**Analisis Cuaca di Szeged pada Tahun 2006-2016 Menggunakan Model Naïve Bayes**

**Viqram Ananta Wataf1, Ken Dahana2**

1,2Program Studi Teknik Informatika, Universitas Surabaya, Jawa Timur

Email: 1s160420119@student.ubaya.ac.id, 2s160420115@student.ubaya.ac.id

(Naskah masuk: 14 Des 2022, direvisi: dd mmm yyyy, diterima: dd mmm yyyy)

**Abstrak**

Peramalan cuaca adalah proses teknologi dan ilmu pengetahuan yang digunakan untuk memprediksi kondisi atmosfer untuk lokasi yang diberikan. Ilmuwan juga telah berusaha memprediksi cuaca secara formal dan informal. Analisis cuaca di Szeged menggunakan model naive bayes merupakan sebuah studi yang bertujuan untuk mengklasifikasikan cuaca di Szeged berdasarkan data historisnya. Model Naïve Bayes Classifier digunakan untuk memodelkan hubungan sebab-akibat ini dari data survei lapangan menggunakan library Scikit-learn. Model naive bayes dipilih karena kemudahannya dalam menangani data yang bersifat kategorikal dan memiliki tingkat akurasi yang cukup tinggi. Data yang ada di dataset dibagi menjadi data training dan testing. Data input yang digunakan pada penelitian ini berasala dari data Kaggle Cuaca di Szeged pada Tahun 2006-2016 terdiri dari 96453 data dan 12 fitur. Hasil dari analisis ini menunjukkan bahwa model naive bayes mendapatkan akurasi terbaik sebesar 92.964% dengan metode yang efektif untuk mengklasifikasikan cuaca di Szeged. Selain itu, hasil analisis juga menunjukkan bahwa faktor-faktor seperti suhu, kelembaban, dan tekanan udara berperan penting dalam mempengaruhi cuaca di Szeged.

# **Kata Kunci:**Naïve Bayes, Cuaca, Klasifikasi, Analisis, Hujan.

***Weather analysis in Szeged at 2006-2016 using* Naïve Bayes *Model***

***Abstract***

Weather forecasting is a technological and scientific process used to predict atmospheric conditions for a given location. Scientists have also sought to predict the weather formally and informally. Weather analysis in Szeged using the naïve bayes model is a study that aims to classify the weather in Szeged based on historical data. The Naïve Bayes Classifier model was used to model this cause-and-effect relationship from field survey data using the Scikit-learn library. The naïve bayes model was chosen because of its ease in handling categorical data and has a high level of accuracy. The data in the dataset is divided into training and testing data. The input data used in this study is derived from the Weather Kaggle data in Szeged in 2006-2016 consisting of 96453 data and 12 features. The results of this analysis showed that the naïve bayes model got the best accuracy of 92.964% with an effective method for classifying the weather in Szeged. In addition, the results of the analysis also show that factors such as temperature, humidity, and air pressure play an important role in influencing the weather in Szeged.

# ***Keywords:*** *Naïve Bayes, Weather, Classification, Analysis, Rain.*

# **I. PENDAHULUAN**

Prakiraan cuaca telah menjadi yang paling menonjol di antara yang paling masalah yang menyusahkan secara eksperimental dan teknologi dunia pada abad terakhir [1]. Lingkungan Perubahan telah mencari banyak pertimbangan sejak lama karena perubahan mendadak yang terjadi. Ada beberapa batasan dalam pelaksanaan cuaca yang lebih baik meramalkan sehingga akhirnya sulit memprediksi cuaca di sini dan sekarang dengan efektivitas [2][3]. Peramalan cuaca mengasumsikan peran penting dalam meteorologi [4]. Untuk membuat yang tepat prediksi adalah salah satu masalah signifikan yang dihadapi ahli meteorologi di seluruh dunia. Untuk membuat yang tepat prediksi adalah salah satu masalah signifikan yang dihadapi ahli meteorologi di mana pun di seluruh dunia. Peringatan cuaca sangat penting mengingat fakta bahwa mereka digunakan untuk memastikan kehidupan sehari-hari. Prakiraan tergantung pada suhu, kondisi hujan, kelembaban dan angin penting untuk pertanian dan pedagang pasar. Prakiraan suhu digunakan oleh perusahaan utilitas untuk menilai permintaan selama beberapa hari mendatang. Karena kegiatan di luar ruangan adalah sangat berkurang oleh hujan besar, salju dan angin dingin, perkiraan dapat digunakan untuk merancang aktivitas di sekitar kegiatan, dan untuk mempersiapkan dan bertahan hidup mereka. Jika prediksi cuaca tidak akurat maka tiap individu dapat berakhir di situasi yang berbahaya karena mereka tidak siap akhirnya dirugikan atau lebih buruk [1]. Kesulitan prakiraan cuaca, antara lain, adalah mempelajari representasi cuaca menggunakan volume yang sangat besar himpunan data cuaca. Untuk tujuan ini, dilakukan analisis *data mining* yang berbeda. Teknik *data mining* memungkinkan pengguna untuk menganalisis data dari berbagai dimensi atau sudut, mengklasifikasikannya. Beberapa istilah mendasar yang terkait dengan Data Mining adalah: Klasifikasi, Pembelajaran dan Prediksi.

Szeged merupakan salah satu kota di Hungaria yang memiliki iklim yang cukup ekstrim, dengan musim panas yang panas dan musim dingin yang sangat dingin. Untuk memprediksi cuaca di Szeged, kami menggunakan metode analisis naive bayes. Naïve Bayes Algorithm adalah teknik klasifikasi berdasarkan Teorema Bayes. Naïve Bayes mudah dibangun dan sangat banyak berguna untuk himpunan data besar. Dengan menggunakan persamaan Naïve Bayes kita dapat menemukan probabilitas masa depan [6]. Seperti namanya Naive menunjukkan bahwa semua variabel berkontribusi terhadap klasifikasi dan saling menguntungkan berkorelasi atau independen. Asumsi ini disebut kelas Independensi Bersyarat. Pengklasifikasi ini menganggap bahwa ada (atau tidak adanya) fitur(atribut) kelas tertentu tidak terkait dengan keberadaan (atau absensi) dari fitur lain ketika variabel kelas. Metode ini merupakan salah satu metode klasifikasi yang populer digunakan dalam bidang data mining dan machine learning. Metode ini menggunakan sebuah teori probabilitas yang mengasumsikan bahwa setiap atribut yang ada pada data memiliki kemungkinan yang independen satu sama lain. Klasifikasi adalah Metode *data mining* *(machine learning)* yang digunakan untuk memprediksi partisipasi agregat untuk kasus informasi. Misalnya klasifikasi dapat digunakan untuk memprediksi apakah cuaca pada hari tertentu akan "cerah", "hujan" atau "berawan" [1]. Pembelajaran mengacu pada kontribusi pelatihan dan pemetaan untuk informasi hasil. Ini cenderung dilakukan dalam dua yang berbeda cara: Pembelajaran yang diawasi dan Tidak Diawasi. Seorang yang diawasi algoritma pembelajaran menganalisis data pelatihan dan menghasilkan kapasitas turunan menggunakan Pengklasifikasi [5]. Dalam machine learning, *unsupervised learning* menyinggung masalah mencoba menyembunyikan struktur dalam informasi yang tidak berlabel [1]. Sejak preseden diberikan kepada peserta didik tidak berlabel, tidak ada kesalahan atau sinyal hadiah untuk menilai solusi potensial. Ini membuktikan unsupervised learning dari supervised learning. Prediksi mengidentifikasi dengan pemodelan dan hubungan logis dari model cepat atau lambat. Menemukan pola dan data mungkin meminta prediksi yang masuk akal [5]. Penelitian serupa telah dilakukan oleh. Anusha et.al, Asha P, et.al) algoritma yang berbeda yang dapat digunakan untuk prediksi curah hujan dan mengklasifikasi ke dalam berbagai jenis seperti berawan, sebagian keruh, mendung penuh, dll., Dari mana beberapa metode memprediksi nilai numerik tetapi setiap metode sebagai kelebihan dan kekurangannya sendiri. Dalam sistem ini, keadaan cuaca diklasifikasikan dalam beberapa atribut seperti suhu, kondisi hujan, kelembaban dan angin. Dalam sistem terdapat dua fungsi dasar yaitu klasifikasi (training) dan prediksi (testing).

# **II. METODOLOGI PENELITIAN**

Terdapat lima tahapan yang dilakukan dalam penelitian ini. Tahapan-tahapan tersebut meliputi pengumpulan *dataset*, *pre-processing dataset*, pembentukan model, *training-testing*, serta perhitungan performa. Gambar 1, disajikan diagram blok alur proses penelitian secara lebih detail.

Sebuah gambar berisi teks

Deskripsi dibuat secara otomatis

Gambar 1. Diagram Blok Alur Proses Penelitian

## Pengumpulan Dataset

Tahap awal dalam penelitian ini adalah Pengumpulan *dataset*. Namun tahap krusial adalah *pre-processing* data, karena hanya data yang valid yang dapat menghasillkan *output* yang akurat. Dalam penelitian ini *Dataset* yang digunakan diperoleh dari *Kaggle* *Data*. Berisi kumpulan 96453 data cuaca yang berbeda-beda dengan 12 fitur. Meskipun kumpulan data-data berisi banyak fitur/atribut, maka dilakukan proses pre-processing untuk memfilter data-data yang penting dan relevan akan di ambil untuk proses selanjutnya. Data-data di ambil dari tahun 2006-2016. Pada tabel 1 disajikan jumlah data dan fitur yang ada dalam *dataset*.

Tabel 1. *Dataset* Cuaca di Szeged Tahun 2006-2016

|  |  |
| --- | --- |
| **Jumlah Data** | **Fitur/Atribut** |
| 96453 | 12 |

Sedangkan pada tabel 2 disajikan visualisasi detail data pada *dataset* yang digunakan untuk penelitian ini.

Tabel 2. Detail Fitur/Atribut *Dataset* Cuaca di Szeged Tahun 2006-2016

Tabel 2. Detail Data Pada *Dataset* yang Di Analisis

|  |  |  |
| --- | --- | --- |
| **No** | **Nama** | ***Class*** |
| 1 | Summary | Partly Cloudy, Mostly Cloudy, Overcast, Foggy, Breezy and Mostly Cloudly, Clear |
| 2 | Precip Type | Rain, Snow, Hail |

## Pembuatan Model dan Training-Testing

Data yang harus diklasifikasikan disebut data *training*. Dengan menggunakan data ini dengan data *testing*, dapat memperkirakan kondisi cuaca. Model *training* set pelatihan dan pengetahuan itu digunakan sebagai data *testing* untuk memprediksi nilai.Jumlah data yang digunakan untuk data *training* adalah 70% sedangkan sisanya digunakan untuk data *testing.*

Berikut detail konfigurasi *environment* perangkat keras yang dipakai selama uji coba.

Sebuah gambar berisi meja

Deskripsi dibuat secara otomatis  
Gambar 5. Spesifikasi Perangkat Keras

## Perhitungan Performa

Pada tahapan ini, dilakukan proses perhitungan metriks performa Naïve Bayes yang didasarkan pada Teorema Bayesian. Persamaan (1) menunjukkan perhitungan atau rumus dari teorema bayesian. Kata naive mengacu pada kondisi independen antar fitur. Dua bilangan A dan B adalah iff independen. .

(1)

Persamaan (2) menunjukkan dari masalah yang akan di klasifikasikan. Mengingat masalah klasifikasi kelas K dengan kelas dan pengamatan yang tidak diketahui . adalah posterior, yaitu probabilitas bahwa suatu pengamatan berada di kelas mengingat nilai pengamatan . Mengingat kelasnya . adalah bukti atau probabilitas marjin dari .

(2)

Persamaan (3) menunjukkan rumus dari perhitungan berdasarkan asumsi naïve independen.

(3)

Karena tidak bergantung dari , maka pengklasifikasi Naïve Bayes memberikan pengamatan ke kelas untuk beberapa k yang memenuhi syarat.

Persamaan (4) menunjukkan rumus dari perkirakan probabilitas menggunakan *MLE (maximum likelihood estimator)*.

(4)

Jika beberapa hasil perhitungan sama dengan nol, perkirakan probabilitas konditinal menggunakan *MAP (maximum a posteriori)* kebalikan dari *MLE (maximum likelihood estimator).*

Persamaan (5) menunjukkan perhitungan dari naïve bayes training untuk fitur kategori. Tambahkan 1 ke semua jumlah *MLE (maximum likelihood estimator)* dan normalkan untuk mendapatkan distribusi yang sebenarnya (jumlah probabilitas adalah 1).

(5)

adalah jumlah kategori yang tersedia di

Persamaan (6) menunjukkan perhitungan dari *Gaussian Naïve Bayes*. Jika adalah variabel lanjutan, maka asumsikan bahwa distribusi bersyarat adalah distribusi normal (*Gaussian*).

(6)

Dimana:

adalah rata-rata yang

adalah varians dari yang

# **III. HASIL DAN PEMBAHASAN**

1. Percipitation Type

Seperti yang telah diungkit sebelumnya, ada 12 fitur yang terdapat pada dataset namun hanya 7 fitur yang dapat membantu kami dalam menentukan nilai variabel apa saja yang sesuai. Lima fitur yang tidak terpakai itu berisi string yang tidak relevan atau kolom hanya berisi 0 saja.

Setelah membuang fitur-fitur yang tidak dibutuhkan, kami akan membersihkan setiap kolom dari nilai null atau NaN. Pada dataset yang kami miliki, kolom “Percip Type” yang akan digunakan sebagai variabel yang diprediksi memiliki beberapa nilai null pada training dan testnya. Setelah kami analisis baris-baris yang memiliki nilai “Percip Type” null, mereka semua memiliki suhu yang tidak cukup tinggi untuk terjadinya hujan air namun tidak cukup rendah untuk terjadinya hujan salju, sehingga kami memutuskan untuk mengganti nilai null ini sebagai “hail”, “hail” itu kondisi dimana air yang turun berada diantara cair dan salju sehingga hanya terbentuk es layaknya es batu bening biasa.

Agar dapat dibaca oleh library Python, nilai-nilai dari kolom “Percip Type” yang berbentuk string harus dikodifikasi menjadi angka. Contoh hasil kodifikasi Python sebagai berikut:

|  |  |
| --- | --- |
| **Original Percip Type** | **Coded Percip Type** |
| *Rain* | 0 |
| *Hail* | 1 |
| *Snow* | 2 |

Tabel 3. Contoh kodifikasi Percipitation Type

Setelah membersihkan dataset dari fitur-fitur yang tidak dibutuhkan dan nilai yang memiliki null serta mengkodifikasi nilai-nilai yang muncul pada kolom “Percip Type”, kami akan menghitung keseluruhan probabilitas setiap variabel yang ada, caranya dengan mengkali setiap probabilitas dari suatu variabel sehingga terjadi ikatan “dan” pada seluruh variabel yang ada, kemudian hasil yang akan diambil yaitu nilai prediksi yang memiliki nilai probabilitas yang terbesar.

Output yang kami harapkan pada penelitian kali ini yaitu Percipitation Type dan Summary yang sesuai. Fitur-fitur input dan dapat menyokong hasil output adalah, Temperature, Apparent Temperature, Humidity, Wind Speed, Wind Bearing, Visibility, dan Pressure. Kesemua fitur-fitur ini adalah komponen yang penting dalam mengetahui tipe-tipe hujan apa saja yang akan muncul.

Temperature dan Apparent Temperature memiliki peran yang sangat penting dikarenakan semakin tinggi suhu maka uap air yang mengembun ketika turun akan berbentuk cair karena petikan hujan tersebut akan meleleh selama perjalanannya turun ke permukaan bumi.

Humidity memiliki peran yang lumayan penting juga dikarenakan dari nilai tersebut dapat diperkirakan kira-kira pengembunan dapat terjadi dengan cepat atau tidak. Semakin lembab udara, ditunjukkan dengan angka yang mendekati 1, maka semakin banyak partikel air yang ada di udara yang menyebabkan air dipermukaan kesulitan untuk mengembun dan mencapai lapisan bumi tempat awan terbentuk.

Wind speed juga berpengaruh karena semakin cepat angin bergerak maka udara, partikel air, dan awan akan semakin cepat berpindah tempat. Visibility juga berpengaruh karena visibility dapat menunjukkan seberapa tebal awan yang ada di suatu lokasi, hal ini mengasumsikan bahwa kabut yang terjadi berasal dari uap air dan bukan partikel lain seperti asap motor. Pressure pada hal ini cukup berperan walau tidak seberpengaruh fitur-fitur yang lain karena pressure mempengaruhi seberapa banyak energi yang dibutuhkan air untuk menguap ke atmosfer.

Setelah mengetahui fungsi masing-masing fitur, kita dapat memulai perhitungan. Pertama, kami harus menentukan berapa persen dari total data di dataset yang akan menjadi training data dan berapa yang menjadi testing data. Pada penelitian kali ini, kami memutuskan untuk menggunakan perbandingan 30%-70%, pada data testing akan diambil sebanyak 30% dari dataset yang ada sedangkan 70% digunakan untuk testing.

Kemudian kami melakukan perhitungan korelasi antara variabel-variabel yang ada dengan variabel “Percip Type” dan kami dapatkan bahwa seluruh variabel memiliki nilai bobot absolute kurang dari 1 sehingga dapat digunakan semestinya. Cara yang kami gunakan yaitu looping setiap fitur input testing satu-persatu, kemudian kami bandingan hubungan fitur input testing tersebut dengan fitur output testing. Baru setelah itu kami mendapatkan nilai korelasinya

Tabel 4. Hasil korelasi fitur-fitur yang ada dengan Percipitation Type

|  |  |
| --- | --- |
| **Fitur** | **Nilai Korelasi** |
| *Temperature (C)* | -0.5422 |
| *Apparent Temperature (C)* | -0.5453 |
| *Humidity* | 0.2239 |
| Wind Speed (km/h) | -0.0683 |
| Wind Bearing (degrees) | -0.0404 |
| Visibility (km) | -0.292 |
| Pressure (millibars) | 0.0065 |

Sesudah melakukan pembagian mana saja yang menjadi training dan testing, kami akan memproses model training yang telah dipisah menggunakan library Python sklearn yang bernama Naive Bayes. Pada package naive bayes ini, terdapat satu fungsi yang bernama Gaussian Naive Bayes yang akan digunakan pada penelitian kali ini. Pertama kita harus menginisialisasi objek gaussian naive bayes menggunakan .GaussianNB(), kemudian kami akan memasukkan data training tersebut ke fungsi yang bernama .fit(x,y). Setelah model Gaussian Naive Bayes terbentuk, maka kita akan memprediksi jenis hujan apa saja yang cocok dengan dataset testing. Setelah diprediksi, kami akan mengevaluasi seberapa cocok nilai Precip Type yang terprediksi dengan Precip Type yang sebenarnya. Skor akurasi yang kami dapatkan berupa 92% yang mana sangat cocok untuk memprediksi perkiraan tipe hujan. Dengan ini, hasil yang kami peroleh memiliki tingkat akurasi yang cukup tinggi untuk dijadikan pada kota Szeged.

Prediksi jenis hujan sangatlah membantu para penduduk yang tinggal di daerah itu serta bisnis-bisnis yang beroperasi di daerah sekitar. Hal ini dapat meningkatkan efisiensi perekonomian di daerah sekitar dikarenakan tidak ada bahan baku atau tenaga masyarakat dan/atau perusahaan yang terbuang sia sia dikarenakan tidak mengetahui kondisi cuaca yang akan muncul kedepannya. Mereka dapat mempersiapkan dengan baik-baik sesuai dari ramalan yang akan muncul mengenai jenis hujan yang akan muncul sehingga tidak terjadi kerusakan akibat hujan es, hujan salju, dan juga hujan air yang muncul.

Kerusakan yang diakibatkan oleh hujan es yaitu kemungkinkan munculnya korban jiwa dan kemungkinan munculnya kerugian harta material seperti es yang menghantam jendela mobil. Kerusakan yang muncul diakibatkan oleh hujan salju yaitu kendaraan yang tergelincir akibat jalanan yang membeku, kemungkinan lansia tergelincir dengan akibat yang sama, serta kemungkinan mesin yang tidak dapat jalan akibat mesin yang beku akibat hujan salju. Terakhir, kerusakan yang diakibatkan oleh hujan air yaitu mesin yang mogok akibat banjir di kota tersebut, adanya kemungkinan badai dan petir yang dapat membahayakan nyawa, serta kemacetan yang diakibatkan oleh jalanan yang banjir.

Model yang dihasilkan pada pemrediksian tipe hujan ini hanya menggunakan dataset dari Szeged saja, namun tidak menutup kemungkinan dapat digunakan oleh berbagai daerah lain dan bahkan negara tropis seperti Indonesia. Namun, karena Indonesia mayoritas suhu permukaan berkisar 26-28°C maka hujan yang dihasilkan sebagian besar akan berupa hujan air atau hujan es jika suhu permukaan mencapai dibawah 20°C

2. Summary

Pada bagian ini, kami mendapatkan satu fitur yang patut dicoba untuk mendapatkan hasil prediksinya yaitu variabel “Summary”. Pada dataset Szeged ini, fitur Summary tidak memiliki nilai yang terlalu banyak, hanya sebanyak 27. Fitur “Summary” ini hanya berisi catatan yang sangat singkat, hanya 3-4 kata saja. Jika hasilnya akurat, maka akan dapat membantu pembaca perkiraan cuaca dalam menentukan kondisi cuaca secara garis besar, misalnya mengetahui ketebalan awan, mengetahui kecepatan udara, mengetahui jenis hujan, mengetahui bagaimana sensasi cuacanya, dan sebagai macam.

Hal ini dapat bermanfaat bagi pembaca perkiraan cuaca yang bisa jadi adalah masyarakat umum yang tidak dapat menentukan secara pasti kondisi cuaca jika variabel lain hanya menunjukkan detail berupa angka, kecepatan angin, kelembaban, dan berbagai macam. Mereka dapat langsung menarik kesimpulan sehingga dapat mengurangi beban kognitif pembaca awam.

Sama seperti perkiraan variabel “Percip Type”, hal ini dapat berguna bagi perusahaan dan masyarakat umum dalam menentukan aktifitas harian apa yang akan dilakukan sehingga dapat meningkatkan efisiensi ekonomi masyarakat sekitar.

Sebelumnya, sudah kami singgung bahwa nilai-nilai unik dari variabel “Summary” ada 27. Semua nilai-nilai tersebut yaitu: Partly Cloudy, Mostly Cloudy, Overcast, Foggy,Breezy and Mostly Cloudy, Clear, *Breezy and Partly Cloudy*, *Breezy and Overcast*, *Humid and Mostly Cloudy*, *Humid and Partly Cloudy*, *Windy and Foggy*, *Windy and Overcast*, *Breezy and Foggy*, *Windy and Partly Cloudy* , *Breezy*, *Dry and Partly Cloudy*, *Windy and Mostly Cloudy*' *Dangerously Windy and Partly Cloudy*, *Dry*, *Windy*, *Humid and Overcast*, *Light Rain*, *Drizzle*, *Windy and Dry*, *Dry and Mostly Cloudy*, *Breezy and Dry*, dan *Rain.*

Jika melihat nilai-nilai yang disediakan pada fitur “Summary” ini, yang perlu diperhatikan adalah cuaca berpetir, berangin keras, berhujan lebat, bersuhu dingin, dan bersuasana terlalu kering atau lembab, dan terlau bertampak kabut. Hal itu dikarenakan, banyak sekali kejadian yang yang memakan korban jiwa jika terjadi situasi seperti diatas.

Pada penelitian fitur “Summary” ini, kami mencoba tidak membutuhkan *preprocessing data* dikarenakan fitur-fitur input dan fitur output tidak memiliki nilai null. Lanjut ke bagian pemrosesan, kami menggunakan teknik yang sama seperti bagian “Percipitation Type” yaitu mengodifikasi nilai-nilai string yang ada pada pada fitur “Summary”.

Setelah itu, kami membagi dataset untuk training dan testing sebanyak 60%-40%. Setelah membagi dataset menjadi training dan testing, kami akan melakukan perhitungan angka korelasi antara fitur-fitur yang ada dengan fitur Summary.

|  |  |
| --- | --- |
| **Fitur** | **Nilai Korelasi** |
| *Temperature (C)* | 0.1467 |
| *Apparent Temperature (C)* | 0.1428 |
| *Humidity* | -0.1131 |
| Wind Speed (km/h) | 0.0043 |
| Wind Bearing (degrees) | 0.0228 |
| Visibility (km) | 0.1401 |
| Pressure (millibars) | 0.1434 |

Tabel 5. Hasil korelasi fitur-fitur yang ada dengan Summary

Setelah kami proses hasilnya menggunakan Gaussian Naive Bayes, sayang sekali akurasi yang didapatkan hanya 44.44%. Ternyata, yang menyebabkan hasilnya cukup kecil adalah karena banyaknya jumlah value unik yang ada serta jumlah dataset yang tergolong terbatas untuk pengategorian nilai yang memiliki banyak variasi. Dengan ini, kami mendapatkan keakurasian prediksi cukup rendah sehingga model yang didapat ini tidak dapat digunakan sebagai prediksi Summary yang tepat jika memiliki fitur-fitur input seperti yang telah dijelaskan.

Padahal, jika hasilnya cukup baik, hal ini dapat digunakan secara lebih luas agar pembaca awam dapat langsung mengetahui peringatan-peringatan yang penting terkait dengan prediksi cuaca hari itu. Jika ingin prediksi Summary ini semakin tepat, mungkin jumlah nilai-nilai uniknya dapat dirampingkan sehingga pilihannya hanya 10 saja, kemudian untuk sampel datanya bisa dapat dikumpulkan lebih banyak lagi. Namun, jika datanya terlalu banyak, maka dapat dipastikan waktu komputasinya akan memakan waktu yang lebih lama. Dibutuhkan perangkat yang lebih kuat jika ingin memproses data yang lebih banyak.

# **IV. KESIMPULAN**

Setelah meneliti kedua fitur yang akan diprediksi, Percipitation Type mendapatkan hasil yang cukup baik dikarenakan nilai-nilai yang perlu diperhatikan hanya sedikit, hanya terdapat 3 data saja yang perlu diprediksi. Dengan nilai akurasi yang mencapai 92%, dapat dipastikan bahwa prediksi Percipitation Type dapat diandalkan dengan kesalahan yang minimum. Sedangkan pada fitur Summary, kami tidak mendapatkan hasil yang baik dengan akurasi hanya mencapai 44.44%. Hal ini dapat dipungkiri dikarenakan banyaknya nilai yang harus diprediksi dengan jumlah data yang hanya terbatas, kurang dari 100000 baris data yang ada pada dataset ini, sehingga pengamatan data lebih lanjut diperlukan agar dapat memprediksi Summary dengan lebih tepat. Kemudian, dari sisi nilai-nilai yang ada di Summary, dapat dirampingkan berapa banyak data unik yang ada di Summary. Mungkin bisa dirampingkan menjadi hanya 10 atau kurang nilai yang unik sehingga dapat diprediksi dengan lebih baik dan dapat diandalkan. Prediksi cuaca adalah hal yang sangat penting bagi kelancaran perekonomian dan keamanan masyarakat secara umum. Jika terjadi kesalahan prediksi cuaca, apalagi salah memprediksi cuaca terburuk, maka bisa jadi pemerintah dan perusahaan kehilangan uang yang cukup banyak, dan untuk masyarakat umum dapat membahayakan nyawa. Prediksi Percipitation Type dan diikuti dengan prediksi Summary dapat membantu masyarakat umum dengan menyediakan informasi siap saji yang dapat langsung diterapkan dan diantisipasikan oleh pengguna. Prediksi Percipitation Type dapat memberikan masyarakat sebuah peringatan mengenai jenis hujan apa yang akan muncul jika memang terjadi hujan dan dengan itu masyarakat dapat bersiap sesuai dengan tipe hujan yang akan muncul. Jika hujan salju, mungkin masyarakat dapat mempersiapkannya dengan melakukan modifikasi ban kendaraan agar lebih bergerigi, memakai baju yang lebih tebal agar tidak hipotermia, dan memakai sepatu anti selip. Kemudian untuk hujan air, masyarakat dapat mempersiapkan jas hujan, mempersiapkan baju ganti, dan juga mempersiapkan kondisi kendaraan agar siap kelicinan jalan yang berair. Pada hujan es, masyarakat dapat bersiap dengan memasukkan mobil ke dalam garasi atau menyiapkan atap yang dapat melindungi kendaraan dari hantaman batu es, kemudian untuk masyarakat agar tetap tinggal di rumah sehingga tidak terkena jatuhan bebatuan es yang bergerak dengan kecepatan yang tinggi. Untuk prediksi Summary, masyarakat umum akan sangat terbantu jika memang terbukti efektif agar mereka dapat secara cepat mengambil kesimpulan dari data-data yang disediakan. Hampir kebanyakan masyarakat tidak dapat memperkirakan (walaupun secara kasar) kira-kira cuaca disekitar akan seperti apa. Jadi, dengan bantuan Summary ini, masyarakat dapat membaca singkat dan dapat menghubungkan data-data perkiraan cuaca yang disediakan dengan kondisi sekitar.

##### **REFERENSI**

1. Heart Disease Prediction System using Naive Bayes, Medhekar D, Bote M, Deshmukh S, 2013.
2. Weather Prediction: A novel approach for measuring and analyzing weather data, Navadia S, Thomas J, Yadav P, Shaikh S.
3. MR-VSM: Map Reduce based vector SpaceModel for user profiling-an empirical study on News data, Gautam A and Bedi P, 2015.
4. The Weather Forecast Using Data Mining Research Based on Cloud Computing, Wang, ZhanJie & Mujib, A B M, 2017.
5. Data Mining Concepts and Techniques, Han J, Kamber M, 2006.
6. An ensemble of neural networks for weather forecasting, Maqsood I, Khan R, and Abraham A, 2004.
7. Classification and Forecasting of Weather using ANN, k-NN and Naïve Bayes Algorithms, Barde N, Patole M, 2013, 2319-7064.
8. A Naïve Bayes Classifier Approach to Incorporate Weather to Predict Congestion at Intersections, Barua S, 2020, 72-76.
9. Using naïve bayes algorithm to students' bachelor academic performances analysis, Razaque F, Soomro N, Shaikh S et al., 4th IEEE International Conference on Engineering Technologies and Applied Sciences, ICETAS 2017, 2018, 1-5.
10. Naïve Bayes Classifier for Journal Quartile Classification, Wibawa A, Kurniawan A, Murti D et al., International Journal of Recent Contributions from Engineering, Science & IT (iJES), 2019, 91, 7(2).
11. Data Mining Application on Weather Prediction Using Classification Tree, Naïve Bayes and K-Nearest Neighbor Algorithm With Model Testing of Supervised Learning Probabilistic Brier Score, Confusion Matrix and ROC, Prasetya A, Ridwan A, 2019.
12. The naÏve bayes algorithm for learning data analytics, Viet T, Minh H, Hieu L et al., Indian Journal of Computer Science and Engineering, 2021, 1038-2043, 12(4).
13. Solar irradiance forecast using naïve bayes classifier based on publicly availableweather forecasting variables, Kwon Y, Kwasinski A, Kwasinski A, Energies, 2019, 12(8)
14. Analisis Sentimen Multi-Kelas Untuk Film Berbasis Teks Ulasan Menggunakan Model Regresi Logistik, Averina A, Hadi H, Siswantoro J, 2022.
15. Naive bayes classification of uncertain data, Ren J, et al., Ninth IEEE International Conference on in Data Mining, 2009.
16. Keoptimalan Naïve Bayes Dalam Klasifikasi, Shadiq A, no. 1, p. 31, 2009.
17. Prediksi Keputusan Menggunakan Metode Klasifikasi Naïve Bayes , One-R , Dan Decision Tree, Prediction D, Classification U, Naive M, and Tree D, pp. 1–10, 2016.
18. Classification and Forecasting of Weather using ANN, k-NN and Naïve Bayes Algorithms, Barde N and Patole M, Int. J. Sci. Res., vol. 5, no. 2, pp. 1740–1742, 2016.
19. Comparative Study of K-NN , Naive Bayes and Decision Tree Classification Techniques, Jadhav S and Channe H, vol. 5, no. 1, pp. 2014–2017, 2016.