluargaku dan almamaterku tercinta.

MOTTO

"Sebaik-baik manusia diantaramu adalah yang paling banyak manfaatnya bagi orang lain." (H.R. Bukhari)

UCAPAN TERIMAKASIH

Puji syukur penulis ucapkan kehadiran Tuhan Yang Maha Esa, karena atas rahmat dan hidayah-Nya tesis ini dapat diselesaikan.

Tesis dengan judul “Prakiraan Cuaca Kota Bandar Lampung Dengan Menggunakan *Random Forest* Dan C4.5” merupakan salah satu syarat untuk memperoleh gelar Magister Teknik di Universitas Lampung.

Dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Bapak Dr. Eng. Ir. Helmy Fitriawan, S.T., M.Sc. selaku Dekan Fakultas Teknik Unila sekaligus Penguji tesis ini yang telah memberikan arahan dan bantuannya dalam proses penyelesaian tesis ini.
2. Ibu Herlinawati, S.T., M.T. selaku Ketua Jurusan Teknik Elektro Unila.
3. Bapak Misfa Susanto, S.T., M.Sc., P.hD. selaku Ketua Program Studi Magister Teknik Elektro dan Anggota Penguji tesis ini yang telah memberikan arahan dan bantuannya dalam proses penyelesaian tesis ini.
4. Ibu Dr. Ir. Sri Ratna Sulistiyanti, M.T. selaku pembimbing utama yang telah bersedia untuk memberikan bimbingan, saran, masukan dalam proses penyelesaian tesis ini.
5. Bapak Dr. Eng. F. X. Arinto Setyawan, S.T., M.T. selaku pembimbing pendamping yang telah bersedia untuk memberikan bimbingan, saran, masukan dalam proses penyelesaian tesis ini.
6. Bapak dan Ibu dosen Magister Teknik Elektro yang sudah memberikan ilmu selama proses KBM berlangsung.

7. Bapak dan Ibu staff administrasi Magister Teknik Elektro Unila.
8. Teman-teman angkatan 2019, Yudi Eka Putra, M Ridho Yoga Permono, Ubaidah, Sitronella Nurfitriani Hasim, Sri Suryadi dan Nadia Julian Putri yang telah membantu dan selalu memberikan semangat.
9. Kakak tingkat dan adik tingkat yang telah membantu dan selalu memberikan semangat.

Bandar Lampung, 19 Juni 2023

Rahma Ferika Shaumi

ABSTRAK

PRAKIRAAN CUACA KOTA BANDAR LAMPUNG DENGAN MENGUNAKAN *RANDOM FOREST* DAN *C4.5*

Oleh

RAHMA FERIKA SHAUMI

Prakiraan cuaca merupakan informasi yang penting bagi berbagai instansi maupun masyarakat luas. Prakiraan cuaca biasanya digunakan untuk kepentingan berbagai sektor seperti transportasi, pariwisata perkebunan dan lain-lain. Penelitian ini bertujuan untuk membuat model baru mengenai prakiraan cuaca dengan menggunakan algoritma random forest dan C4.5 dengan menggunakan aplikasi WEKA. Dataset menggunakan data Stasiun Meteorologi Maritim Panjang dengan banyaknya data 365 hari dengan 6 atribut.

Dataset yang digunakan terlebih dahulu melalui *preprocessing* data, *replace missing values*, *training* data dan penerapan kedua algoritma hingga menghasilkan prediksi yang kemudian dievaluasi dengan MAE dan RMSE. Hasil dari penelitian ini yaitu algoritma *C.4.5* lebih baik pada uji coba dengan *cross-validation folds* 5, 10 dan 15 dibandingkan dengan algoritma *random forest*, yaitu dengan nilai MAE 0.0284, 0.0277, dan 0.0285 serta nilai RMSE 0.1408, 0.1392 dan 0.1414. Sedangkan hasil dari *random forest* untuk nilai MAE adalah 0.1131, 0.112 dan 0.1119 serta nilai RMSE 0.216, 0.2146 dan 0.2146. Kemudian algoritma *C4.5* juga lebih baik pada uji coba *percentage split* 25%, 66% dan 90% dibandingkan algoritma *random forest* yaitu dengan nilai MAE 0.0613, 0.0329 dan 0.0234 serta nilai RMSE 0.2287, 0.1463 dan 0.102. Sedangkan hasil dari *random forest* untuk nilai MAE adalah 0.111, 0.1202 dan 0.1295 serta nilai RMSE 0.2218, 0.2428 dan 0.2523.

Kata kunci: prakiraan cuaca, *random forest*, *C4.5*, WEKA.

ABSTRACT

WEATHER FORECAST FOR BANDAR LAMPUNG CITY USING RANDOM FOREST AND C4.5

Oleh

RAHMA FERIKHA SHAUMI

Weather forecasts are essential information for various agencies and the wider community. Weather forecasts are usually used to benefit multiple sectors such as transportation, tourism, plantations, and others. This study aims to create a new model regarding weather forecasting using the random forest and C4.5 algorithms using the WEKA application. The dataset uses data from the Long Maritime Meteorological Station with 365 days of data with six attributes.

The dataset used first goes through data preprocessing, replacing missing values, training data, and applying the two algorithms to produce predictions which are then evaluated with MAE and RMSE. The results of this study are that the C4.5 algorithm is better in trials with cross-validation folds 5, 10, and 15 compared to the random forest algorithm, with MAE values of 0.0284, 0.0277, and 0.0285 and RMSE values of 0.1408, 0.1392 and 0.1414. While the results of the random forest for the MAE values are 0.1131, 0.112, and 0.1119, and the RMSE values are 0.216, 0.2146, and 0.2146. Then the C4.5 algorithm is also better in the percentage split trials of 25%, 66%, and 90% compared to the random forest algorithm, with MAE values of 0.0613, 0.0329, and 0.0234 and RMSE values of 0.2287, 0.1463 and 0.102. While the results of the random forest for the MAE values are 0.111, 0.1202, and 0.1295, and the RMSE values are 0.2218, 0.2428, and 0.2523.

Keywords: weather forecast, random forest, C4.5, WEKA.

Judul Tesis : **PRAKIRAAN CUACA KOTA BANDAR
LAMPUNG DENGAN MENGGUNAKAN
RANDOM FOREST DAN C4.5**

Nama Mahasiswa : **Rahma Ferika Shaumi**

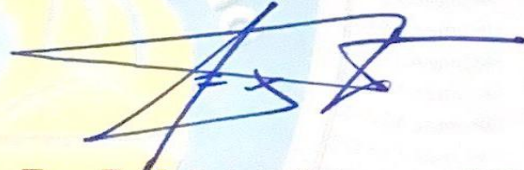
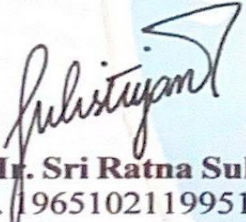
Nomor Pokok Mahasiswa : 1925031007

Program Studi : Magister Teknik Elektro

Fakultas : Teknik


MENYETUJUI

1. Komisi Pembimbing



Dr. Ir. Sri Ratna Sulistiyanti, M.T. **Dr. Eng. F. X. Arinto Setyawan, S.T., M.T.**
NIP. 196510211995122001 NIP. 196912191999031002

2. Ketua Program Studi Magister Teknik Elektro

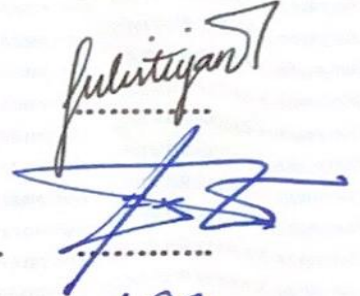


Misfa Susanto, S.T., M.Sc., Ph.D.
NIP. 197105251999031001

MENGESAHKAN

1. Komisi Penguji

Ketua : Dr. Ir. Sri Ratna Sulistiyanti, M.T.



Sekretaris : Dr. Eng. F. X. Arinto Setyawan, S.T., M.T.



Penguji : Dr. Eng. Ir. Helmy Fitriawan, S.T., M.Sc.



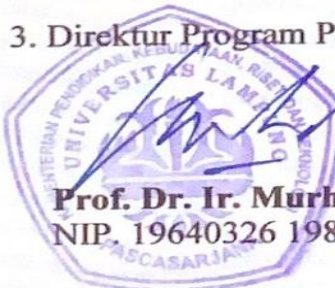
Anggota : Misfa Susanto, S.T., M.Sc., Ph.D.

2. Dekan Fakultas Teknik



Dr. Eng. Ir. Helmy Fitriawan, S.T., M.Sc.
NIP. 19750928 200112 1 002

3. Direktur Program Pascasarjana



Prof. Dr. Ir. Murhadi, M.Si.
NIP. 19640326 198902 1 001

Tanggal Lulus Ujian Tesis : 17 Juni 2023

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tesis yang saya buat tidak terdapat karya orang lain dan diterbitkan orang lain kecuali secara tertulis diacu dalam naskah sebagaimana telah dituliskan dalam daftar pustaka. Selain dari itu saya menyatakan bahwa skripsi ini saya buat sendiri.

Apabila pernyataan saya tidak benar maka, saya bersedia terkena sanksi sesuai dengan hukum yang berlaku.

Bandar Lampung, 19 Juni 2023



Rahma Ferika Shaumi
1925031007

RIWAYAT HIDUP



Penulis dilahirkan di Marga Agung pada tanggal 19 Februari 1996, anak dari bapak Haryono dan ibu Mariyaningsih, anak ke pertama dari tiga bersaudara. Pendidikan sekolah dasar, diselesaikan pada tahun 2008 di SDN 1 Marga Agung, sekolah menengah pertama diselesaikan di MTs. Al- Hidayah Jati Agung pada tahun 2011, kemudian penulis melanjutkan pendidikan sekolah menengah atas di SMA Gajah Mada Bandar Lampung.

Pada tahun 2014 penulis melanjutkan pendidikan di Universitas Lampung jurusan Teknik Elektro dengan jalur masuk SNMPTN (Seleksi Nasional Masuk Perguruan Tinggi Negeri). Penulis menyelesaikan studi S1 nya pada tanggal 12 November 2018. Pada tahun 2019 penulis melanjutkan pendidikan Magister di Universitas Lampung dengan Jurusan Teknik Elektro. Pada tahun 2021 penulis mulai bekerja sebagai PNS di Instansi Kemenkominfo pada Satuan Kerja LPP TVRI Stasiun Lampung. Program Studi Magister Teknik Elektro Penulis menyelesaikan studinya pada tanggal 17 Juni 2023 dengan judul tesis “Prakiraan Cuaca Kota Bandar Lampung Dengan Menggunakan *Random Forest* dan *C4.5*”.

Ku persembahkan tesis ini untuk:

Suamiku Ardhi Istiadi.

Kedua orang tuaku, Bapak Haryono dan Ibu Mariyaningsih.

Kedua mertuaku, Bapak Mislam Ismail (Alm) dan Ibu Nunsiyanti.

Adikku Ahmad Taufik Fadhani, Dicky Arya Novandi,

Artha Maulana Rahman,

serta untuk keluargaku dan almamaterku tercinta.

MOTTO

"Sebaik-baik manusia diantaramu adalah yang paling banyak manfaatnya bagi orang lain." (H.R. Bukhari)

UCAPAN TERIMAKASIH

Puji syukur penulis ucapkan kehadiran Tuhan Yang Maha Esa, karena atas rahmat dan hidayah-Nya tesis ini dapat diselesaikan.

Tesis dengan judul “Prakiraan Cuaca Kota Bandar Lampung Dengan Menggunakan *Random Forest* Dan C4.5” merupakan salah satu syarat untuk memperoleh gelar Magister Teknik di Universitas Lampung.

Dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Bapak Dr. Eng. Ir. Helmy Fitriawan, S.T., M.Sc. selaku Dekan Fakultas Teknik Unila sekaligus Penguji tesis ini yang telah memberikan arahan dan bantuannya dalam proses penyelesaian tesis ini.
2. Ibu Herlinawati, S.T., M.T. selaku Ketua Jurusan Teknik Elektro Unila.
3. Bapak Misfa Susanto, S.T., M.Sc., P.hD. selaku Ketua Program Studi Magister Teknik Elektro dan Anggota Penguji tesis ini yang telah memberikan arahan dan bantuannya dalam proses penyelesaian tesis ini.
4. Ibu Dr. Ir. Sri Ratna Sulistiyanti, M.T. selaku pembimbing utama yang telah bersedia untuk memberikan bimbingan, saran, masukan dalam proses penyelesaian tesis ini.
5. Bapak Dr. Eng. F. X. Arinto Setyawan, S.T., M.T. selaku pembimbing pendamping yang telah bersedia untuk memberikan bimbingan, saran, masukan dalam proses penyelesaian tesis ini.
6. Bapak dan Ibu dosen Magister Teknik Elektro yang sudah memberikan ilmu selama proses KBM berlangsung.

7. Bapak dan Ibu staff administrasi Magister Teknik Elektro Unila.
8. Teman-teman angkatan 2019, Yudi Eka Putra, M Ridho Yoga Permono, Ubaidah, Sitronella Nurfitriani Hasim, Sri Suryadi dan Nadia Julian Putri yang telah membantu dan selalu memberikan semangat.
9. Kakak tingkat dan adik tingkat yang telah membantu dan selalu memberikan semangat.

Bandar Lampung, 19 Juni 2023

Rahma Ferika Shaumi

DAFTAR ISI

	Halaman
DAFTAR GAMBAR	iii
DAFTAR TABEL	v
I. PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Tujuan Penelitian	4
1.3 Manfaat Penelitian	4
1.4 Perumusan Masalah	4
1.5 Batasan Masalah	5
1.6 Hipotesis.....	5
1.7 Sistematika Penulisan	5
II. TINJAUAN PUSTAKA	
2.1 BMKG.....	7
2.1.1 Meteorologi.....	7
2.1.2 Klimatologi	8
2.1.3 Geofisika	8
2.2 Prakiraan Cuaca	9
2.3 <i>Data Mining</i>	10
2.4 <i>Machine Learning</i>	11
2.5 <i>Random Forest</i>	12
2.6 <i>C4.5</i>	16
2.7 <i>Mean Absolute Error</i>	19
2.8 <i>Root Mean Square Error</i>	19
2.9 WEKA.....	20
2.10 <i>Cross-Validation</i>	21
2.11 <i>Percentage Split</i>	22
III. METODE PENELITIAN	
3.1 Waktu dan Tempat Penelitian	23
3.2 Alat dan Bahan	23
3.3 Metode yang Digunakan	23
3.3.1 Garis Besar Metode yang Digunakan	24
3.3.2 Diagram Alir	24

IV. HASIL DAN PEMBAHASAN	
4.1 Pengujian <i>Cross-validation Folds 5</i>	32
4.2 Pengujian <i>Cross-validation Folds 10</i>	35
4.3 Pengujian <i>Cross-validation Folds 15</i>	37
4.4 Pengujian <i>Percentage Split 25%</i>	40
4.5 Pengujian <i>Percentage Split 66%</i>	43
4.6 Pengujian <i>Percentage Split 90%</i>	46
4.7 Analisis Data Hasil Pengujian.....	49
V. KESIMPULAN DAN SARAN	
5.1 Kesimpulan	52
5.2 Saran	52

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR GAMBAR

Gambar	Halaman
2.1 Diagram <i>random forest</i>	13
2.2 Contoh pohon keputusan dengan <i>random forest</i>	15
2.3 Pohon keputusan node 1	18
2.4 <i>Software</i> WEKA.....	20
2.5 Ilustrasi <i>k-fold cross-validation</i>	21
2.6 Ilustrasi <i>percentage split</i>	22
3.1 Diagram alir penelitian	25
3.2 Diagram proses prakiraan cuaca	26
3.3 Sampel dataset.....	27
3.4 <i>Preprocessing</i> data	27
3.5 Grafik data seluruh atribut	28
3.6 Hasil proses <i>replace missing values</i>	28
3.7 <i>Training</i> dataset <i>random forest</i>	29
3.8 Evaluasi <i>training</i> data <i>random forest</i>	29
3.9 <i>Training</i> dataset C4.5	30
3.10 Evaluasi <i>training</i> data C4.5.....	31
4.1 <i>Cross-validation folds</i> 5 dengan <i>random forest</i>	32
4.2 Evaluasi <i>cross-validation folds</i> 5 dengan <i>random forest</i>	33
4.3 <i>Cross-validation folds</i> 5 dengan C4.5.....	33
4.4 Evaluasi <i>cross-validation folds</i> 5 dengan C4.5.....	34
4.5 <i>Cross-validation folds</i> 10 dengan <i>random forest</i>	35
4.6 Evaluasi <i>cross-validation folds</i> 10 dengan <i>random forest</i>	35
4.7 <i>Cross-validation folds</i> 10 dengan C4.5	36
4.8 Evaluasi <i>cross-validation folds</i> 10 dengan C4.5.....	37
4.9 <i>Cross-validation folds</i> 15 dengan <i>random forest</i>	38
4.10 Evaluasi <i>cross-validation folds</i> 15 dengan <i>random forest</i>	38
4.11 <i>Cross-validation folds</i> 15 dengan C4.5	39
4.12 Evaluasi <i>cross-validation folds</i> 15 dengan C4.5.....	40
4.13 <i>Percentage split</i> 25% dengan <i>random forest</i>	41
4.14 Evaluasi <i>percentage split</i> 25% dengan <i>random forest</i>	41
4.15 <i>Percentage split</i> 25% dengan C4.5	42
4.16 Evaluasi <i>percentage split</i> 25% dengan C4.5	43
4.17 <i>Percentage split</i> 66% dengan <i>random forest</i>	44
4.18 Evaluasi <i>percentage split</i> 66% dengan <i>random forest</i>	44
4.19 <i>Percentage split</i> 66% dengan C4.5.....	45
4.20 Evaluasi <i>percentage split</i> 66% dengan C4.5	46
4.21 <i>Percentage split</i> 90% dengan <i>random forest</i>	47
4.22 Evaluasi <i>percentage split</i> 90% dengan <i>random forest</i>	47
4.23 <i>Percentage split</i> 90% dengan C4.5.....	48
4.24 Evaluasi <i>percentage split</i> 90% C4.5.....	49

4.25 Hasil pengujian <i>cross-validation</i>	50
4.26 Hasil pengujian <i>percentage split</i>	51

DAFTAR TABEL

Tabel	Halaman
2.1 Dataset untuk <i>random forest</i>	14
2.2 Dataset untuk C4.5 dengan kasus keputusan bermain sepak bola	17
2.3 Hasil perhitungan node ke 1	18

I. PENDAHULUAN

1.1. Latar Belakang

Dalam perkembangan teknologi yang semakin berkembang pesat dari waktu ke waktu sehingga sangat membantu pekerjaan ataupun kegiatan manusia agar dapat berjalan dengan baik dan lancar. Salah satu teknologi yang membantu kelancaran pekerjaan manusia teknologi di bidang Meteorologi, Klimatologi dan Geofisika, yaitu pemanfaatan teknologi untuk memprakirakan cuaca di suatu tempat dalam kurun waktu tertentu.

Prakiraan cuaca biasanya juga dapat disebut dengan ramalan cuaca, menurut KBBI prakiraan cuaca adalah prakiraan keadaan cuaca yang mungkin terjadi di suatu daerah. Cuaca menurut Kamus Besar Bahasa Indonesia (KBBI) yaitu keadaan udara seperti temperatur, cahaya matahari, kelembapan, kecepatan angin pada suatu daerah tertentu dengan jangka waktu yang terbatas [1]. Keadaan cuaca dapat dengan mudah berubah-ubah pada waktu tertentu dan di wilayah yang relatif tidak luas. Keadaan cuaca ini dapat berubah-ubah dikarenakan ada beberapa yang mempengaruhi, antara lain angin, curah hujan, suhu, tekanan udara, kelembapan udara dan lain-lain.

Prakiraan cuaca sangat bermanfaat dan dapat berperan penting dalam pekerjaan manusia, seperti memungkinkan untuk seseorang atau instansi mengambil suatu tindakan atau merencanakan terhadap kegiatannya maupun untuk mencoba mencegah dari berbagai kejadian yang tidak diinginkan seperti bencana alam, banjir dan lain-lain. Sehingga dampaknya dapat diminimalisir dan tidak menimbulkan kerugian yang lebih banyak lagi.

Prakiraan cuaca juga dapat bermanfaat dalam sistem transportasi. Terdapat beberapa alat transportasi yang memanfaatkan hasil prakiraan cuaca yang digunakan untuk proses berjalannya suatu sistem transportasi, seperti pesawat udara [2]. Jenis moda transportasi ini membutuhkan cuaca yang cerah demi berjalan lancar serta demi keselamatan awak dan penumpangnya. Ketika sedang beroperasi karena jenis transportasi ini menggunakan navigasi, jika cuaca sedang tidak baik maka navigasi dapat terganggu yang dapat berakibat perjalanan menjadi lambat atau ditunda karena tidak berani berlayar [3].

Kemudian cuaca juga dapat mempengaruhi pada bidang pariwisata. Apabila cuaca cerah maka pariwisata dapat dinikmati oleh wisatawan, sehingga penghasilan dari pariwisata dapat meningkat. Kemudian apabila cuaca dalam keadaan kurang baik maka pariwisata dapat menjadi sepi dan menurunkan pendapatan dari sektor pariwisata [4].

Prakiraan cuaca juga bermanfaat bagi bidang pertanian, faktor yang terdapat dalam cuaca penting bagi pertumbuhan tanaman dan waktu tanam, seperti waktu penanaman tembakau. Hasil produksi tanaman tembakau sangat dipengaruhi oleh cuaca [5].

Di Indonesia memiliki badan yang bertugas untuk memberikan informasi mengenai prakiraan cuaca, seperti di Provinsi Lampung terdapat Badan Meteorologi, Klimatologi dan Geofisika (BMKG) Provinsi Lampung. BMKG Provinsi Lampung merupakan lembaga pemerintah non departemen yang memberikan pelayanan data dan informasi dibidang meteorologi, klimatologi dan geofisika dengan menggunakan teknologi AWS (*Automatic Weather Station*), satelit dan lain-lain. Data prakiraan cuaca yang disajikan dalam waktu beberapa hari kedepan melalui kanal *website* BMKG Lampung ataupun media sosial serta beberapa media lain yang bekerja sama dengan BMKG Provinsi Lampung [6].

Saat ini untuk mendapatkan hasil suatu informasi cuaca terdapat serangkaian proses dengan menggunakan beberapa peralatan oleh peramal cuaca (*forecaster*) dengan membutuhkan beberapa waktu [7]. Kemudian untuk memenuhi kebutuhan informasi masyarakat terkait hal tersebut dengan memanfaatkan perkembangan teknologi yang lebih mudah, murah dan dapat dijangkau oleh masyarakat maka prakiraan cuaca dapat dihasilkan dengan menggunakan algoritma yang dipakai untuk penelitian ini yaitu *Random forest* dan *C4.5*.

Random forest memiliki kelebihan yaitu dapat meningkatkan hasil akurasi apabila terdapat data yang hilang dan untuk *resisting outliers*, serta efisien untuk penyimpanan sebuah data. *Random forest* juga memiliki fitur proses seleksi yang mampu mengambil fitur terbaik sehingga dapat meningkatkan performa terhadap model klasifikasi [8]. Kemudian juga memiliki kekurangan yaitu untuk pembelajaran dapat berjalan dengan lambat, ini terjadi karena tergantung dengan parameter yang digunakan serta tidak dapat memperbaiki model yang dihasilkan secara berulang. Kelebihan *C4.5* yaitu dapat membuat pohon keputusan sehingga menjadi efisien, dimana pohon keputusan menangani atribut tipe diskrit dan diskrit-numerik, serta mudah untuk diinterpretasikan dan memiliki tingkat akurasi yang dapat diterima. *C4.5* memiliki kelemahan yaitu terdapat pada skalabilitas *data training* hanya dapat digunakan dan disimpan secara keseluruhan pada waktu yang bersamaan di memori [9].

Berdasarkan penelitian yang sudah dilakukan kinerja RF (*Random Forest*) memiliki kinerja yang baik dibandingkan dengan beberapa model seperti *K-Nearest Neighbors*, *Least Medium Square Regression*, jaringan RBF (*Radial Basic Function*) dan jaringan saraf MLP (*Multilayer Perceptron*) [10]. Pada penelitian yang lain RF mendapatkan hasil lebih baik dan lebih akurat dari ANN (*Artificial Neural Network*) dan SVM (*Support Vector Machine*) [11].

Algoritma *C4.5* telah digunakan dalam beberapa penelitian sebelumnya seperti Implementasi *Data mining* Pada Dataset Prakiraan Cuaca Menggunakan *C4.5* dengan memperoleh tingkat akurasi sebesar 81,94% [12]. Kemudian pada penelitian yang lain penggunaan *C4.5* dengan Naïve Bayes didapatkan hasil bahwa *C4.5* mendapatkan hasil akurasi yang lebih besar dari Naïve Bayes [13]. Penelitian selanjutnya yaitu pada penelitian analisis perbandingan Naïve Bayes, KNN dan *C4.5* yang menghasilkan *C4.5* memiliki akurasi tertinggi dari ketiganya [14].

Pada penelitian ini akan melakukan prakiraan cuaca dengan menggunakan dua algoritma yaitu *random forest* dan *C4.5* dengan menggunakan dataset dari Stasiun Meteorologi Maritim Panjang sebanyak 365 data dengan 6 atribut berupa 1 atribut keluaran yaitu intensitas hujan dan 5 atribut input yaitu *temperature* rata-rata, kelembapan rata-rata, curah hujan, lamanya penyinaran matahari dan kecepatan angin rata-rata. Hasil dari penelitian ini yaitu dapat menghasilkan prakiraan Kota Bandar Lampung serta dapat membandingkan hasil evaluasi algoritma *random forest* dan *C4.5* dengan parameter *error*.

1.2. Tujuan Penelitian

Tujuan dari penelitian ini yaitu:

1. Mengaplikasikan algoritma *Random forest* dan *C4.5* untuk mamprakirakan cuaca di Kota Bandar Lampung.
2. Menghasilkan prakiraan cuaca dengan mengevaluasi hasil tingkat *error* pada algoritma yang digunakan.

1.3. Manfaat Penelitian

Manfaat dari penelitian ini yaitu:

1. Mengetahui keefektifan metode prakiraan cuaca yang digunakan dalam penelitian ini.
2. Mendapatkan alternatif prakiraan cuaca selain dari BMKG Provinsi Lampung.

1.4. Perumusan Masalah

Perumusan masalah pada penelitian ini yaitu:

1. Bagaimana cara mengaplikasikan Algoritma *Random forest* dan *C4.5* dengan hasil yang baik?
2. Bagaimana cara mengukur nilai *error* terhadap metode yang digunakan?

1.5. Batasan Masalah

Batasan masalah dalam penelitian ini yaitu:

1. Membahas prakiraan cuaca di Kota Bandar Lampung.
2. Membahas metode Algoritma *Random forest* dan *C4.5* yang digunakan dalam memprakirakan cuaca di Kota Bandar Lampung.
3. Dataset prakiraan cuaca dari Stasiun Meteorologi Maritim Panjang Kota Bandar Lampung.

1.6. Hipotesis

Metode yang digunakan untuk memprakirakan cuaca di Provinsi Lampung dengan menggunakan algoritma yaitu Algoritma *C4.5* dan Algoritma *Random forest*. Kedua algoritma ini diharapkan mampu menghasilkan prakiraan cuaca dengan ketepatan yang baik. Dataset menggunakan data prakiraan cuaca Kota Bandar Lampung, kemudian hasil dari penelitian ini nantinya akan mendapatkan hasil prakiraan cuaca dan dapat membandingkan hasil dari metode RF dan *C4.5*.

1.7. Sistematika Penulisan

Untuk memperjelas dalam penulisan serta pemahaman mengenai materi dalam penelitian ini, maka dibagi menjadi 5 bab yaitu:

BAB I PENDAHULUAN

Memuat latar belakang, tujuan penelitian, manfaat penelitian, perumusan masalah, batasan masalah, hipotesis dan sistematika laporan.

BAB II TINJAUAN PUSTAKA

Menjelaskan teori-teori yang mendukung tentang proses memprakirakan cuaca, algoritma *Random forest* serta algoritma *C4.5*.

BAB III METODE PENELITIAN

Memaparkan waktu dan tempat penelitian, alat dan bahan yang digunakan, garis besar metode yang digunakan, serta diagram alir metode yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Membahas metode yang digunakan, proses dalam penelitian dan perhitungan nilai *error* dari hasil penelitian.

BAB V KESIMPULAN DAN SARAN

Berisi kesimpulan yang didapat dari proses penelitian, serta saran-saran untuk pengembangan lebih lanjut.

II. TINJAUAN PUSTAKA

2.1. BMKG

Badan Meteorologi, Klimatologi dan Geofisika merupakan sebuah Lembaga pemerintah non departemen. Tugas dari BMKG yaitu melaksanakan tugas pemerintahan di bidang Meteorologi, Klimatologi, Kualitas Udara dan Geofisika sesuai dengan ketentuan perundang-undangan yang berlaku [15].

2.1.1. Meteorologi

Meteorologi berasal dari Bahasa Yunani yaitu *meteoros* yang memiliki arti ruang atas/atmosfer dan *logos* yang berarti sebuah kajian atau ilmu. Sehingga dapat diartikan yaitu ilmu yang membahas tentang atmosfer. Adapun arti lain dari meteorologi yaitu ilmu yang mempelajari proses fisis dan gejala cuaca yang terjadi di dalam atmosfer yang biasanya terjadi pada lapisan bawah (troposfer) [16]. Kemudian meteorologi juga dapat diartikan dengan sebuah ilmu yang mempelajari masalah atmosfer, misalnya udara, suhu, angin, cuaca, dan beberapa sifat fisika serta sifat kimia atmosfer yang digunakan dalam prakiraan cuaca.

Meteorologi sangat erat kaitannya dengan kegiatan pengamatan atau observasi. Observasi ini dilakukan demi mendapatkan data pada perubahan cuaca dengan parameter-parameter tertentu yang kemudian data tersebut dianalisis sehingga menghasilkan sebuah prakiraan cuaca. Parameter-parameter dalam meteorologi dapat dilihat secara langsung dengan mata/panca indera maupun menggunakan peralatan tergantung dengan tingkat kesulitan. Parameter-parameter yang digunakan dalam ilmu meteorologi yaitu:

1. Tekanan
2. Suhu/temperatur
3. Angin
4. Awan
5. Penglihatan mendatar
6. Penyinaran matahari
7. Kelembapan
8. Keadaan tanah
9. Penguapan
10. Cuaca

2.1.2. Klimatologi

Klimatologi berasal dari Bahasa Yunani yaitu *Klima* yang memiliki arti tempat, wilayah, zoba dan dapat diartikan dengan slope planet bumi yang berhubungan dengan lintang tempat atau kemiringan khayal dari bumi. Kemudian kata logos yang berarti ilmu/ mempelajari. Sehingga klimatologi dapat berarti ilmu yang mencari sebuah gambaran dan penjelasan iklim dan cuaca di bumi dapat berbeda-beda, dan menjelaskan tentang hubungan antara iklim dengan kehidupan manusia.

Klimatologi juga berarti ilmu yang mempelajari jenis-jenis iklim yang terdapat di bumi dan berbagai faktor penyebabnya [16].

2.1.3. Geofisika

Geofisika memiliki pengertian yaitu ilmu bumi yang mempelajari bumi dengan menggunakan prinsip-prinsip atau dengan menggunakan kaidah fisika [17]. Prinsip-prinsip fisika yang dipelajari yaitu bentuk bumi, medan magnet dan gravitasi (medan potensial bumi) serta reaksi terhadap suatu

gaya. Terdapat beberapa bidang ilmu geofisika seperti udara, geofisika bumi padat dan laut.

2.2. Prakiraan Cuaca

Dalam kehidupan sehari-hari cuaca dapat dengan mudah berubah-ubah, hal ini dikarenakan memiliki beberapa faktor seperti tekanan udara, suhu, kelembapan udara, angin serta curah hujan. Cuaca sendiri merupakan keadaan atmosfer yang sifatnya berubah-ubah pada waktu tertentu dari waktu ke waktu [18]. Cuaca ini terjadi pada zona yang tidak luas dan dalam jangka waktu yang singkat. Biasanya cuaca dihitung dengan satuan hari yang mencakup wilayah yang sempit, serta dalam hari berikutnya dapat terjadi perubahan.

Cuaca memiliki beberapa unsur-unsur, yaitu:

1. Suhu Udara

Suhu udara merupakan rata-rata ukuran energi kinetik dari sebuah pergerakan molekul-molekul. Suhu udara dapat diukur dengan menggunakan alat termometer. Jika keadaan suhu udara pada sepanjang hari dapat dianalisa dengan menggunakan termograf dan termogram.

2. Tekanan Udara

Tekanan udara merupakan kerapatan molekul udara. Yang menimbulkan tekanan ke berbagai arah dan disebut dengan hidrostatis. Tekanan udara ini dapat diukur dengan peralatan yang disebut Barometer. Tekanan di suatu wilayah dapat berbeda-beda. Jika terjadi perbedaan tekanan udara maka akan terjadi Gerakan udara dari daerah yang memiliki tekanan udara yang tinggi ke arah yang memiliki tekanan udara yang rendah dan Gerakan ini dapat disebut dengan angin. Suatu wilayah yang semakin tinggi dari permukaan laut, maka akan semakin rendah tekanan udaranya.

3. Kelembapan Udara

Kelembapan udara merupakan tingkat kebasahan udara dikarenakan didalam udara selalu terkandung uap air. Kandungan udara yang hangat

memiliki uap air yang lebih banyak dibandingkan dengan kandungan uap air di dalam udara yang dingin. Terdapat dua jenis kelembapan udara yaitu kelembapan udara mutlak dan kelembapan udara relatif.

4. Awan

Awan merupakan sebuah gumpalan uap air yang terbentuk oleh siklus daur air yang terus terjadi atau biasanya disebut dengan siklus hidrologi. Kristal es yang membentuk butiran awan bukanlah air murni, tetapi titik-titik air yang mengumpul di sekeliling inti kondensasi. Inti kondensasi ini berupa kristal-kristal garam yang berasal dari deburan ombak, debu, dan asap. Terdapat 4 jenis awan yaitu awan tinggi, awan sedang, awan rendah dan awan vertikal.

5. Angin

Angin merupakan udara yang bergerak oleh rotasi bumi dan karena adanya perbedaan tekanan udara yaitu dari tekanan yang tinggi ke tekanan yang rendah di wilayah sekitarnya.

6. Hujan

Hujan merupakan suatu peristiwa jatuhnya suatu cairan dari atmosfer yang berbentuk cair/beku ke arah permukaan bumi. Awan yang semakin padat titik-titik airnya akan bergabung menjadi satu dengan yang lain sehingga menjadi tetesan air [18].

Prakiraan Cuaca dapat diartikan sebagai rangkuman informasi suatu kondisi cuaca harian hingga mingguan. Prakiraan cuaca dapat menggunakan semua jenis unsur cuaca [19].

2.3. Data mining

Data mining merupakan sebuah proses untuk mencari informasi atau pola menarik pada data terpilih dengan menggunakan metode maupun Teknik

tertentu. Metode atau teknik dalam *Data mining* bervariasi dan sangat bergantung pada tujuan [20]. *Data mining* merupakan pemisahan dari model informasi yang bermanfaat dalam penyimpanan database data [21].

Data mining dibagi menjadi beberapa berdasarkan tugas yang dapat dilakukan yaitu:

1. Deskripsi

Deskripsi merupakan cara untuk mencoba menemukan sebuah gambaran pola dan trend yang tersembunyi di dalam data.

2. Prediksi

Prediksi akan memberikan nilai dari hasil yang akan datang.

3. Estimasi

Variabel target estimasi lebih condong ke arah numerik dibandingkan dengan kategori.

4. Klasifikasi

Klasifikasi variable tujuannya bersifat kategorik

5. *Clustering*

Teknik pengelompokkan *record* data, pengamatan dalam kelas yang memiliki tingkat kemiripan satu dengan yang lain dan memiliki ketidakmiripan dengan *record* lain dalam *cluster*.

6. Asosiasi

Mengidentifikasi hubungan antara beberapa peristiwa yang terjadi dalam satu waktu.

2.4. Machine Learning

Machine learning adalah salah satu cabang ilmu kecerdasan buatan, yang mempelajari tentang bagaimana komputer dapat belajar dari data untuk meningkatkan kecerdasannya [22]. *Machine learning* dapat juga diartikan sebuah tipe dari kecerdasan buatan yang menyediakan komputer dengan

kemampuan untuk belajar dari data, tanpa secara eksplisit harus mengikuti instruksi terprogram [23].

Selain itu *machine learning* merupakan komputer yang mempunyai kemampuan untuk melakukan belajar dari pengalaman terhadap tugas-tugasnya dan mengalami peningkatan kerja [24].

Machine learning mempunyai fokus untuk pengembangan sistem yang dapat belajar sendiri untuk memutuskan sesuatu, tanpa harus diprogram oleh manusia berulang-ulang. Mesin dapat beradaptasi dengan perubahan yang terjadi dan dapat menemukan aturan untuk perilaku yang optimal dalam pengambilan keputusan. Perbedaan antara pemrograman tradisional dengan *machine learning*, pada pemrograman tradisional data dan program yang dijalankan pada komputer untuk dapat menghasilkan sebuah keluaran, sedangkan pada *machine learning* data dan keluaran dijalankan pada komputer untuk membuat sebuah program, yang kemudian program tersebut dapat digunakan dalam pemrograman tradisional [25].

2.5. *Random forest*

Random forest secara rekursif membagi kumpulan data dengan memilih, pada setiap simpul, variabel dan ambang batas yang memaksimalkan metrik ketidaksamaan (misalnya, perolehan informasi) hingga kriteria penghentian tercapai (misalnya, jumlah sampel kumpulan data berada di bawah jumlah yang ditentukan, dll) [26].

Untuk membuat sebuah prediksi:

$$f(x) = \frac{1}{K} \sum_{k=1}^K T_k(x) \quad (1)$$

Dimana:

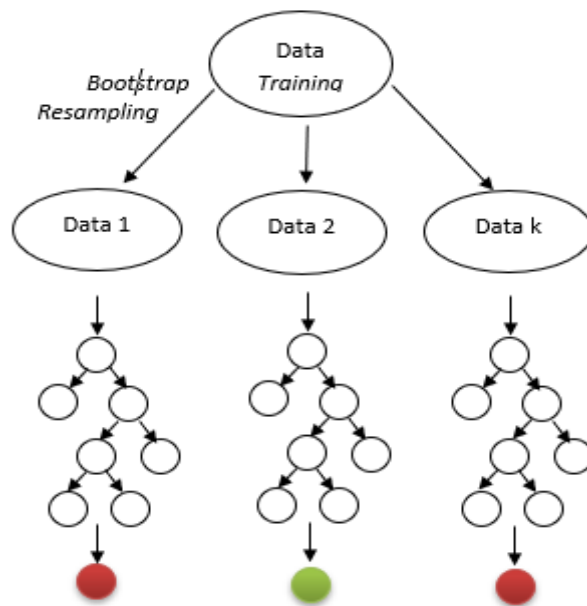
K = Jumlah pohon

f = Jumlah input variabel yang dipilih secara acak secara berurutan

T = Variabel Keluaran

x = Angka prediksi yang baru

Metode ini banyak digunakan di berbagai bidang peramalan deret waktu, seperti biostatistik, pemantauan iklim, perencanaan di industri energi, dan peramalan cuaca [10]. *Random forest* adalah metode pembelajaran *ensemble* yang menghasilkan banyak regresi dan menggabungkan hasilnya. Model beroperasi pada pola siklus musiman deret waktu yang menyederhanakan masalah peramalan terutama ketika deret waktu menunjukkan nonstasioneritas, heteroskedastisitas, tren, dan siklus musiman ganda. Keuntungan utama dari model ini adalah kemampuannya untuk melakukan generalisasi, validasi silang bawaan dan sensitivitas yang rendah terhadap nilai parameter [10].



Gambar 2.1. Diagram *random forest* [27]

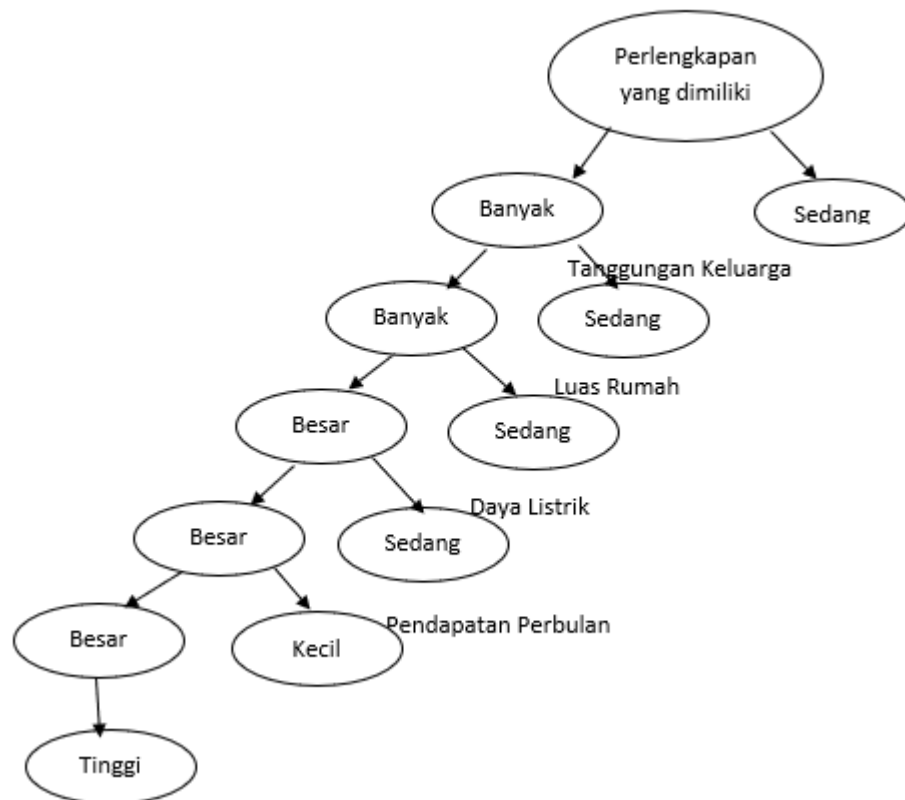
Cara kerja *random forest* berdasarkan pada Gambar 2.1. yaitu apabila memiliki sebuah dataset maka tidak dapat langsung mendapatkan data *training* dari pohon ini, tetapi akan mengambil sampel secara *bootstrap* yaitu sampel dengan replacement untuk mendapatkan data1 kemudian dari pohon 1 dapat

dijalankan pohon klasifikasi untuk mendapatkan pohon yang pertama. Kemudian diulangi untuk pembuatan pohon ke 2 sampai pohon k. Umumnya k yang digunakan dalam *random forest* cukup besar. Untuk melakukan prediksi dari setiap pohon akan didapatkan prediksi untuk setiap individu dengan prediksi yang berbeda. Kemudian prediksi kelas akhir bagi individu tersebut digabungkan dengan cara mengakreditasi suara terbanyak (*majority vote*). *Random forest* memiliki perbedaan yaitu pada proses *splitting* dalam pembentukan pohon. Dataset yang akan dilakukan *split* menggunakan variabel terbaik, tidak menggunakan semua variabel sehingga pohon yang dihasilkan akan berbeda-beda.

Tabel 2.1. Dataset untuk *random forest*

NO	Tanggungan Keluarga	Luas Rumah	Pendapatan perbulan	Daya Listrik	Perlengkapan yang Dimiliki	Penggunaan Listrik
1	Banyak	Besar	Besar	Sedang	Banyak	Tinggi
2	Kecil	Standar	Besar	Sedang	Banyak	Tinggi
3	Sedang	Besar	Besar	Sedang	Sedang	Tinggi
4	Banyak	Standar	Kecil	Rendah	Banyak	Sedang

Tabel 2.1. adalah dataset untuk membuat *bootstrap* dengan 4 amatan dengan jumlah yang sama maka secara acak memilih sampel dari dataset. Misalnya amatan pertama menjadi amatan ke dua dalam *bootstrap*, amatan ke empat bisa menjadi amatan ke pertama dalam *bootstrap* dan amatan ke empat dapat menjadi amatan ke tiga. Pada tahap ini ada amatan yang digunakan dua kali dan juga amatan yang tidak dipakai. Langkah selanjutnya yaitu membuat pohon keputusan menggunakan *bootstrap* dataset dalam setiap langkahnya, artinya proses *splitting* data hanya menggunakan beberapa variabel.



Gambar 2.2. Contoh pohon keputusan dengan *random forest*

Gambar 2.2. merupakan contoh pohon keputusan dengan menggunakan data sampel untuk mendapatkan prediksi penggunaan listrik rumah tangga. Untuk data *training* dipilih secara acak yaitu perlengkapan yang dimiliki, apabila perlengkapan yang dimiliki banyak maka selanjutnya menggunakan data banyaknya tanggungya keluarga, kemudian mendapatkan hasil banyak. Selanjutnya yaitu luas rumah besar, daya listrik besar dan pendapatan perbulan besar. Hasil prediksi akhir didapatkan berdasarkan suara terbanyak yaitu penggunaan listrik adalah tinggi. Kemudian seterusnya untuk pohon keputusan hingga jumlah k.

2.6. C4.5

Pada dasarnya algoritma C4.5 adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan [12]. Algoritma ini merupakan metode klasifikasi dan prediksi yang sangat terkenal dan kuat [28]. Pohon keputusan berfungsi untuk menemukan hubungan tersembunyi antara beberapa langkah dalam membuat pohon keputusan yaitu memilih akar dari atribut yang digunakan dengan berdasarkan pada nilai *gain* tertinggi. Kemudian membuat cabang untuk masing-masing nilai. Selanjutnya membagi setiap kasus dalam cabang yang berdasarkan perhitungan nilai tertinggi. Kemudian mengulangi proses dalam setiap cabang sehingga semua kasus memiliki kelas yang sama. Sebelum menghitung nilai *gain*, maka harus menghitung nilai *entropy* dahulu

$$Entropy(S) = \sum_{i=1}^N -p_i \log_2 p_i \quad (2)$$

Dimana:

S = Himpunan kasus

N = Jumlah partisi S

Pi = Proporsi Si terhadap S

Kemudian menentukan *gain*:

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * entropy(S_i) \quad (3)$$

Dimana:

S = Himpunan kasus

A = Fitur

N = Jumlah partisi atribut A

|Si| = Proporsi Si terhadap S

|S| = Jumlah kasus dalam S

Tabel 2.2. Dataset untuk C4.5 dengan kasus keputusan bermain sepak bola

<i>No</i>	<i>Outlook</i>	<i>Temperature</i>	<i>Humidity</i>	<i>Windy</i>	<i>Play</i>
1	<i>Sunny</i>	<i>Hot</i>	<i>High</i>	<i>False</i>	<i>No</i>
2	<i>Sunny</i>	<i>Hot</i>	<i>High</i>	<i>True</i>	<i>No</i>
3	<i>Cloudy</i>	<i>Hot</i>	<i>High</i>	<i>False</i>	<i>Yes</i>
4	<i>Rainy</i>	<i>Mild</i>	<i>High</i>	<i>False</i>	<i>Yes</i>
5	<i>Rainy</i>	<i>Cool</i>	<i>Normal</i>	<i>False</i>	<i>Yes</i>
6	<i>Rainy</i>	<i>Cool</i>	<i>Normal</i>	<i>True</i>	<i>Yes</i>
7	<i>Cloudy</i>	<i>Cool</i>	<i>Normal</i>	<i>True</i>	<i>Yes</i>
8	<i>Sunny</i>	<i>Mild</i>	<i>High</i>	<i>False</i>	<i>No</i>
9	<i>Rainy</i>	<i>Cool</i>	<i>Normal</i>	<i>False</i>	<i>Yes</i>
10	<i>Rainy</i>	<i>Mild</i>	<i>Normal</i>	<i>False</i>	<i>Yes</i>
11	<i>Sunny</i>	<i>Mild</i>	<i>Normal</i>	<i>True</i>	<i>Yes</i>
12	<i>Cloudy</i>	<i>Mild</i>	<i>High</i>	<i>True</i>	<i>Yes</i>
13	<i>Cloudy</i>	<i>Hot</i>	<i>Normal</i>	<i>False</i>	<i>Yes</i>
14	<i>Rainy</i>	<i>Mild</i>	<i>High</i>	<i>True</i>	<i>No</i>

Pertama-tama menghitung jumlah kasus untuk keputusan *yes* dan untuk jumlah kasus untuk keputusan *no*. Kemudian *entropy* semua kasus yang dibagi berdasarkan seluruh atribut yaitu *outlook*, *temperature*, *humidity* dan *windy*. Setelah itu dilakukan perhitungan dengan masing-masing atribut. Cara perhingan *entropy* dengan Persamaan 2:

$$Entropy = \left(-\frac{4}{14} * \log_2 \left(\frac{4}{14} \right) \right) + \left(-\frac{10}{14} * \log_2 \left(\frac{10}{14} \right) \right) = 0,86$$

Kemudian menghitung *gain* pada baris *outlook* dengan menggunakan Persamaan 3:

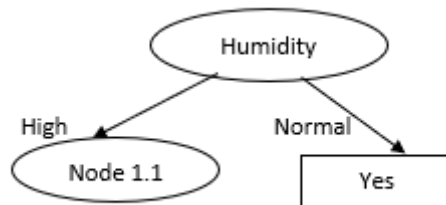
$$Gain(Total, Outlook) = 0,86 - \left(\frac{4}{14} * 0 \right) - \left(\frac{5}{14} * 0,723 \right) - \left(\frac{15}{14} * 0,97 \right) = 0,25$$

Berdasarkan perhitungan atribut *gain* tertinggi yaitu *humidity* yaitu sebesar 0,37. Maka *humidity* akan menjadi node akar dengan nilai atribut *humidity* yaitu *high* dan *normal*. Dari nilai atribut tersebut mengklarifikasi kasus menjadi 1 yaitu keputusan *yes*, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Tetapi untuk perhitungan nilai *high* perlu dilakukan perhitungan.

Tabel 2.3 Hasil perhitungan node ke 1

Node			Jumlah Kasus (S)	Tidak (S ₁)	Ya (S ₂)	Entropy	Gain
1	Total		14	4	10	0,86	
	<i>Outlook</i>						0,25
		<i>Cloudy</i>	4	0			
		<i>Rainy</i>	5	1	0,72		
		<i>Sunny</i>	5	3	0,97		
	<i>Temperature</i>						0,18
		<i>Cool</i>	4	0	4	0	
		<i>Hot</i>	4	2	2	1	
		<i>Mild</i>	6	2	4	0,91	
	<i>Humidity</i>						0,37
		<i>High</i>	7	4	3	0,98	
		<i>Normal</i>	7	0	7		
	<i>Windy</i>						0,0059
		<i>False</i>	8	2	6	0,81	
		<i>True</i>	6	4	2	0,91	

Berdasarkan Tabel 2.3. maka pohon keputusannya adalah:



Gambar 2.3. Pohon keputusan node 1

Berdasarkan hasil perhitungan *humidity* menjadi akar node, kemudian *humidity* memiliki isian yaitu *high* dan *normal*. Ketika *humidity* maka akan *yes*, tetapi ketika *high* akan melalui proses berikutnya dengan menggunakan persamaan-persamaan sebelumnya.

2.7. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) merupakan salah satu metode yang digunakan untuk mengukur tingkat keakuratan sebuah model peramalan. Nilai MAE menunjukkan rata-rata kesalahan absolute [29].

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (4)$$

Dimana:

n = jumlah data

f_i = nilai hasil peramalan

y_i = nilai sebenarnya

2.8. Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) merupakan besarnya tingkat kesalahan hasil prediksi, dimana semakin kecil (mendekati 0) nilai RMSE maka hasil prediksi akan semakin akurat.

Kriteria yang digunakan untuk mengukur kebaikan model setelah diperoleh suatu model adalah *root mean square error* (RMSE) [30]. RMSE merupakan alat seleksi model berdasarkan pada *error* hasil estimasi. *Error* yang ada menunjukkan seberapa besar perbedaan hasil estimasi dengan nilai yang akan diestimasi. Nilai ini akan digunakan untuk menentukan model mana yang terbaik. Definisi RMSE dapat ditulis sebagai berikut [31]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i - F_i)^2}{n}} \quad (5)$$

Dimana:

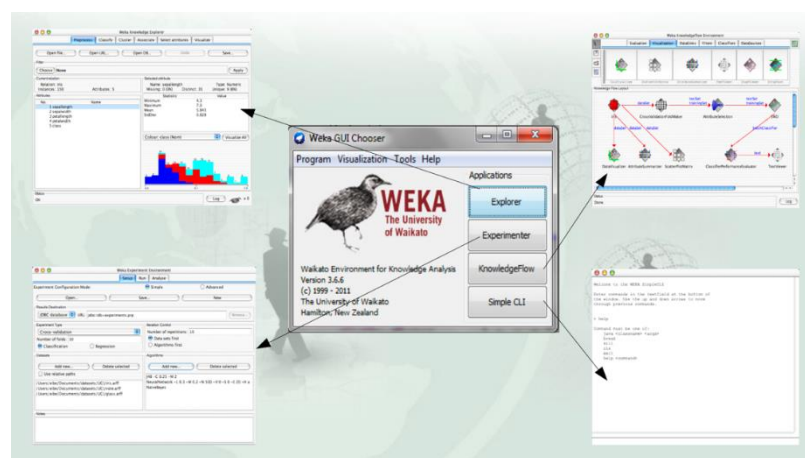
n : Jumlah data

X_t : Data actual pada periode t
 F_t : Nilai peramalan pada periode t

2.9. WEKA

Waikato Environment for Knowledge Analysis (WEKA) merupakan sebuah *tools machine learning* praktis yang dibuat di Universitas Waikato, New Zealand untuk penelitian, Pendidikan dan berbagai aplikasi. WEKA dapat membantu menyelesaikan berbagai masalah *Data mining*, khususnya klasifikasi yang mendasari pendekatan *machine learning*.

WEKA dapat dengan mudah digunakan dan dapat diterapkan pada beberapa tingkatan yang berbeda.. WEKA mempunyai *tools* untuk *pre processing* data, klasifikasi, regresi, *clustering*, aturan asosiasi, dan visualisasi. Contoh penggunaan pada WEKA yaitu dengan menerapkan sebuah metode pembelajaran dataset dan menganalisis hasilnya untuk memperoleh informasi tentang data tersebut, atau dapat menerapkan beberapa metode dan membandingkan performanya untuk dipilih. *Preprocessing* dataset dapat menggunakan *tools* yang membuat *user* dapat berfokus pada algoritma yang digunakan tanpa perlu melihat detail [32].



Gambar 2.4. Software WEKA

2.10. Cross-Validation

Cross-validation atau dapat disebut juga dengan estimasi rotasi merupakan sebuah Teknik validasi model untuk menilai bagaimana hasil statistik analisis akan menggeneralisasi kumpulan data independen. Teknik ini biasanya digunakan untuk melakukan prediksi model dan memperkirakan seberapa akurat sebuah model prediktif ketika dijalankan. Salah satu contohnya adalah *k-fold cross-validation* [33].

Cara kerja *k-fold cross-validation* yaitu dataset yang digunakan akan dipartisi secara merata menjadi k -sub-bagian atau lipatan. Contohnya ada 12 data dengan $k=3$ akan split menjadi data *training* dan data *testing*. Kemudian akan melakukan iterasi ke 1, iterasi ke 2 dan iterasi ke 3.

Iterasi 1	1, 2, 3, 4	5, 6, 7, 8	9, 10, 11, 12
Iterasi 2	1, 2, 3, 4	5, 6, 7, 8	9, 10, 11, 12
Iterasi 3	1, 2, 3, 4	5, 6, 7, 8	9, 10, 11, 12

Gambar 2.5. Ilustrasi *k-fold cross-validation*

Pada Gambar 2.5. iterasi 1 yaitu data yang berada di kolom biru (1, 2, 3, 4) merupakan data *testing*, sedangkan data yang berada di kolom hijau (5, 6, 7, 8, 9, 10, 11, 12) merupakan data *training*. Untuk iterasi 2 data yang berada di kolom biru (5, 6, 7, 8) menjadi data *testing*, dan sisanya menjadi data *training*. Kemudian untuk iterasi 3 data di kolom warna biru (9, 10, 11, 12) akan menjadi data *testing*, dan sisanya menjadi data *training*. Proses ini diulangi sebanyak k , setiap kelompok diperlakukan sebagai data validasi dan sisanya sebagai data latih. Setiap iterasi dicari akurasi, kemudian akan dirata-rata, dan inilah yang akan menjadi sebuah metode *k-fold cross-validation*. *Cross-validation* akurasi lebih stabil karena setiap data akan diolah menjadi data *testing*.

2.11. *Percentage Split*

Percentage split merupakan salah satu test option yang dapat digunakan di WEKA. Hasil klasifikasi *percentage split* akan dites dengan menggunakan k% dari data set sebagai data *training*. Pada *percentage split* dataset dibagi secara acak menjadi dua kumpulan data yang terpisah. Set yang pertama akan mengekstrak pengetahuan dari data *training*. Pengetahuan yang diekstrak dapat diuji terhadap set kedua yang disebut uji coba, biasanya sistem membagi secara acak kumpulan data menjadi 2 bagian [34].

60% - Data <i>training</i>	2,4,5,6,7,8
40% - Data <i>testing</i>	1,3,9,10

Gambar 2.6. Ilustrasi *percentage split*

Gambar 2.6 merupakan ilustrasi contoh penggunaan *percentage split* misalkan terdapat dataset sejumlah 10 data kemudian akan dilakukan *percentage split* yaitu 60% maka artinya 60% dari total dataset sebagai data *training* dan 40% dari total dataset menjadi data *testing*. Sistem ini akan membagi data secara acak. *Percentage split* akurasi tidak stabil karena tidak semua data menjadi data *testing*.

V. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Adapun kesimpulan dari penelitian ini yaitu:

1. Algoritma *C.4.5* lebih baik pada uji coba dengan *cross-validation folds* 5, 10 dan 15 dibandingkan dengan algoritma *random forest*, yaitu dengan nilai MAE 0.0284, 0.0277, dan 0.0285 serta nilai RMSE 0.1408, 0.1392 dan 0.1414. Sedangkan hasil dari *random forest* untuk nilai MAE adalah 0.1131, 0.112 dan 0.1119 serta nilai RMSE 0.216, 0.2146 dan 0.2146.
2. Algoritma *C4.5* lebih baik pada uji coba *percentage split* 25%, 66% dan 90% dibandingkan algoritma *random forest* yaitu dengan nilai MAE 0.0613, 0.0329 dan 0.0234 serta nilai RMSE 0.2287, 0.1463 dan 0.102. Sedangkan hasil dari *random forest* untuk nilai MAE adalah 0.111, 0.1202 dan 0.1295 serta nilai RMSE 0.2218, 0.2428 dan 0.2523.

5.2. Saran

Adapun saran untuk penelitian selanjutnya yaitu:

1. Menggunakan dataset yang lebih lengkap sehingga meminimalisir terjadinya *missing data*.
2. Melakukan uji validasi terhadap prediksi yang didapatkan dengan menggunakan data cuaca di waktu selanjutnya.
3. Uji coba dengan metode lainnya guna mendapatkan hasil yang lebih baik dari penelitian ini.

DAFTAR PUSTAKA

- [1] <https://kbbi.web.id/cuaca.html>, diakses pada 10 Desember 2022 pukul 14.10.
- [2] Saputra, A. D., et al. (2015). Pengaruh Kondisi Cuaca Penerbangan Terhadap Beban Kerja Mental Pilot. *Jurnal Transportasi*, 15 (3), 159-168.
- [3] Andromeda, V. F., et al. (2022). Strategi Transportasi Laut Untuk Mendukung Pengembangan Pariwisata di Karimunjawa. *Jurnal Penelitian Transportasi Laut*, 24, 11-20.
- [4] Becken, L. (2010). The Importance of Climate And Weather for Tourism. *Researchgate*, 1-18.
- [5] Rahmadani, R. E. N., et al. (2019). Pemanfaatan Big Database BMKG untuk Memprediksi Masa Tanam Tembakau (Studi Kasus Di Desa Jatiguwi). *Jurnal Teknologi & Manajemen Informatika*, 5(3).
- [6] <https://lampung.bmkg.go.id>, diakses pada 10 Desember 2022 pukul 15.00.
- [7] Diani, F., Permana, H., Sarah N, P. (2012). Kajian Sistem Informasi Prakiraan Cuaca BMKG Pada BMKG Bandung. *Seminar Nasional Aplikasi Informasi 2012(SNATI 2012)*, B-16-B21. ISSN:1907-5022.
- [8] Supriyadi, R., et al. (2020). Penerapan Algoritma Random Forest Untuk Menemukan Kualitas Anggur Merah. *Jurnal Ilmiah Ekonomi dan Bisnis*, 13(2), 67-75.
- [9] Siahaan, S. W., et al. (2020). Penerapan Algoritma C4.5 Dalam Meningkatkan Kemampuan Bahasa Inggris Pada Mahasiswa. *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*, 13(2), 229-239.
- [10] Naing, W.Y.N., and Htike, Z.Z. (2015). Forecasting of Monthly Temperature Variations using *Random forest*. *Journal of Arp*, 10(21), 10109-10112.
- [11] Lahouar, A., Slama, J.B.H. (2015). Day-Ahead Load Forecast using *Random forest* and Expert Input Selection. *Journal of Elsevier*, 103, 1040-1051.
- [12] Novandya, A. (2017). Penerapan Algoritma Klasifikasi *Data mining C4.5* Pada Data Set Cuaca Wilayah Bekasi. *Konferensi Nasional Ilmu Sosial & Teknologi (KNiST)*. 368-372.
- [13] Bamaruckmani, P., Kausalya, R. (2019). Efficient Analysis of Weather Prediction Using *C4.5* Decision Tree and Naïve Bayes Algorithm. *International Journal of Research in Advent Technology*, 7(5S), 253-260.
- [14] Findawati, Y., et al. (2019). Comparative Analysis of Naïve Bayes, K Nearest Neighbor and *C4.5* Method in Weather Forecast. *Journal of Physics : Conference Series*, 1402, 1-6.
- [15] (<https://www.bmkg.go.id/profil/?p=tugas-fungsi>), diakses pada tanggal 29 Agustus 2022.
- [16] Tjasyono, B. (1999). *Klimatologi Umum*. Bandung. ITB.
- [17] Bisri, M. (2008). *Aliran Air Tanah*. Malang. UPT Penerbit Teknik Universitas Brawijaya.
- [18] Kertasapoetra, A.G. (2010). *Teknologi Konservasi Tanah dan Air*. Rineka Cipta.
- [19] Winarso, P. A. (2003). Variabilitas/ Penyimpangan Iklim Atau Musim di Indonesia dan Pengembangannya. *Makalah Seminar Nasional Ilmu Tanah KMIT Jurusan Tanah Fakultas Pertanian UGM*.

- [20] Mardi, M. *Data mining: Klasifikasi Menggunakan Algoritma C4.5*. *Jurnal Edik Informatika*, v2 (i2), 213-219.
- [21] Tarigan, T. M. S., et al. (2022). Implementasi *Data mining* Menggunakan ALgoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap). *Jurnal Sistem Informasi, Teknologi Informasi dan Komputer (JUST IT)*. 12 (2), 51-61.
- [22] Wahyono, T. (2018). *Python for Machine Learning*. Yogyakarta. Penerbit Gava Media.
- [23] Budiharto, W. (2017). *Machine Learning dan Komputasional Intelligence*. Yogyakarta. Penerbit Andi.
- [24] Mitchell, T. (1997). *Machine Learning*. New York. McGrawHill Publishing.
- [25] Brownlee, J. (2015). *Basic Concepts in Machine Learning, Machine Learning Mastery*, diakses dari <https://machinelearningmastery.com> pada 30 Agustus 2022.
- [26] Loken, E.D., et al. (2019). Postprocessing Next-Day Ensemble Probabilistic Precipitation Forecast Using *Random forest*. *American Meteorological Society*, 34, 2017-2033.
- [27] <https://researchgate.com>, diakses pada 14 Juni 2023.
- [28] Pramudito, D.K. (2022). *Data mining* Implementation on Java North Coast Weather Forecast Dataset Using C4.5 Algorithm. *Jurnal Teknologi Pelita Bangsa*, 13(3), 139-148
- [29] Suryanto, A. A., Muqtadir, A. Penerapan Metode Mean Absolute Error Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi. *SAINTEKBU: Jurnal Sains dan Teknologi*, 11(1), 78-83.
- [30] Lestari, N., Wahyuningsih, N. (2012) Peramalan Kunjungan Wisata dengan Pendekatan Model SARIMA (Studi kasus : Kusuma Agrowisata. *JURNAL SAINS DAN SENI ITS*. 1(1), A-29-A33.
- [31] Azmi, U., Hadi, Z. N., Soraya, S. (2020). ARDL METHOD: Forecasting Data Jumlah Hari Terjadinya Hujan Di NTB. *Jurnal Varian*, 3 (2), 73-82.
- [32] Purnamasari, D., et al. (2013). *Get Easy Using WEKA*. Indonesia. Dapur Buku.
- [33] Azis, H., et al. (2020). Performa Klasifikasi K-NN dan *Cross-Validation* pada Data Pasien Pengidap Penyakit Jantung. *ILKOM Jurnal Ilmiah*, 12 (2), 81-86.
- [34] Syahfitri, I. N. 2017. Penerapan Data Mining untuk Menentukan Besar Pinjaman pada Koperasi Simpan Pinjam dengan Algoritma C4.5 (Studi Kasus: Koperasi Wanita XYZ. *JUISI*, 03 (02), 18-26.