

**IMPLEMENTASI *BOOTSTRAP SAMPLING* PADA METODE
ENSEMBLE BERDASARKAN METODE *NAÏVE BAYES*,
RANDOM FOREST DAN REGRESI LOGISTIK BINER UNTUK
PREDIKSI KEJADIAN HUJAN**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
di Jurusan Matematika pada Fakultas MIPA**

Oleh:

NURAFNI RAHAYU KHOTIMAH

08011181823020



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA**

2022

HALAMAN PENGESAHAN

**IMPLEMENTASI *BOOTSTRAP SAMPLING* PADA METODE *ENSEMBLE*
BERDASARKAN METODE *NAÏVE BAYES*, *RANDOM FOREST* DAN
REGRESI LOGISTIK BINER UNTUK PREDIKSI KEJADIAN HUJAN**

SKRIPSI

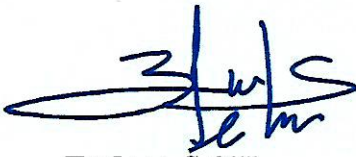
**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
di Jurusan Matematika pada Fakultas MIPA**

Oleh:

NURAFNI RAHAYU KHOTIMAH

08011181823020

Pembimbing Kedua



Endang Sri Kresnawati, M.Si.

NIP.197702082002122003

Indralaya, 01 Agustus 2022

Pembimbing Utama



Dr. Yulia Resti, M.Si.

NIP. 197307191997022001

Mengetahui,

Ketua Jurusan Matematika



Dr. Sugandi Yahdin, M.M.

NIP. 195807271986031003

PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Nurafni Rahayu Khotimah

NIM : 08011181823020

Jurusan : Matematika

Menyatakan dengan ini saya bersungguh-sungguh bahwa skripsi yang berjudul “Implementasi *Bootstrap Sampling* Pada Metode *Ensemble* Berdasarkan Metode *Naïve Bayes*, *Random Forest* Dan Regresi Logistik Biner Untuk Prediksi Kejadian Hujan” merupakan karya yang saya susun sendiri dan saya tidak melakukan penjiplakan dari karya manapun serta saya melakukan pengutipan sesuai dengan pedoman keilmuan yang berlaku seperti tertuang dalam Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 17 Tahun 2010 terkait Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi.

Apabila dikemudian hari, terdapat pelanggaran yang ditemukan dalam skripsi saya ataupun adanya pengaduan dari pihak lain terhadap keaslian skripsi saya, maka saya bersedia menanggung sanksi yang dijatuhkan kepada saya.

Demikianlah pernyataan ini dibuat dengan sungguh-sungguh tanpa adanya paksaan dari pihak manapun.

Indralaya, 28 Juli 2022

Yang membuat pernyataan,



Nurafni Rahayu Khotimah

NIM. 08011181823020

HALAMAN PERSEMBAHAN

MOTTO

“Hasbunallah wa ni'mal wakil ni'mal maula wa ni'man nasiir”

(HR. Bukhari no.4563)

Skripsi ini kupersembahkan kepada:

- 1. Allah SWT**
- 2. Ayah dan Ibu tercinta serta Saudara-saudaraku tersayang**
- 3. Dosen**
- 4. Almamater**
- 5. Sahabat-sahabat yang luar biasa**

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT yang Maha Pengasih Lagi Maha Penyayang atas berkat rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Implementasi *Bootstrap Sampling* Pada Metode *Ensemble* Berdasarkan Metode *Naïve Bayes*, *Random Forest* Dan Regresi Logistik Biner Untuk Prediksi Kejadian Hujan**” dengan baik. Skripsi ini merupakan salah satu persyaratan yang harus dipenuhi penulis untuk meraih gelar Sarjana Sains bidang Studi Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Selama menyelesaikan skripsi penulis telah mendapat banyak bantuan, motivasi, bimbingan, dan pengarahan dari berbagai pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, dengan segala hormat dan kerendahan hati penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini terkhusus kepada orang tua tercinta, Ibu **Nurlaila** dan Bapak **Romli** atas segala doa terbaik, kasih sayang dan nasihat yang selalu diberikan kepada penulis. Penulis juga mengucapkan terima kasih kepada:

1. Bapak **Drs. Sugandi Yahdin, M.M**, selaku Ketua Jurusan Matematika FMIPA Universitas Sriwijaya yang telah memberikan bimbingan dan ilmu selama masa perkuliahan.
2. Ibu **Dr. Dian Cahyawati Sukanda, M.Si** selaku Sekretaris Jurusan Matematika FMIPA Universitas Sriwijaya yang telah membantu memberikan bimbingan selama masa perkuliahan.
3. Ibu **Dr. Yulia Resti, M.Si** selaku Dosen Pembimbing Utama yang telah

bersedia meluangkan waktu, tenaga dan pikiran untuk memberikan bimbingan, pengetahuan, motivasi, dan pengarahan yang sangat bermanfaat bagi penulis dalam penyusunan skripsi ini.

4. Ibu **Endang Sri Kresnawati, M.Si** selaku Dosen Pembimbing Kedua yang telah bersedia meluangkan waktu untuk memberikan bimbingan, saran, ilmu dan motivasi kepada penulis selama menyelesaikan skripsi.
5. Ibu **Dr. Ir. Herlina Hanum, M.Si** selaku Dosen Penguji I dan Ibu **Dr. Evi Yuliza, M.Si** selaku Dosen Penguji II yang telah memberikan tanggapan, ilmu, kritik dan saran yang bermanfaat untuk perbaikan dan penyelesaian skripsi ini.
6. Ibu **Dra. Ning Eliyati, M.Pd** selaku Dosen Pembimbing Akademik yang telah memberikan semangat, motivasi dan pengarahan kepada penulis selama masa perkuliahan.
7. Ibu **Indrawati, M.Si** selaku Ketua Seminar dan Ibu **Des Alwine Zayanti, M.Si** selaku Sekretaris Seminar Skripsi yang telah memberikan arahan, saran, membimbing penulis serta mengatur jalannya seminar sehingga dapat berjalan dengan baik.
8. **Bapak dan Ibu Dosen Jurusan Matematika FMIPA Universitas Sriwijaya** yang telah memberikan pengetahuan yang bermanfaat, nasihat serta motivasi selama penulis menjalani perkuliahan.
9. Bapak **Irwansyah** dan Ibu **Hamidah** selaku Pegawai tata usaha Jurusan Matematika FMIPA yang telah membantu penulis dalam hal administrasi selama perkuliahan.
10. Keluarga yang selalu memberikan dukungan kepada penulis.

11. Sahabat-sahabat penulis yang baik hati.
12. Kakak-kakak tingkat angkatan 2016, 2017, dan seluruh teman-teman angkatan 2018 atas bantuan selama perkuliahan.
13. Pihak-pihak lain yang membantu penulisan skripsi penulis.

Penulis menyadari bahwa penulisan skripsi ini masih memiliki kekurangan sehingga penulis sangat mengharapkan kritik dan saran yang membangun agar skripsi ini dapat menjadi lebih baik.

Indralaya, Mei 2022

Penulis

**IMPLEMENTATION OF BOOTSTRAP SAMPLING IN ENSEMBLE
METHOD BASED ON NAÏVE BAYES METHOD, RANDOM FOREST
AND BINARY LOGISTIC REGRESSION FOR RAIN EVENT
PREDICTION**

By:

**Nurafni Rahayu Khotimah
08011181823020**

ABSTRACT

Rain is an event of precipitation or water vapor that condenses in the atmosphere and falls to the earth's surface in the form of a liquid. Rain is highly dependent on topography and weather factors. In this case, the forecast for rain can be seen based on weather elements such as air temperature, humidity, air pressure, wind speed and so on. The purpose of this study was to predict rainfall events using the Ensemble Majority Voting method based on the Naïve Bayes method, Random Forest and Binary Logistics Regression with the implementation of Bootstrap Sampling. This study uses weather data in Australia for 2008-2017 sourced from the Kaggle website, which consists of 18 variables with a total of 145399 data entries. The level of accuracy obtained in predicting rain events using the Naïve Bayes method with the implementation of Bootstrap Sampling produces an average value of accuracy, precision, recall, and fscore of 77.94%, 50.66%, 56.37%, and 53.37%, respectively. The Random Forest method with the implementation of Bootstrap Sampling produces an average value of accuracy, precision, recall, and fscore of 82.58%, 67.04%, 43.71%, and 52.92%, respectively. The Binary Logistics Regression Method with the implementation of Bootstrap Sampling produces an average value of accuracy, precision, recall, and fscore of 82.71%, 69.93%, 39.98%, and 50.87%, respectively. The Ensemble Majority Voting method with the implementation of Bootstrap Sampling produces an average value of accuracy, precision, recall, and fscore of 82.83%, 67.51%, 44.98%, and 53.98%, respectively. The results of this study indicate that the prediction of rain events using the Ensemble Method has better accuracy than the other three methods because it has a higher accuracy value and fscore. However, the Naïve Bayes method has a better recall value and Binary Logistics Regression has a better precision value than other methods.

Keywords: Rain Forecast, Ensemble Majority Voting, Naïve Bayes, Random Forest, Binary Logistics Regression

**IMPLEMENTASI BOOTSTRAP SAMPLING PADA METODE
ENSEMBLE BERDASARKAN METODE NAÏVE BAYES, RANDOM
FOREST DAN REGRESI LOGISTIK BINER UNTUK PREDIKSI
KEJADIAN HUJAN**

Oleh:

**Nurafni Rahayu Khotimah
08011181823020**

ABSTRAK

Hujan adalah sebuah peristiwa presipitasi atau uap air yang mengkondensasi di atmosfer dan jatuh ke permukaan bumi berwujud cairan. Hujan sangat tergantung pada faktor topografi dan cuaca. Dalam hal ini, prakiraan terjadinya hujan bisa dilihat berdasarkan unsur cuaca seperti suhu udara, kelembaban, tekanan udara, kecepatan angin dan lain sebagainya. Tujuan penelitian ini adalah memprediksi kejadian hujan menggunakan metode *Ensemble Majority Voting* berdasarkan metode *Naïve Bayes*, *Random Forest* dan Regresi Logistik Biner dengan implementasi *Bootstrap Sampling*. Penelitian ini menggunakan data cuaca di Australia tahun 2008-2017 yang bersumber dari website Kaggle, yang terdiri dari 18 variabel dengan jumlah data sebanyak 145399 entri. Tingkat ketepatan yang diperoleh dalam memprediksi kejadian hujan menggunakan metode *Naïve Bayes* dengan implementasi *Bootstrap Sampling* menghasilkan nilai rata-rata akurasi, presisi, *recall*, dan *fscore* masing-masing sebesar 77,94%, 50,66%, 56,37%, dan 53,37%. Metode *Random Forest* dengan implementasi *Bootstrap Sampling* menghasilkan nilai rata-rata akurasi, presisi, *recall*, dan *fscore* masing-masing sebesar 82,58%, 67,04%, 43,71%, dan 52,92%. Metode Regresi Logistik Biner dengan implementasi *Bootstrap Sampling* menghasilkan nilai rata-rata akurasi, presisi, *recall*, dan *fscore* masing-masing sebesar 82,71%, 69,93%, 39,98%, dan 50,87%. Metode *Ensemble Majority Voting* dengan implementasi *Bootstrap Sampling* menghasilkan nilai rata-rata akurasi, presisi, *recall*, dan *fscore* masing-masing sebesar 82,83%, 67,51%, 44,98%, dan 53,98%. Hasil penelitian ini menunjukkan bahwa prediksi kejadian hujan menggunakan Metode *Ensemble* memiliki keakuratan yang lebih baik dibandingkan ketiga metode lainnya karena mempunyai nilai akurasi dan *fscore* yang lebih tinggi. Namun metode *Naïve Bayes* memiliki nilai *recall* yang lebih baik dan Regresi Logistik Biner memiliki nilai presisi yang lebih baik dari metode lainnya.

Kata Kunci: Prediksi Hujan, *Ensemble Majority Voting*, *Naïve Bayes*, *Random Forest*, Regresi Logistik Biner

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERSEMBAHAN	iii
KATA PENGANTAR.....	iv
ABSTRACT	vii
ABSTRAK	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan.....	6
1.5 Manfaat.....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 <i>Statistical Machine Learning</i>	8
2.2 Peluang	9
2.3 Klasifikasi.....	9
2.4 <i>Bootstrap Sampling</i>	10
2.5 Metode <i>Naïve Bayes</i>	11
2.6 <i>Laplace Smoothing</i>	13
2.7 Metode <i>Random Forest</i>	13
2.8 Analisis Regresi.....	16
2.8.1 Analisis Regresi Logistik Biner	16
2.8.2 Pendugaan Parameter.....	17
2.8.3 Uji Serentak	19
2.8.4 Uji Parsial	20
2.8.5 Uji Kesesuaian Model.....	21
2.8.6 <i>Odds Ratio</i>	22
2.9 Metode <i>Ensemble</i>	22
2.10 <i>Confusion Matrix</i>	23
BAB III METODOLOGI PENELITIAN	26
3.1 Waktu Penelitian	26
3.2 Tempat Penelitian.....	26
3.3 Data Penelitian	26

3.4 Metode Penelitian.....	27
BAB IV HASIL DAN PEMBAHASAN	30
4.1 Deskripsi Data	30
4.2 Diskritisasi Data	31
4.3 Ringkasan Data	33
4.4 Partisi Data	36
4.5 <i>Bootstrap Sampling</i>	36
4.6 Metode <i>Naïve Bayes</i>	37
4.7 Metode <i>Random Forest</i>	44
4.8 Analisis Regresi Logistik Biner	53
4.8.1 Pendugaan Parameter	53
4.8.2 Uji Serentak	54
4.8.3 Uji Parsial	55
4.8.4 Uji Kesesuaian Model.....	56
4.8.5 Interpretasi <i>Odds Ratio</i>	57
4.8.6 Probabilitas	58
4.9 Metode <i>Ensemble</i>	61
4.10 Analisis Hasil	63
BAB V KESIMPULAN DAN SARAN	66
5.1 Kesimpulan.....	66
5.2 Saran.....	67
DAFTAR PUSTAKA	68
LAMPIRAN.....	72

DAFTAR TABEL

Tabel 2.1 Metode <i>Ensemble Majority Voting</i>	23
Tabel 2.2 <i>Confusion Matrix</i>	23
Tabel 2.3 Kategori Ketepatan Klasifikasi	25
Tabel 3.1 Variabel penelitian	26
Tabel 4.1 Deskripsi variabel	30
Tabel 4.2 Diskritisasi Data.....	31
Tabel 4.3 Ringkasan Data Variabel Respon (Y)	33
Tabel 4.4 Ringkasan Data Variabel Prediktor.....	33
Tabel 4.5 Data latih asli	36
Tabel 4.6 Data uji	36
Tabel 4.7 Data latih <i>Bootstrap Sampling</i> ke-29	37
Tabel 4.8 Nilai Peluang <i>Likelihood</i>	40
Tabel 4.9 Hasil prediksi metode <i>Naïve Bayes</i>	42
Tabel 4.10 <i>Confusion Matrix</i> Metode <i>Naïve Bayes</i>	43
Tabel 4.11 Sampel Pohon Pertama	45
Tabel 4.12 Perhitungan <i>entropy</i> dan <i>gain</i> variabel X_4, X_{12}, X_{14} dan X_{17}	47
Tabel 4.13 <i>Rules</i> Pohon Keputusan	51
Tabel 4.14 Hasil prediksi Metode <i>Random Forest</i>	51
Tabel 4.15 <i>Confusion Matrix</i> Metode <i>Random Forest</i>	52
Tabel 4.16 Pendugaan parameter	54
Tabel 4.17 Hasil Uji Serentak	55
Tabel 4.18 Perhitungan statistik uji Wald.....	56
Tabel 4.19 Uji <i>Hosmer</i> dan <i>Lemeshow</i>	56
Tabel 4.20 Hasil Prediksi Metode Regresi Logistik Biner	59
Tabel 4.21 <i>Confusion Matrix</i> metode Regresi Logistik Biner	59
Tabel 4.22 Hasil prediksi metode <i>Ensemble Majority Voting</i>	61
Tabel 4.23 <i>Confusion Matrix</i> metode <i>Ensemble Majority Voting</i>	62
Tabel 4.24 Perbandingan ketepatan hasil prediksi dengan implementasi <i>Bootstrap</i>	64

DAFTAR GAMBAR

Gambar 2.1 Contoh Visualisasi Sederhana Metode <i>Bootstrap</i>	10
Gambar 2.2 Metode <i>Random Forest</i>	14
Gambar 4.1 Pohon keputusan <i>root node</i>	48
Gambar 4.2 Pohon keputusan <i>node</i> cabang 1.0	49
Gambar 4.3 Pohon keputusan sampai <i>node</i> 1.0.2.0	50

DAFTAR LAMPIRAN

Lampiran 1. Pendugaan Parameter Regresi Logistik Biner	72
--	----

BAB I

PENDAHULUAN

1.1 Latar Belakang

Hujan adalah sebuah peristiwa presipitasi atau uap air yang mengkondensasi di atmosfer dan jatuh ke permukaan bumi berwujud cairan. Presipitasi adalah proses uap air yang mengkondensasi dan jatuh dari atmosfer ke bumi dalam segala bentuknya dalam rangkaian siklus hidrologi (Susilowati & Sadad, 2015). Menurut Kamus Besar Bahasa Indonesia Daring (KBBI Daring) hujan adalah tetesan air yang jatuh dari udara karena adanya proses pendinginan. Air yang terdapat di perairan akan mengalami penguapan akibat panasnya matahari, kemudian uap air akan berkumpul di atas dan berpindah tempat karena terbawa angin. Pada proses ini adanya perubahan suhu udara atau penambahan uap air ke udara yang mengubah uap menjadi embun dan selanjutnya membentuk awan. Proses hujan akan berlangsung jika kondisi fisik awan baik di dalam maupun diluar mendukung. Oleh karena itu sifat dan kondisi suatu hujan sangat tergantung pada faktor topografi dan faktor cuaca (Mulyono, 2016).

Cuaca merupakan suatu kondisi udara di suatu tempat yang sifatnya berubah-ubah atau relatif singkat. Menurut Dewi et al. (2014), cuaca terbentuk dari beberapa unsur-unsur diantaranya suhu udara, kelembaban, kecepatan angin, arah angin dan lain sebagainya. Kehidupan manusia tidak dapat terlepas dari peranan cuaca itu sendiri. Pertanian dan perkebunan bergantung dengan cuaca karena cuaca yang baik akan membuat hasil pertanian dan perkebunan yang baik pula. Jika cuaca mendukung maka pelayaran dan penerbangan dapat berjalan dengan baik.

Pentingnya hasil pengukuran unsur-unsur cuaca ialah untuk digunakan sebagai proses prediksi atau prakiraan cuaca. Prakiraan cuaca adalah memperkirakan keadaan atau kondisi udara pada masa datang yang terjadi di suatu daerah (Diani et al., 2012). Prakiraan cuaca dimanfaatkan sebagai langkah antisipasi untuk memperkecil risiko yang akan terjadi. Prediksi diharapkan mempunyai keakuratan tinggi terhadap cuaca supaya kegiatan manusia tidak terhambat (Siregar, 2020).

Pada penelitian ini, akan dilakukan proses klasifikasi untuk memprediksi kejadian hujan dengan menggunakan proses *statistical machine learning*. *Statistical machine learning* mengacu pada teknik untuk memprediksi masa yang akan datang serta melakukan inferensi dan interpretasi dari model (James et al., 2013). Metode yang akan digunakan dalam penelitian ini adalah metode *Ensemble* berdasarkan metode *Naïve Bayes*, *Random Forest* dan Regresi Logistik Biner.

Prediksi kejadian hujan dilakukan dengan menggunakan data cuaca yang berlokasi di Australia. Data diperoleh dari website [Kaggle.com](https://www.kaggle.com) dan kemudian akan dilakukan proses klasifikasi terhadap hujan menggunakan 18 variabel yaitu variabel Location, MinTemp, MaxTemp, Rainfall, WindGustDir, WindGustSpeed, WindDir9am, WinDir3pm, WindSpeed9am, WindSpeed3pm, Humidity9am, Humidity3pm, Cloud9am, Cloud3pm, Temp9am, Temp3pm, RainToday dan RainTomorrow.

Metode *Naïve Bayes* merupakan metode klasifikasi yang sederhana dan berakurasi tinggi. Penentuan estimasi parameter yang diperlukan dalam proses pengklasifikasian dengan metode *Naïve Bayes* hanya membutuhkan jumlah data pelatihan yang kecil sehingga memiliki alur perhitungan yang tidak panjang (Putro

et al., 2020). Penelitian terdahulu oleh Imandasari et al. (2019) membahas Algoritma *Naïve Bayes* dalam Klasifikasi Lokasi Pembangunan Sumber Air menghasilkan akurasi sebesar 76,92%. Penelitian lainnya yang dilakukan oleh Pramudita (2018) membahas Klasifikasi Berita Olahraga Menggunakan Metode *Naïve Bayes* dengan *Enhanced Confix Stripping Stemmer* menghasilkan akurasi 77%.

Metode *Random Forest* merupakan kombinasi dari beberapa *decision tree* yang digabungkan kedalam suatu model dan kemudian keputusan akhir ditentukan dari hasil keputusan terbanyak (Adrian et al., 2021). Penelitian terdahulu oleh Ratnawati & Sulistyaningrum (2019) membahas Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel menghasilkan akurasi sebesar 75,3191%. Penelitian lainnya yang dilakukan oleh Primajaya & Sari (2018) membahas *Random Forest Algorithm for Prediction of Precipitation* menghasilkan akurasi sebesar 71,09%.

Menurut Hosmer et al. (2013) model regresi logistik biner digunakan untuk menganalisis hubungan antara satu variabel respon dan beberapa variabel prediktor, dimana variabel responnya berbentuk data kualitatif bersifat biner atau dikotomus. Pada penelitian Rumaenda et al. (2016), membahas Perbandingan Klasifikasi Penyakit Hipertensi Menggunakan Regresi Logistik Biner dan Algoritma C4.5 memberikan hasil terbaik pada Regresi Logistik Biner dengan akurasi mencapai 72,54%. Penelitian lainnya yang dilakukan oleh Ramli et al. (2013) membahas Perbandingan Metode Klasifikasi Regresi Logistik Dengan Jaringan Saraf Tiruan (Studi Kasus: Pemilihan Jurusan Bahasa dan IPS pada SMAN 2 Samarinda Tahun

Ajaran 2011/2012) memberikan hasil ketepatan Regresi Logistik biner sebesar 78,34%.

Metode *Ensemble* merupakan algoritma *machine learning* (pembelajaran mesin) dalam pencarian keputusan prediksi terbaik dengan menggunakan beberapa metode untuk dapat mencapai akurasi tertinggi. *Ensemble learning* sangat populer untuk meningkatkan akurasi dengan melakukan penggabungan metode (Siregar, 2020). Pada penelitian Siregar (2020) membahas Prediksi Cuaca Menggunakan *Ensemble Learning* menghasilkan akurasi sebesar 81,21%. Penelitian lainnya oleh Sidik & Sen (2019) membahas Penggunaan Teknik *Ensemble Stacking Classifier* Untuk Prediksi Curah Hujan menghasilkan akurasi sebesar 78,25%.

Kadangkala dalam prediksi sering dihadapkan dengan jumlah sampel yang relatif sedikit, sehingga tidak dapat mewakili data populasi sebenarnya. Oleh karena itu solusi yang dapat digunakan ialah *Bootstrap Sampling* (Sari et al., 2017). *Bootstrap Sampling* merupakan teknik *resampling* data sampel dengan syarat pengembalian dari data asli (Kuhn & Johnson, 2013). Ukuran *resampling* diambil sebanyak berulang kali untuk mendapatkan ukuran sampel yang lebih besar. Hal ini dilakukan agar dapat mewakili data populasinya (Rachman et al., 2018).

Penelitian-penelitian terdahulu yang telah dijelaskan pada paragraf sebelumnya memperlihatkan bahwa metode *Naïve Bayes*, *Random Forest* dan Regresi Logistik Biner memberikan ketepatan yang baik. Oleh karena itu, penulis tertarik untuk melakukan penelitian mencari prediksi terbaik dengan menerapkan metode *Bootstrap* dalam prediksi kejadian hujan menggunakan metode *Ensemble* berdasarkan metode *Naïve Bayes*, *Random Forest* dan Regresi Logistik Biner.

Metode *Ensemble* yang digunakan adalah metode *Ensemble* jenis *Majority Voting* yang berarti pengambilan prediksi dengan suara terbanyak dari hasil prediksi berbagai metode klasifikasi.

1.2 Perumusan Masalah

Berdasarkan latar belakang, perumusan masalah dalam penelitian ini adalah:

1. Berapa tingkat ketepatan dalam memprediksi kejadian hujan menggunakan metode *Naïve Bayes*, *Random Forest* dan Regresi Logistik Biner dengan implementasi *Bootstrap Sampling*?
2. Berapa tingkat ketepatan dalam memprediksi kejadian hujan menggunakan metode *Ensemble* berdasarkan metode *Naïve Bayes*, *Random Forest* dan Regresi Logistik Biner dengan implementasi *Bootstrap Sampling*?
3. Bagaimana perbandingan tingkat ketepatan dalam memprediksi kejadian hujan antara metode *Naïve Bayes*, *Random Forest*, Regresi Logistik Biner dengan implementasi *Bootstrap Sampling* dan metode *Ensemble* dengan implementasi *Bootstrap Sampling*?

1.3 Batasan Masalah

Penelitian ini menggunakan data cuaca di Australia pada tahun 2008-2017 yang bersumber dari website Kaggle yang dapat diakses melalui link:

<https://www.kaggle.com/jsphyg/weather-dataset-rattle-package>.

1.4 Tujuan

Berdasarkan perumusan masalah, tujuan penelitian ini adalah:

1. Menentukan tingkat ketepatan dalam memprediksi kejadian hujan menggunakan metode *Naïve Bayes*, *Random Forest* dan Regresi Logistik Biner dengan implementasi *Bootstrap Sampling*.
2. Menentukan tingkat ketepatan dalam memprediksi kejadian hujan menggunakan metode *Ensemble* berdasarkan metode *Naïve Bayes*, *Random Forest* dan Regresi Logistik Biner dengan implementasi *Bootstrap Sampling*.
3. Mengetahui perbandingan tingkat ketepatan dalam memprediksi kejadian hujan antara metode *Naïve Bayes*, *Random Forest*, Regresi Logistik Biner dengan implementasi *Bootstrap Sampling* dan metode *Ensemble Majority Voting* dengan implementasi *Bootstrap Sampling*.

1.5 Manfaat

Manfaat dari penelitian ini adalah sebagai berikut.

1. Dapat menambah pengetahuan dan wawasan mengenai keilmuan statistika dalam persoalan riil khususnya pada penggunaan Metode *Ensemble* berdasarkan metode *Naïve Bayes*, *Random Forest* dan Regresi Logistik Biner.
2. Diharapkan memberikan manfaat kepada instansi tertentu agar mengetahui faktor-faktor yang mempengaruhi terjadinya hujan sehingga bisa dilakukan antisipasi untuk memperkecil risiko yang akan terjadi.

3. Sebagai referensi bagi penelitian yang membahas mengenai prediksi hujan berdasarkan unsur cuaca.

DAFTAR PUSTAKA

- Adrian, M. R., Putra, M. P., Rafialdy, M. H., & Rakhmawati, N. A. (2021). Perbandingan Metode Klasifikasi Random Forest dan SVM. *Jurnal Informatika UPGRIS*, 7(1).
- Ardianto, R. (2021). *Prediksi Hujan Berdasarkan Unsur Cuaca Menggunakan Metode Decision Tree*. Universitas Sriwijaya.
- Chairunisa, R., Adiwijaya, & Astuti, W. (2021). Perbandingan CART dan Random Forest untuk Deteksi Kanker berbasis Klasifikasi Data Microarray. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(5), 805–812.
- Dewi, C., Kartikasari, D. P., & Mursityo, Y. T. (2014). Prediksi Cuaca Pada Data Time Series Menggunakan Adaptive Neuro Fuzzy Inference System (Anfis). *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 1(1), 18–24.
- Diani, F., Permana, H., & Sarah, P. N. (2012). Kajian Sistem Informasi Prakiraan Cuaca BMKG Pada BMKG Bandung. *Faks*, 40012(022).
- Ghozi, S., Ramli, R., & Setyani, A. (2018). Analisis Keputusan Nasabah dalam Memilih Jenis Bank: Penerapan Model Regresi Logistik Biner (Studi Kasus Pada Bank Bri Cabang Balikpapan). *Media Statistika*, 11(1), 17–26.
- Hijriani, A., Muludi, K., & Andini Ain, E. (2016). Implementasi Metode Regresi Linier Sederhana Pada Penyajian Hasil Prediksi Pemakaian Air Bersih PDAM Way Rilau Kota Bandar Lampung dengan Sistem Informasi Geografis. *Jurnal Informatika Mulawarman*, 11(2), 37–42.
- Hosmer, D. W., Lemeshow, Stanley., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (Third Edition). A John Wiley and Sons, Inc.
- Imandasari, T., Irawan, E., Windarto, A. P., & Wanto, A. (2019). Algoritma Naive Bayes dalam Klasifikasi Lokasi Pembangunan Sumber Air. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1, 750–761.
- Iniesta, R., Stahl, D., & McGuffin, P. (2016). Machine learning, statistical learning and the future of biological research in psychiatry. *Psychological Medicine*, 46(12), 2455–2465.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning : with applications in R*. Springer.

- Kresna, A., Putra, B. A., Fauzi, M. A., Setiawan, B. D., & Setiawati, E. (2018). Identifikasi Ujaran Kebencian Pada Facebook dengan Metode Ensemble Feature dan Support Vector Machine. In *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* (Vol. 2, Issue 12).
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. usa.
- Kusyanti, A. (2019). Metode Ensemble Classifier Untuk Mendeteksi Jenis Attention Deficit Hyperactivity Disorder (ADHD) Pada Anak Usia Dini. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 6(3), 301–308.
- Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge In Data: An Introduction to Data Mining*. In *IEEE Computer Society* (Second Edition). A John Wiley and Sons, Inc.
- Marcus, G. L., Wattimanela, H. J., & Lesnussa, Y. A. (2012). Analisis Regresi Komponen Utama Untuk Mengatasi Masalah Multikolinieritas dalam Analisis Regresi Linier Berganda. *Jurnal Barekeng*, 6, 31–40.
- Mishra, S., Vanli, O. A., Huffer, F. W., & Jung, S. (2016). *Regularized Discriminant Analysis for Multi-sensor Decision Fusion and Damage Detection with Lamb-waves*.
- Mulyono, D. (2016). Analisis Karakteristik Curah Hujan di Wilayah Kabupaten Garut Selatan. *Jurnal Konstruksi*, 12(1), 1–9.
- Pramudita, Y. D., Putro, S. S., & Makhmud, N. (2018). Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(3), 269.
- Primajaya, A., & Sari, B. N. (2018). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIM)*, 1(1), 27–31.
- Putro, H. F., Vlandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi Dan Komunikasi (TIKOMSiN)*, 8(2).
- Rachman, F. P. A. P., Goejantoro, R., & Hayati, M. N. (2018). Penentuan Jumlah Replikasi Bootstrap Menggunakan Metode Pretest Pada Independent Sampel T Test (Pendapatan Asli Daerah Kabupaten/Kota di Provinsi Kalimantan Timur dan Kalimantan Utara Tahun 2015). *Jurnal Ekspansional*, 9(1).

- Ramadhani, N., & Fajarianto, N. (2020). Sistem Informasi Evaluasi Perkuliahan dengan Sentimen Analisis Menggunakan Naïve Bayes dan Smoothing Laplace. *Jurnal Sistem Informasi Bisnis*, 02.
- Ramli, Yuniarti, D., & Goejantoro, R. (2013). Perbandingan Metode Klasifikasi Regresi Logistik Dengan Jaringan Saraf Tiruan (Studi Kasus: Pemilihan Jurusan Bahasa dan IPS pada SMAN 2 Samarinda Tahun Ajaran 2011/2012). *Jurnal Eksponensial*, 4(1).
- Ratnawati, L., & Sulistyaningrum, D. R. (2019). Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel. *Jurnal Sains Dan Seni ITS*, 8(2).
- Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156.
- Rumaenda, W., Wilandari, Y., & Safitri, D. (2016). Perbandingan Klasifikasi Penyakit Hipertensi Menggunakan Regresi Logistik Biner Dan Algoritma C4.5 (Studi Kasus Upt Puskesmas Ponjong I, Gunungkidul). *Studi Puskesmas, U P T*, 5(2), 299–309.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Creative Information Technology Journal*, 2(3), 207–217.
- Saputro, D. S. R., & Widyaningsih, P. (2016). Algoritme Pendugaan Parameter Model Regresi Logistik Biner (RLB) dengan Maksimum Likelihood dan Broyden-Fletcher-Goldfarb-Shanno (BFGS). *Seminar Nasional Matematika Dan Pendidikan Matematika UNY*, 97–104.
- Sari, S. K. P., Santoso, R., & Suparti. (2017). Prediksi Simpanan Berjangka Pada Bank Umum dan BPR Menggunakan Metode ARIMA dengan Outliers dan ARIMA Bootstrap. *Jurnal Gaussian*, 6(3), 459–468.
- Sidik, D. D., & Sen, T. W. (2019). *Penggunaan Stacking Classifier Untuk Prediksi Curah Hujan*. 04(01).
- Siregar, A. M. (2020). Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning. *Petir*, 13(2), 138–147.
- Sitinjak, B. J. B. M. (2021). *Estimasi Parameter Regresi Logistik Multinomial Menggunakan Maksimum Likelihood*. Universitas Sumatera Utara.

- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 45(4), 427–437.
- Sugiyarto. (2021). *Pengantar Statistika Matematika I*. Magnum Pustaka Utama.
- Susilowati, & Sadad, I. (2015). Analisa Karakteristik Curah Hujan di Kota Bandar Lampung. *Jurnal Konstruksia*, 7(1), 13–26.
- Tampil, Y. A., Komalig, H., & Langi, Y. (2015). *Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado*.
- Tyasti, A. E., Ispriyanti, D., & Hoyyi, A. (2015). Algoritma Iterative Dichotomiser 3 (Id3) Untuk Mengidentifikasi Data Rekam Medis (Studi Kasus Penyakit Diabetes Mellitus Di Balai Kesehatan Kementerian Perindustrian, Jakarta). *Jurnal Gaussian*, 4(2), 237–246.
- Wibawa, A. P., Guntur, M., Purnama, A., Fathony Akbar, M., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1).
- Zuhdi, S., & Saputro, D. (2015). *Pendugaan Parameter Regresi Logistik Biner Dengan Spreadsheet Solver (Add-In Microsoft Excel) Analysis Bayesian View project R Programming for Geographically Weighted Multinomial Logistic Regression View project*.