

PERBANDINGAN ALGORITME KLASIFIKASI UNTUK PREDIKSI CUACA

Amril Mutoi Siregar

Universitas Buana Perjuangan karawang
amrilmutoi@ubpkarawang.ac.id

Sutan Faisal

Universitas Buana Perjuangan karawang
sutan.faisal@ubpkarawang.ac.id

Yana Cahyana

Universitas Buana Perjuangan karawang
yana.cahyana@ubpkarawang.ac.id

Bayu Priyatna

Universitas Buana Perjuangan karawang
bayu.priyatna@ubpkarawang.ac.id

ABSTRACT

Weather conditions is an air condition in a place with a relatively short time, which is expressed by the value of parameters such as temperature, wind speed, pressure, rainfall, which is another atmospheric phenomenon as the main component. Human activities can be influenced by weather conditions, such as transportation, agriculture, plantation, construction or even sports. Therefore, for determining the weather, getting weather information needs to be made so that it can be utilized by the community. Problems that arise how to make weather predictions automatically so that it can be done by everyone. In this study proposed several algorithms Navie Bayes, Decision Tree, Random Forest to calculate the opportunities of one class from each of the existing group attributes and determine which class is the most optimal, meaning that grouping can be done based on the categories that users enter in the application. The prediction system has been made to obtain an accuracy rate of Navie Bayes of 77.22% with a standard deviation of 29%, a Decision Tree accuracy rate of 79.46% with a standard deviation of 15%, a random forest accuracy rate of 82.38% with a standard deviation of 43%..

Keywords: *Classification, Decision Tree, Naïve Bayes, Random Forest, Prediction, Weather*

ABSTRAK

Kondisi cuaca merupakan kondisi udara di suatu tempat dengan waktu relatif singkat, yang dinyatakan dengan nilai parameter seperti suhu, kecepatan angin, tekanan, curah hujan, yang menjadi fenomena atmosfer lainnya sebagai komponen utama. Aktivitas manusia dapat dipengaruhi oleh kondisi cuaca, seperti bidang transportasi, bidang pertanian, bidang perkebunan, bidang bangunan atau bahkan bidang olah raga dan lain lain. Karena itu untuk penentuan cuaca mendapatkan informasi cuaca perlu dibuat sehingga dapat dimanfaatkan oleh masyarakat. Permasalahan yang muncul bagaimana membuat prediksi cuaca dengan otomatis sehingga dapat dilakukan oleh setiap orang. Dalam penelitian ini mengajukan beberapa algoritme Navie Bayes, Decision Tree, Random forest untuk menghitung peluang dari satu kelas dari masing-masing atribut kelompok yang ada dan menentukan kelas mana yang paling optimal, artinya pengelompokan dapat dilakukan berdasarkan kategori yang pengguna masukkan pada aplikasi. Sistem prediksi telah dibuat mendapatkan tingkat akurasi Navie Bayes sebesar 77.22% dengan standar deviasi sebesar 29 %, tingkat akurasi Decision Tree

sebesar 79.46% dengan standar deviasi sebesar 15 %, tingkat akurasi Random forest sebesar 82.38% dengan standar deviasi sebesar 43 %.

Kata Kunci: *Cuaca, Decision Tree, Klasifikasi, Naïve bayes, Random forest, Prediksi,*

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Cuaca adalah suatu tempat kondisi udara pada waktu yang relatif singkat, dinyatakan oleh nilai-nilai parameter seperti kecepatan angin, suhu, tekanan, curah hujan, dan fenomena atmosfer lainnya sebagai komponen utama. Cuaca merupakan hal penting yang tidak akan pernah lepas dari kehidupan manusia. Terutama aktivitas manusia dapat dipengaruhi oleh kondisi cuaca, seperti pertanian, bidang perkebunan. Misalnya waktu keputusan yang baik untuk menanam tanaman. Besarnya pengaruh iklim yang disebabkan mendorong pengembangan sistem cuaca menentukan yang menentukan kondisi cuaca. Pendekatan ini didukung dengan menggunakan teknologi yang tersedia adalah penerapan sistem cuaca penentuan saat (R. Chaniago, 2014)

Penentuan cuaca adalah proses penerapan ilmu pengetahuan dan teknologi untuk menentukan keadaan atmosfer bumi. Proses cuaca penentuan menggunakan metode untuk menghasilkan cuaca penentuan keluaran. Penggunaan metode yang baik, baru-baru ini menjadi suatu kegiatan yang sering dilakukan oleh para peneliti cuaca atau atmosfer. Untuk mendapatkan hasil yang lebih baik dan lebih akurat, peneliti didorong untuk terus meneliti atmosfer yang berhubungan dengan cuaca dan mengembangkan metode untuk memenuhi kebutuhan informasi tentang cuaca atau kondisi atmosfer.

Naïve bayes classifier adalah metode pengklasifikasian probabilistik sederhana, dapat diandalkan, mudah, efektif, efisien, dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset. Naive Bayes dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Saleh, 2015)

Karena cuaca ini banyak parameter yang mempengaruhi sehingga para petani sulit meramalkannya. Sering petani di pedesaan mengalami gagal panen sehingga mengalami kerugian yang besar, diakibatkan tidak hujan atau tingkat hujan terlalu tinggi, mengakibatkan tanaman tidak berkembang dengan baik. Solusinya ada banyak teknik data mining yang digunakan untuk prediksi cuaca, tapi evaluasi pohon keputusan dapat diukur. ramalan cuaca adalah menonjol diantara persyaratan alami terbaik dalam hidup kita. Saat ini, prakiraan cuaca dibuat melalui penerapan ilmu pengetahuan dan teknologi.

1.2 Tinjauan Pustaka

Studi literatur memainkan peran yang sangat penting dalam pengembangan proyek. Survei literatur memberikan pengetahuan yang diperlukan tentang proyek dan latar belakang. Hal ini juga membantu dalam mengikuti praktik terbaik dalam pengembangan proyek. Survei literatur juga membantu dalam memahami risiko dan kelayakan proyek. Kelayakan proyek tergantung pada risiko proyek. Jika sumber daya, waktu dan uang tidak tersedia untuk pengembangan proyek risikonya lebih tinggi. Survei literatur juga memberikan cahaya pada berbagai alat, platform dan sistem operasi yang cocok untuk pengembangan proyek. Setelah pemrograman dimulai programmer membutuhkan banyak dukungan dan saran.

Penelitian ini memiliki tujuan untuk dibandingkan metode klasifikasi yang berbeda yaitu Pohon Keputusan, Peraturan berbasis Metode, *Naïve Bayes, decision tree, Random forest.*

Banyak teknik data mining yang digunakan untuk peramalan cuaca dalam skenario ini, dengan berbagai tingkat akurasi. Dari literatur atas itu mengungkapkan bahwa ada karya-karya yang dilakukan mempertimbangkan Metode berbasis *Naïve Bayes, decision tree, Random forest.* Tapi tak satu pun dari mereka telah berusaha mengidentifikasi untuk pohon

Keputusan menggunakan dataset maka dalam pekerjaan ini dilakukan usaha untuk memprediksi cuaca di masa depan.

1.2.1 Algoritme Naïve Bayes

Navie Bayes classifier yang didasarkan pada teorema probabilitas yang disebut dengan teorema Bayes. Ini adalah algoritma yang baik untuk pemodelan prediktif. Naïve bayes banyak digunakan untuk mengklasifikasikan dataset pelatihan dimensi tinggi. Ini adalah *classifier* probabilistik sebagai penggunaannya teorema probabilitas untuk klasifikasi dan berfokus pada pembuatan anggapan bahwa keberadaan fitur tertentu independen dari keberadaan fitur, yang demikian disebut sebagai Navie.

Bayes adalah berasal dari teorema Bayes atau hukum yang memberi metode untuk probabilitas bersyarat bahwa probabilitas dari suatu peristiwa didasarkan pada beberapa pengetahuan sebelumnya yang tersedia pada peristiwa. Teknik naïve bayes didasarkan pada Bayes' teorema yang menyediakan posterior probabilitas:

$(P(C|A) \text{ dari } P(C), P(A) \text{ dan } P(A|C))$:

$$P(C|A) = \frac{P(A|C)P(C)}{P(A)} \quad (1)$$

Dalam rumus di atas, kita dapat mengetahui probabilitas (probabilitas bersyarat) dari terjadinya peristiwa C diberikan karena A adalah benar sama dengan probabilitas terjadinya peristiwa A mengingat karena C benar mengalikan dengan probabilitas C pada probabilitas A.

1.2.2. Algoritme Random Forest

Random Forest adalah metode pembelajaran diawasi di mana pohon klasifikasi yang dihasilkan. Dalam input algoritma ini berupa data vektor diletakan disetiap random forest untuk mengklasifikasikan objek baru dari input vektor fitur. Setiap pohon forest menyediakan setiap pohon dan pohon dengan tertinggi yang dianggap untuk klasifikasi. pseudo-code. Acak memilih “k” fitur dari total “m” fitur. Letakkan elemen kedalam beberapa elemen dengan menggunakan split terbaik. Algoritma ini mampu preforming yang baik untuk klasifikasi dan regresi.

1.2.3. Algoritme Decision Tree

Decision Tree adalah teknik pemodelan prediktif yang digunakan dalam klasifikasi, tugas prediksi dan clustering. Sebuah pohon keputusan menggunakan teknik yang disebut Divide and Conquer untuk membagi masalah ke jumlah subset. Pembangunan pohon keputusan mengandung dua hal yang umum:

- Pertumbuhan pohon memungkinkannya secara akurat mengkategorikan dataset pelatihan.
- Tahap pemangkasan, kita menghilangkan anomali dalam data pelatihan untuk meningkatkan akurasi klasifikasi.

Berdasarkan klaster disesuaikan dengan klasifikasi analisis disebut klasifikasi dengan pengelompokan diperkenalkan oleh Aviad dan Roy pada tahun 2011. Ini ditemukan kesamaan antara contoh menggunakan algoritma pengelompokan dan atribut target yang juga dipilih. Kemudian dihitung distribusi atribut target untuk setiap cluster. Ketika ambang batas untuk jumlah kasus yang disimpan dalam sebuah cluster tercapai, semua contoh disetiap klaster diklasifikasikan sehubungan dengan nilai yang sesuai dari atribut target.

Prediksi adalah proses untuk memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat

diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Herdianto, 2013: 8).

Pengertian Prediksi sama dengan ramalan atau perkiraan. Menurut kamus besar bahasa Indonesia, prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu.

Cuaca adalah keadaan udara pada saat tertentu dan di wilayah tertentu yang relatif sempit (tidak luas) dan pada jangka waktu yang singkat. Atau definisi cuaca ialah keadaan udara harian pada suatu tempat tertentu dan meliputi wilayah yang sempit, keadaan cuaca ini dapat berubah setiap harinya. Pengertian cuaca yang lainnya yaitu suatu keadaan rata-rata udara sehari-hari disuatu tempat tertentu & meliputi wilayah yang sempit dalam jangka waktu yang singkat. Keadaan dari cuaca mudah berubah – ubah, karena disebabkan oleh tekanan udara, suhu, angin, kelembaban udara, dan juga curah hujan. Seperti contohnya: Ketika pada pukul 07.00 WIB, keadaan cuaca di kota Bandung cerah suhu udaranya berkisar antara 23.C – 27.C, dan angin yang berhembus dari arah barat laut berkisar dengan kecepatan 25 km/jam, Akan tetapi ketika pada pukul 14.00 WIB keadaan cuacanya berubah menjadi berwarna tebal dengan suhu udara berkisaran 30.C, dan angin bertiup dari arah barat dengan kecepatan 20 km/jam.

1.2.3. Accuracy dan Confusion Matrix

Dalam bidang pembelajaran mesin dan khususnya masalah klasifikasi statistik, confusion matriks, juga dikenal sebagai kesalahan matriks. Confusion matriks adalah tabel yang sering digunakan untuk menggambarkan kinerja model klasifikasi (atau "classifier") pada satu set data pengujian yang nilai sebenarnya. Hal ini memungkinkan visualisasi kinerja algoritma. Sebagian besar tindakan kinerja dihitung dari confusion matriks.

Confusion matriks adalah ringkasan dari hasil prediksi pada masalah klasifikasi. Jumlah prediksi yang benar dan salah diringkas dengan menghitung nilai dan dicapai masing-masing kelas. Ini memberi kita wawasan tidak hanya ke dalam kesalahan yang dibuat oleh classifier tetapi yang lebih penting jenis kesalahan yang dibuat.

Tabel 1. Confusion Matrix

	Class 1 Prediksi Positif	Class 2 Prediksi Negatif
Class 1 Aktual Positif	TP	FN
Class 2 Aktual Negatif	FP	TN

Sumber : <https://www.geeksforgeeks.org/confusion-matrix-machine-learning>

Keterangan:

Class 1 = Positif

Class 2 = Negatif

True positive (TP): pengamatan positif, dan diprediksi akan positif.

False negative (FN): pengamatan positif, tetapi diprediksi negatif.

True negative (TN): pengamatan negatif, dan diperkirakan akan negatif.

False positive (FP): pengamatan negatif, tetapi diprediksi positif

Klasifikasi *rate* atau akurasi yang diberikan oleh relasi:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Recall dapat didefinisikan sebagai rasio jumlah total positif diklasifikasikan dengan benar membagi jumlah total positif. Tinggi menunjukkan kelas dikenali dengan benar (jumlah kecil FN).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Presisi adalah untuk mendapatkan nilai presisi, dengan membagi jumlah total positif yang diklasifikasikan dengan benar dengan jumlah total positif yang diprediksi. Presisi tinggi menunjukkan contoh berlabel positif memang positif (sejumlah kecil FP).

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall tinggi, presisi rendah adalah bahwa sebagian besar positif dikenali dengan benar (FN rendah) tetapi ada banyak false positif. Recall rendah, presisi tinggi adalah menunjukkan bahwa kita melewatkan banyak positif (tinggi FN) tetapi kita memprediksi sebagai positif memang positif (FP rendah)

F-Measure adalah memiliki dua langkah (Precision dan recall) itu membantu untuk memiliki pengukuran yang mewakili keduanya. Kita menghitung *F-Measure* yang menggunakan mean harmonik ditempatkan mean aritmatika. *F-Measure* akan selalu lebih dekat ke nilai yang lebih kecil dari Precision atau Recall.

$$F - Measure = \frac{2*Recall*Precision}{Recall+Precision} \quad (5)$$

1.2.3. Penelitian Terkait

Penelitian lain (Gupta & Ghose, 2015) peramalan curah hujan demikian penting dan diperlukan untuk pertumbuhan negara. Faktor cuaca termasuk suhu rata-rata, suhu titik embun, kelembaban, tekanan dari laut dan kecepatan angin dan telah digunakan untuk prakiraan curah hujan. classifier menggunakan Klasifikasi dan Regresi algoritma, Naive Bayes, K-NN dan Neural Network.

Penelitian lain (Nagendra, 2017) Data mining adalah salah satu bidang utama penelitian. Sebuah penelitian yang luas sedang dilakukan pada klasifikasi, yang merupakan salah satu fungsi dari data mining. Berbagai teknik klasifikasi seperti decision tree, Bayesian Classification, Naïve Bayes Classifiers, Jaringan Syaraf Tiruan, Multilayer Perceptron.

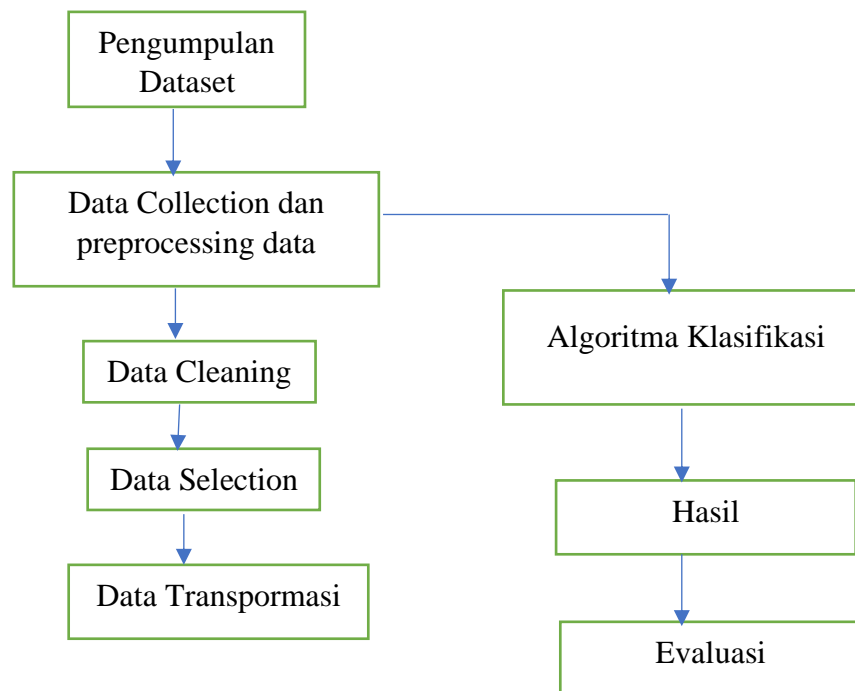
Penelitian lainnya (Riko Herwanto, Rosyana Fitria Purnomo, 2017) dilakukan untuk mengembangkan kerangka kerja metodologis yang memanfaatkan analisis penambangan data yang telah ditentukan (pohon keputusan). Model prediksi curah hujan berbasis keputusan mengembangkan peta variabel iklim, yaitu; a) suhu, b) kelembaban, dan c) kecepatan angin di atas basis data curah hujan yang diamati. Makalah ini menggunakan teknik penambangan data seperti *Teknik Clustering*, *Decision Tree* dan klasifikasi untuk prediksi curah hujan.

Penelitian lainnya (Tharun, Prakash, & Devi, 2018) Perkiraan curah hujan sangat penting dan diperlukan untuk pertumbuhan negara. Faktor cuaca termasuk suhu rata-rata, suhu titik embun, kelembaban, tekanan laut dan kecepatan angin dan telah digunakan untuk meramalkan curah hujan. Kumpulan data dari 2.245 sampel New Delhi dari Juni hingga September (periode curah hujan) dari 1996 hingga 2014 telah dikumpulkan dari sebuah situs web bernama Weather Underground. Dataset pelatihan digunakan untuk melatih pengklasifikasi menggunakan algoritma Klasifikasi dan Regresi, pendekatan Naive Bayes, K Neighbor terdekat dan 5-10-1 pattern recognition Neural Network dan akurasi diuji pada dataset uji. Akurasi pattern recognition Neural Network 82,1%, KNN 80,7% perkiraan

yang benar menempati urutan kedua, Klasifikasi dan Pohon Regresi CART 80,3% sementara Naive Bayes 78,9% sampel yang diklasifikasikan dengan benar.

2. METODOLOGI

Penelitian ini menggunakan beberapa metode yang bertujuan untuk memecahkan masalah yang ada dalam penelitian ini. Data yang digunakan dalam penelitian ini adalah Dataset ini berisi sekitar 10 tahun dan data sebanyak 142.193 dengan 24 atribut, pengamatan cuaca harian dari berbagai stasiun cuaca Australia. Target RainTomorrow berarti, apakah hujan dihari berikutnya? Ya atau tidak. Metode yang diusulkan dalam penelitian ini adalah sebagai berikut:



Sumber : Metode penelitian 2020
Gambar 1. Metode Penelitian

2.1 Pengumpulan Dataset

Dalam penelitian ini, ada beberapa tahap yang harus dilakukan pada data, semua data dan atribut tidak digunakan semua, atribut harus melalui beberapa tahap pengolahan data awal (data persiapan). Mendapatkan data yang berkualitas, beberapa teknik dilakukan sebagai berikut [6]: berikut attribute yang gunakan dalam penelitian ini.

Tabel 2. Atribut yang digunakan

Nama Atribut	Keterangan
<i>Date</i>	Tanggal observasi
<i>Location</i>	Lokasi
<i>MinTemp</i>	<i>Min</i> suhu dalam celcius
<i>MaxTemp</i>	<i>Max</i> suhu dalam celcius
<i>Rainfall</i>	Jumlah Hujan
<i>Evaporation</i>	<i>Evaporasi</i>
<i>Sunshine</i>	Jumlah Jam dalam hari
<i>WindGustDir</i>	<i>direction of the strongest wind gust in the 24 hours to midnight</i>
<i>WindGustSpeed</i>	<i>speed (km/h) of the strongest wind gust in the 24 hours to midnight</i>

<i>WindDir9am</i>	<i>Direction of the wind at 9am</i>
<i>WindDir3pm</i>	<i>Direction of the wind at 3pm</i>
<i>WindSpeed9am</i>	<i>Wind speed (km/hr) averaged over 10 minutes prior to 9am</i>
<i>WindSpeed3pm</i>	<i>Wind speed (km/hr) averaged over 10 minutes prior to 3pm</i>
<i>Humidity9am</i>	<i>Humidity (percent) at 9am</i>
<i>Humidity3pm</i>	<i>Humidity (percent) at 3pm</i>
<i>Pressure9am</i>	<i>Atmospheric pressure (hpa) reduced to mean sea level at 9am</i>
<i>Pressure3pm</i>	<i>Atmospheric pressure (hpa) reduced to mean sea level at 3 pm</i>
<i>Cloud9am</i>	<i>Fraction of sky obscured by cloud at 9am</i>
<i>Cloud3pm</i>	<i>Fraction of sky obscured by cloud (in "oktas": eighths) at 3pm</i>
<i>Temp9am</i>	<i>Temperature (degrees C) at 9am</i>
<i>Temp3pm</i>	<i>Temperature (degrees C) at 3pm</i>
<i>RainToday</i>	<i>Hujan hari ini</i>
<i>RISK_MM</i>	<i>A kind of measure of the "risk".</i>
<i>RainTomorrow</i>	<i>Hujan besok</i>

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

2.2 Data cleaning

Kualitas data yang baik adalah kunci dasar untuk data yang menghasilkan kualitas yang baik, noise data yang masih outlier atau noise, data yang tidak lengkap dihilang nilai atribut, dan data yang tidak konsisten dalam atribut pengisian. Tahapan dalam pembersihan data:

1. Menghilangkan atau mengidentifikasi outlier.
2. Melengkapi nilai-nilai yang tidak lengkap atau hilang (missing value), algoritma naif bayes memiliki keunggulan tersendiri yang dapat menangani data yang tidak lengkap atau hilang (missing value).
3. Data yang tidak konsisten adalah tetap. memecah redundansi yang disebabkan oleh interogasi data.

2.3 Data transformasi

Langkah selanjutnya adalah teknik integrasi data yang digunakan untuk menganalisis korelasi data, atribut berlebihan dan data ganda, dan transformasi digunakan untuk meningkatkan akurasi dan efisiensi algoritma. Kelebihan algoritma naif bayes mampu memproses data yang nominal, terus menerus, dan ordinal. Oleh karena itu nilai setiap atribut yang terkandung dalam dataset tidak harus diubah.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

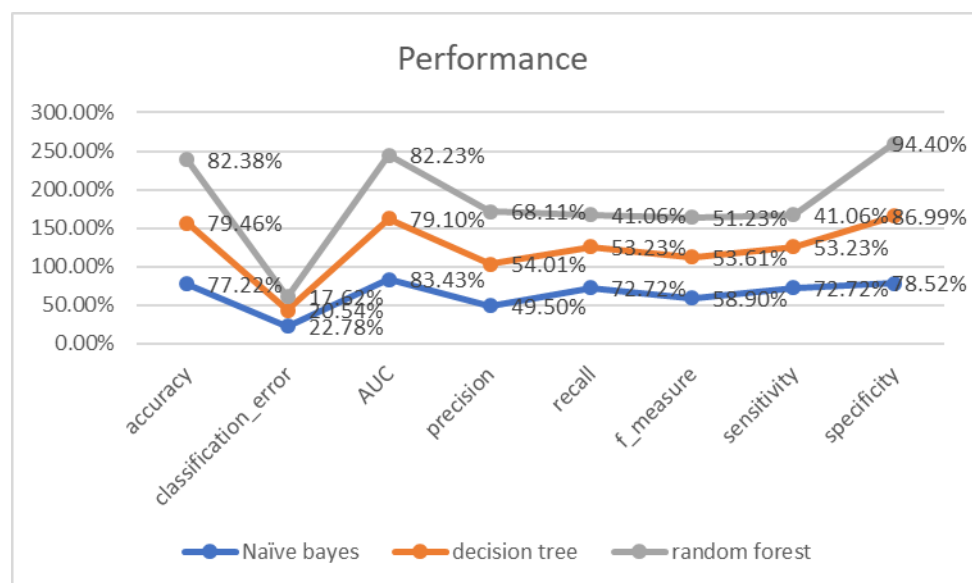
Perubahan iklim merupakan masalah besar bagi seluruh dunia dan membuat prediksi apapun pada yang sekarang, cukup sulit dan tak terduga. Perubahan iklim adalah karena pemanasan global adalah ekspansi manusia yang membuat bertambah sulit memecahkan masalah cuaca. Hasil dari penelitian ini dapat dijadikan untuk mengukur tingkat presisi prediksi terhadap cuaca. Untuk dapat di buat aplikasi prediksi cuaca, dengan melihat tingkat akurasi penelitian ini. Berikut ini adalah tabel dari hasil akurasi algoritma. Tabel berikut random forest muncul sebagai algoritma terbaik untuk prediksi curah hujan.

Tabel 3. Performance Accuracy Algoritme Klasifikasi

Meto de	Accura cy	classificat ion_error	AUC	precisio n	recall	f_meas ure	sensiti vity	specif icity
Naïve bayes	77.22 %	22.78%	83.43%	49.50%	72.72 %	58.90%	72.72%	78.52 %
decisi on tree	79.46 %	20.54%	79.10%	54.01%	53.23 %	53.61%	53.23%	86.99 %
rando m forest	82.38 %	17.62%	82.23%	68.11%	41.06 %	51.23%	41.06%	94.40 %

Sumber : Hasil penelitian 2020

Berdasarkan tabel diatas, menjelaskan ketiga metode yang menggunakan akurasi sebagai ukuran keberhasilan algoritme dimana akurasi semakin tinggi semakin baik algoritme tersebut dimana yang paling adalah *algoritme Random forest*, untuk *classification error* ada 1 – *accuracy*. Jadi semakin tinggi akurasi maka semakin rendah *classification error*.



Sumber : Hasil penelitian 2020

Gambar 2. Grafik Performance Accuracy Algoritme Klasifikasi

Gambar 2 adalah grafik yang menunjukkan semua alat ukur untuk mengukur *performance algoritme* yang digunakan. Untuk AUC lebih baik *naïve bayes* dan presisi *random forest* lebih baik, *recall* yang terbaik adalah *naïve bayes*, *F measure* yang paling baik adalah *radom forest*. Dan *sensitivity* dan *specificity* bisa dilihat dalam gambar 2.

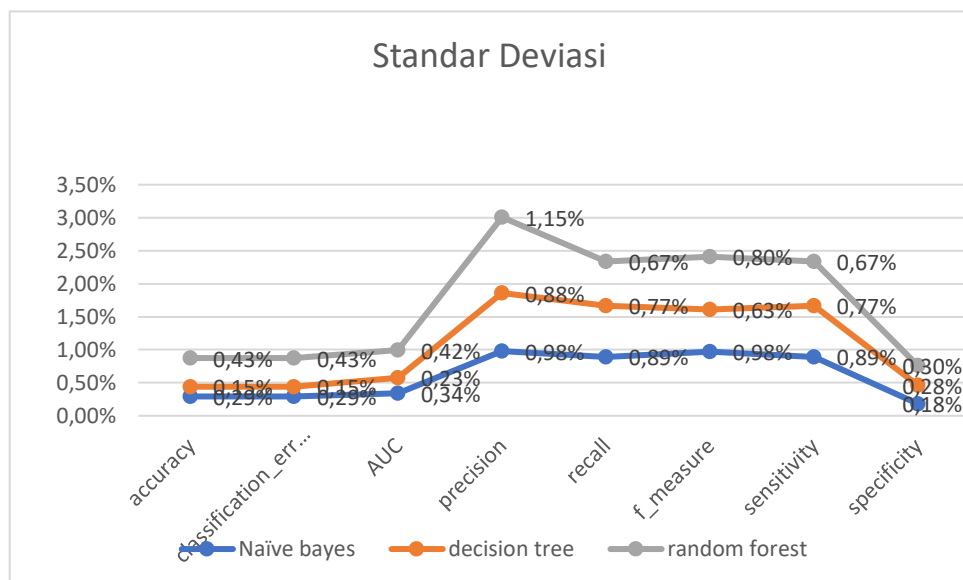
Tabel 4. Performance Standard Deviasi Algoritme Klasifikasi

Metode	accura cy	classificati on_error	AUC	Precisio n	recall	f_measur e	sensitiv ity	specif icity
Naïve bayes	0.29%	0.29%	0.34%	0.98%	0.89%	0.98%	0.89%	0.18%
decision tree	0.15%	0.15%	0.23%	0.88%	0.77%	0.63%	0.77%	0.28%
random	0.43%	0.43%	0.42%	1.15%	0.67%	0.80%	0.67%	0.30%

forest								
--------	--	--	--	--	--	--	--	--

Sumber : Hasil penelitian 2020

Tabel 4. Mengukur standard deviasi setiap algoritme yang digunakan. Pada penelitian ini yang paling besar adalah *random forest*, untuk *classification error*, *AUC* *precision*, *recall*, *f-measure*, *sensitivity* dan *specificity* dapat lihat dalam tabel. Ini semua parameter untuk mengukur algoritme.



Sumber : Hasil penelitian 2020

Gambar 3. Grafik Performance Standard Deviasi Algoritme Klasifikasi

3.2. Pembahasan Penelitian

Berdasarkan hasil penelitian diatas, evaluasi penelitian menggunakan confusion matrix dengan melihat beberapa factor. Accuracy salah satu yang paling populer digunakan, sebelum membuat klasifikasi kita harus mengukur kinerja algoritme yang akan digunakan, seperti dalam penelitian ini dengan membandingkan tiga algoritme yaitu.

1. Algoritme naïve bayes dengan accuracy 72.22 %, presisi 49.50%, F-measure 78.52%
2. Algoritme Decision tree dengan accuracy 54.01 %, presisi 49.50%, F-measure 53.61%
3. Algoritme Random forest dengan accuracy 82.38 %, presisi 68.11%, F-measure 51.230%

Accuracy, Presisi yang paling baik Random forest tetapi memiliki *F-measure* lebih rendah. Dalam klasifikasi tidak lepas dari tingkat deviasi algoritme. Untuk standar deviasi lebih baik dipilih yang lebih kecil. Accuracy, presisi dan *F-measure* faktor penting untuk kesuksesan prediksi cuaca.

4. PENUTUP

4.1 Kesimpulan

Penelitian ini telah diuji dengan dataset curah hujan, model *supervised learning*. Jadi hasil penelitian ini dapat di jadikan referensi untuk membangun *machine learning* dengan nilai accuracy paling tinggi. Dalam prediksi Cuaca cukup membantu untuk memahami algoritme dan jenis data yang digunakan. Sebelum membangun aplikasi prediksi cuaca, seharusnya melihat hasil penelitian ini menggunakan akurasi yang tertinggi ada beberapa kesimpulan yang didapat sebagai berikut:

1. Algoritme Naïve bayes yang paling dari F-measure yaitu 78.52 %
2. Algoritme Random Forest yang paling dari Presisi yaitu 68.11 %
3. Algoritme Random Forest yang paling dari Accuracy yaitu 82.38 %
4. Standard deviasi lebih baik itu yang paling kecil yaitu Algoritme Decision tree

4.2 Saran

Berdasarkan kesimpulan penelitian ini, peneliti memberikan saran yaitu dapat mencari jenis algoritme lain. Dapat menjadi usulan menggunakan metode ini untuk prediksi cuaca yang ada di Indonesia karena pada dasarnya variable yang digunakan sama dengan dataset diluar negeri.

DAFTAR PUSTAKA

- Gupta, D., & Ghose, U. (2015). A comparative study of classification algorithms for forecasting rainfall. *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, 1–6. <https://doi.org/10.1109/ICRITO.2015.7359273>
- Nagendra, K. (2017). A SURVEY ON CLASSIFICATION TECHNIQUES USED FOR RAINFALL FORECASTING. *International Journal of Advanced Research in Computer Science*, 8(8), 226–229. <https://doi.org/10.26483/ijarcs.v8i8.4645>
- R. Chaniago, T. H. L. dan K. R. R. W. (2014). Prediksi Cuaca Menggunakan Metode Case Based Reasoning dan Adaptive Neuro Fuzzy Inference System. *Jurnal Informatika*, 12.
- Riko Herwanto, Rosyana Fitria Purnomo, S. (2017). RAINFALL PREDICTION USING DATA MINING TECHNIQUES. *3rd International Conferences on Information Technology and Business (ICITB)*, 188. Retrieved from https://www.researchgate.net/publication/327933654_Prediction_of_Rainfall_Using_Data_Mining_Techniques
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*, 2(3), 207. Retrieved from <https://citec.amikom.ac.id/main/index.php/citec/article/view/49>
- Tharun, V. ., Prakash, R., & Devi, S. R. (2018). Prediction of Rainfall Using Data Mining Techniques. *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 1507–1512. <https://doi.org/10.1109/ICICCT.2018.8473177>