**KLASIFIKASI DATA MINING UNTUK MENGESTIMASI POTENSI CURAH HUJAN MENGGUNAKAN ALGORITMA NAIVE BAYES**

**Nina Nur Salsabila1, Julian Rifaldi2, Nuzuliarini Nuris3, Sri Diantika4**

1,2,3,4Program Studi Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta

E-mail: 1ninanursalsa@gmail.com, 2julianrifaldi64@gmail.com, 3nuzuliarini.nzn@bsi.ac.id, 4sri.szd@bsi.ac.id

**Abstrak** - Salah satu hal yang sangat penting di Indonesia, adalah iklim. Iklim sendiri sangat memengaruhi kelangsungan hidup manusia. Iklim merupakan kondisi rata-rata cuaca selama periode waktu yang relatif lama. Sedangkan cuaca dapat berubah seiring berjalannya waktu. Cuaca dan iklim sangat erat kaitannya dengan pola curah hujan yang dapat berubah dari waktu ke waktu. Curah hujan yang tinggi dapat menyebabkan banjir, tanah longsor, dan merusak tanaman pertanian. Sebaliknya, curah hujan yang rendah dapat mengakibatkan kekeringan dan gagal panen. Oleh karena itu, pemahaman yang mendalam mengenai pola curah hujan sangat penting dalam perencanaan pembangunan dan pengelolaan sumber daya alam. Dengan adanya klasifikasi potensi curah hujan diharapkan dapat memberikan prediksi yang akurat dan membantu dalam perencanaan serta pengelolaan risiko terkait. Klasifikasi yang digunakan dalam penelitian ini yaitu *Naive Bayes* dengan data curah hujan bulanan pada Kota Jakarta Timur. Berdasarkan hasil perhitungan diperoleh tingkat akurasi sebesar 80% baik menggunakan perhitungan secara manual ataupun menggunakan aplikasi *RapidMiner*.

*Keywords* : curah hujan, klasifikasi, naive bayes, rapidminer

1. **PENDAHULUAN**

**1.1 Latar Belakang**

Salah satu hal yang sangat penting di Indonesia, adalah iklim. Pada beberapa industri, seperti pertanian, pariwisata, dan transportasi, iklim sendiri sangat memengaruhi kelangsungan hidup manusia. Iklim merupakan kondisi rata-rata cuaca selama periode waktu yang relatif lama. Sedangkan cuaca dapat berubah seiring berjalannya waktu. Salah satu faktor yang penting dalam perubahan cuaca yakni curah hujan.

Curah hujan yang tinggi dapat menyebabkan banjir, tanah longsor, dan merusak tanaman pertanian. Sebaliknya, curah hujan yang rendah dapat mengakibatkan kekeringan dan gagal panen. Oleh karena itu, pemahaman yang mendalam mengenai pola curah hujan sangat penting dalam perencanaan pembangunan dan pengelolaan sumber daya alam (Risnawati et al., 2023).

Klasifikasi potensi curah hujan merupakan salah satu metode yang penting dalam memprediksi hujan dengan menggunakan data historis. Salah satu algoritma yang sering digunakan untuk melakukan klasifikasi seperti ini adalah metode *Naive Bayes. Naive Bayes* adalah salah satu metode klasifikasi yang berbasis pada *Teorema Bayes* dengan asumsi bahwa semua fitur input saling independen (Rahmaulidyah et al., 2021). Metode *Naive Bayes* sering digunakan dalam klasifikasi potensi curah hujan karena kemampuannya yang cepat dan efisien. Selain itu, metode ini juga dapat mengatasi masalah *multicollinearity* antara variabel-variabel prediktor yang sering ditemui dalam data curah hujan.

Dengan adanya klasifikasi potensi curah hujan menggunakan metode Naive Bayes, diharapkan dapat memberikan prediksi yang akurat dan membantu dalam perencanaan serta pengelolaan risiko terkait dengan curah hujan. Selain itu, penggunaan metode ini juga dapat membantu dalam memahami pola hujan dan dampaknya terhadap berbagai sektor kehidupan. Dengan demikian, klasifikasi potensi curah hujan menggunakan metode Naive Bayes merupakan pendekatan yang penting dalam analisis data curah hujan. Melalui penerapan metode ini, diharapkan dapat meningkatkan pemahaman dan prediksi terhadap curah hujan untuk kepentingan berbagai sektor, mulai dari pertanian hingga bencana alam.

**1.2 Rumusan Masalah**

Bagaimana cara memahami mengenai pola curah hujan dalam perencanaan pembangunan dan pengelolaan sumber daya guna mengurangi resiko terkait menggunakan data mining khususnya pada wilayah Jakarta Timur.

**1.3 Tujuan Penelitian**

Adapun tujuan dari penelitian ini yaitu:

1. Untuk mengetahui kondisi secara umum curah hujan di Jakarta Timur menurut Seksi Meteorologi Halim Perdana Kusuma tahun 2019 – 2022.
2. Untuk mengetahui hasil klasifikasi curah hujan di Jakarta Timur dengan menggunakan metode *Naive Bayes*.
3. Untuk mengetahui potensi curah hujan di Jakarta Timur berdasarkan variabel tertentu
4. **METODELOGI PENELITIAN**

Penelitian ini dilakukan dengan menggunakan metode Klasifikasi. Untuk jenis klasifikiasi yang digunakan adalah *Naive Bayes*. Adapun langkah-langkah dalam penelitian ini menggunakan proses KDD (*Knowledge Discovery from Data*) yaitu (Kirana et al., 2024):

1. *Data Selection*,
2. *Preprocessing* atau *cleaning*,
3. *Transformation*,
4. *Data Mining*,
5. *Interpretation/Evaluation*,

**2.1 Tahapan Penelitian**

Adapun tahapan penelitian ini dapat digambarkan melalui kerangka kerja penelitian yang dapat dilihat pada Gambar II.1:

Analisis Masalah dan Studi Literatur

Pengumpulan Data

Pengolahan Data

Pembagian Data

*Data Training*

*Modeling Naive Bayes*

*Data Testing*

Evaluasi

Selesai

Sumber: Hasil Penelitian (2024)

Gambar II. 1. Bagan Tahapan Penelitian

1. Analisis Masalah dan Studi Literatur

Tahap ini merupakan langkah awal yang digunakan dalam menganalisis permasalahan yaitu bagaimana cara memahami mengenai pola curah hujan dalam perancangan pembangunan dan pengelolaan sumber daya guna mengurangi risiko terkait menggunakan data mining. Selanjutnya dilakukan analisa untuk mengetahui bagaimana cara menyelesaikan masalah tersebut dengan mempelajari teori yang berkaitan dengan penelitian yang mana dapat diperoleh dari berbagai macam literatur baik nasional maupun internasional.

1. Pengumpulan Data

Dalam tahap ini dilakukan untuk mendapatkan data-data yang berkaitan dengan penelitian yang sedang di lakukan. Pengumpulan data bisa diperoleh dari data primer maupun sekunder. Dalam penelitian menggunakan data sekunder sebagai berikut:

Tabel III.1 Dataset curah hujan bulanan kota Jakarta Timur

| **Tahun** | **Bulan** | **Jumlah Curah Hujan (Mm)** | **Jumlah Hari Hujan (Mm)** | **Kecepatan Angin** | **Kelembapan** | **Suhu** | **Tekanan Udara** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2022 | Januari | 208 | 18 | 30 | 78 | 31,9 | 12,4 |
| 2022 | Februari | 253,3 | 15 | 32 | 80 | 31,6 | 12,3 |
| 2022 | Maret | 186 | 7 | 43 | 80 | 32,7 | 11,9 |
| 2022 | April | 382,2 | 9 | 30 | 78 | 33,7 | 12,2 |
| 2022 | Mei | 329,1 | 1 | 28 | 79 | 33,1 | 11,9 |
| 2022 | Juni | 0 | 0 | 12,1 | 65 | 28,5 | 28 |
| 2022 | Juli | 0 | 0 | 11,9 | 65 | 27,7 | 21 |
| ... | ... | ... | ... | ... | ... | ... | ... |

Sumber: BPS Jakarta Timur

1. Pengolahan data

Pada tahapan pengolahan *data mining* pada penelitian ini terdapat beberapa proses yang dilakukan antara lain :

1. *Data Selection*, melakukan proses pemilihan data yang digunakan dalam dianalisis pada penelitian seperti variabel yang digunakan yaitu curah hujan, hari hujan, kelembaban udara, suhu, tekanan udara serta kecepatan angin.
2. *Pre-processing/cleaning*, melakukan proses pembersihan data apabila terdapat data yang tidak sesuai, *noise* dan data yang tidak konsisten serta menghilangkan duplikasi data.
3. *Transformation*, melakukan proses perubahan data ke dalam bentuk/format yang dibutuhkan pada algoritma *Naive Bayes.* Dalam melakukan *transformasi* menggunakan kriteria sebagai berikut:

Tabel III.2 Kriteria Curah Hujan Bulanan

|  |  |
| --- | --- |
| **Curah Hujan** | **Keterangan** |
| 0 – 100  | Rendah |
| 101 – 300  | Sedang |
| 301 – 500 | Tinggi |
| > 500 | Ekstrim |

Sumber: [www.bmkgsampali.net](http://www.bmkgsampali.net)

Sehingga didapatkan hasil dari pengolahan data sebagai berikut:

Tabel III.3 Dataset setelah pengolahan data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Curah Hujan**  | **Jumlah hari hujan (mm)** | **Kecepatan angin** | **Kelembapan** | **Suhu** | **Tekanan udara** |
| Sedang | 18 | 30 | 78 | 31,9 | 12,4 |
| Sedang | 15 | 32 | 80 | 31,6 | 12,3 |
| Sedang | 7 | 43 | 80 | 32,7 | 11,9 |
| Tinggi | 9 | 30 | 78 | 33,7 | 12,2 |
| Tinggi | 1 | 28 | 79 | 33,1 | 11,9 |
| Rendah | 0 | 12,1 | 65 | 28,5 | 28 |
| Rendah | 0 | 11,9 | 65 | 27,7 | 21 |
| Rendah | 3 | 11,3 | 66 | 28 | 23 |
| ... | ... | ... | ... | ... | ... |

 Sumber: Hasil Penelitian (2024)

1. Pembagian Data

Dalam proses pembagian data dibagi menjadi *data training* dan *data testing*. *Data training* dan *data testing* dibagi secara terpisah untuk memastikan akurasi model klasifikasi. *Data training* digunakan untuk membentuk model probabilitas, sedangkan *data testing* digunakan untuk menguji akurasi model tersebut (Rizki et al., 2023). Dalam pembagian data digunakan rasio 90:10. Di mana 90% untuk data *training* dan 10% untuk data *testing*.

1. *Modeling* *Naive Bayes*

Pembentukan model *Naive Bayes* menggunakan data latih. Model ini menggunakan aturan probabilitas untuk memprediksi kategori curah hujan berdasarkan atribut-atribut yang terkait. Perhitungan untuk *Naive bayes* sendiri terbagi menjadi tiga macam variasi (Ismail et al., 2020), yaitu:

* 1. *Gaussian Naive Bayes*: cocok untuk data dengan distribusi normal
	2. *Multinominal Naive Bayes*: ideal untuk data diskrit seperti hitungan kata dalam teks
	3. *Bernoulli Naive Bayes*: digunkan untuk perhitungan data biner (fitur yang hanya memiliki dua nilai, yaitu 0 atau 1)

Dalam penelitian ini menggunakan variasi perhitungan *Gaussian Naive Bayes* (*Lilkelihood)*

1. *Evaluasi*

Tahap akhir pada sistem ini adalah evaluasi, di mana evaluasi yang dilakukan terhadap data set menggunakan *confusion matrix* yang diharapkan dapat memberikan validasi dari algoritma *Naïve Bayes* yang digunakan. *Confusion matrik* adalah adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah (Normawati & Prayogi, 2021). Dalam evaluasi ini akan di cari *accurcy, precision, recall dan F1-score.*

**2.2 Metode Pengumpulan Data**

Data yang digunakan merupakan data sekunder. Yang mana pengumpulan data dilakukan secara observasi dari *website* Badan Pusat Statistik Kota Jakarta Timur yaitu <https://jaktimkota.bps.go.id> . Dalam penelitian ini menggunakan populasi curah hujan bulanan di seluruh Indonesia tahun 2019 sampai 2022, sedangkan untuk sampel menggunakan curah hujan bulanan pada Kota Jakarta Timur mulai bulan Januari 2019 sampai Desember 2022 pada Stasiun Meteorologi Halim Perdana Kusuma. Adapun data yang dikumpulkan diringkas seperti Tabel II.1:

Tabel II. 1. Rincian data yang dikumpulkan

|  |  |  |  |
| --- | --- | --- | --- |
| **Nama Data** | **Kelas Data** | **Jumlah Data** | **Variabel** |
| **Data Latih** | Rendah | 9 | 1. Jumlah hari hujan
2. Kecepatan Angin
3. Kelembapan
4. Suhu
5. Tekanan Udara
 |
| Sedang | 23 |
| Tinggi | 9 |
| Ekstrem | 2 |
| **Total data** | **43** |
| **Data Uji** | Rendah | 1 |
| Sedang | 2 |
| Tinggi | 1 |
| Ekstrem | 1 |
| **Total data** | **5** |

Sumber: Hasil Penelitian (2024)

**3. Metode Klasifikasi Data**

Penelitian ini menggunakan metode klasifikasi *Naive Bayes* yang kemudian akan diperoleh hasil klasifikasi curah hujan, persentase kesalahan serta ketepatan klasifikasi. Dalam penentuan klasifikasi *Naive Bayes* ada beberapa persamaan yang digunakan yaitu :

Persamaan dari *Teorema Bayes* :

$P\left(h \right| D)=\frac{P\left(D \right| h) . P(h)}{P(D)}$

Keterangan:

*h* : kelas data

*D* : Data yang belum memiliki kelas

*P*(*h*) : Probabilitas kelas

 (disebut *Prior Probability)*

*P*(*h|D*) : Probabilitas *h* terhadap kondisi *D*

 (disebut *Posterior Probability)*

*P*(*D|h*) : Probabilitas *D* berdasarkan kondisi pada

 kelas *h (*disebut *likekihood)*

*P(D)* : Probabilitas *D (*disebut *evidence)*

Rumus di atas juga dapat dituliskan secara sederhana sebagai berikut:

$Posterior= \frac{Prior x Likelihood}{Evidence}$

Untuk menghitung hasil menggunakan *Teorema Bayes* perlu beberapa tahapan-tahapan sebagai berikut:

* 1. Mencari Probabilitas Kelas:

$P\left(h\right)= \frac{n\_{h}}{n\_{s}} $

Keterangan:

*P*(*h*) : Probabilitas kelas h

$n\_{h}$ : Jumlah data yang memiliki kelas h

$n\_{s}$ : Jumlah data pada ruang sampel

* 1. Mencari Rata – Rata (*mean*) untuk setiap variabel dalam setiap kelas

$μ\_{h}= \frac{1}{n\_{h}}\sum\_{i=1}^{n\_{h}}D\_{i}$

* 1. Mencari Standar Deviasi untuk setiap variabel dalam setiap kelas

$σ\_{h}=\sqrt{\frac{1}{n\_{h}-1}\sum\_{i=1}^{n\_{h}}(D\_{i}-μ\_{h})^{2}} $

Untuk mempermudah dalam menghitung dapat digunakan varian dengan persamaan sebagai berikut:

$σ\_{h}^{2}=\frac{1}{n\_{h}-1}\sum\_{i=1}^{n\_{h}}(D\_{i}-μ\_{h})^{2}$

* 1. Menghitung *Likelihood* menggunakan *Distribusi Gaussian* :

$P\left(D | h\right)=\frac{1}{\sqrt{2πσ^{2}}} e^{-\frac{(D-μ)^{2}}{2σ^{2}}}$

* 1. Menghitung Probabilitas Posterior untuk setiap kelas

Dikarenakan nilai *evidence (P(D))* yang bersifat konstan untuk semua kelas, maka dalam implementasi praktis cukup dengan mengalikan nilai *prior* dengan *likelihood* untuk setiap kelas, dan memilih kelas dengan nilai tertinggi.

**III. HASIL DAN PEMBAHASAN**

Dalam menghitung klasifikasi potensi curah hujan kami menggunakan 2 (dua) cara yaitu menggunakan perhitungan secara manual dan menggunakan aplikasi RapidMiner.

**3.1 Perhitungan Manual**

Dalam perhitungan manual menggunakan *Algoritma Naive Bayes* melibatkan beberapa langkah seperti yang dijelaskan dalam metode klasifikasi data. Sehingga didapatkan hasil dari tiap langkah sebagai berikut:

Probabilitas Kelas

Tabel III. 1. Hasil perhitungan probabilitas kelas

|  |  |  |
| --- | --- | --- |
| **Kelas** | **Jumlah** | **Probabilitas** |
| Rendah | 9 | 0,2093 |
| Sedang | 23 | 0,5349 |
| Tinggi | 9 | 0,2093 |
| Ekstrem | 2 | 0,0465 |
| **Total** | **43** | **1** |

Sumber: Hasil Penelitian (2024)

Rata-Rata tiap kelas

Tabel III. 2. Hasil Perhitungan Mean pada setiap kelas

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Hari Hujan** | **Kecepatan Angin** | **Kelembapan** | **Suhu** | **Tekanan Udara** |
| **Rendah** | 6,33 | 11,91 | 70,33 | 28,77 | 15,92 |
| **Sedang** | 14,87 | 17,47 | 76,35 | 29,52 | 14,18 |
| **Tinggi** | 17,33 | 14,72 | 79,22 | 29,22 | 14,12 |
| **Ekstrem** | 25,50 | 9,45 | 85,00 | 27,10 | 10,50 |

 Sumber: Hasil Penelitian (2024)

Varian tiap kelas

Tabel III. 3. Hasil perhitungan varia*n* tiap kelas

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Hari Hujan** | **Kecepatan Angin** | **Kelembapan** | **Suhu** | **Tekanan Udara** |
| **Rendah** | 12,75 | 41,65 | 10,00 | 2,70 | 55,28444 |
| **Sedang** | 42,94 | 99,11 | 23,24 | 4,75 | 38,29605 |
| **Tinggi** | 59,50 | 65,88 | 13,19 | 5,95 | 33,98194 |
| **Ekstrem** | 4,50 | 0,41 | 8,00 | 0,18 | 0,32 |

Sumber: Hasil Penelitian (2024)

*Likelihood* menggunakan *Distribusi Gaussian* tiap data uji

Tabel III. 4. Hasil perhitungan *likelihood* untuk data uji ke-1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **P(A|H)** | **Hari hujan** | **Kecepatan Angin** | **Kelembapan** | **Suhu** | **Tekanan Udara** |
| **Rendah** | 0,02317482 | 0,061819031 | 0,030426136 | 0,196716 | 0,042495 |
| **Sedang** | 0,004637565 | 0,034259591 | 0,005181499 | 0,129282 | 0,03514 |
| **Tinggi** | 0,004141613 | 0,046268096 | 5,15032E-05 | 0,134634 | 0,03412 |
| **Ekstrem** | 7,88002E-33 | 0,00037918 | 1,95886E-12 | 0,345923 | 1,08E-75 |

Sumber: Hasil Penelitian (2024)

Tabel III. 5. Hasil perhitungan *likelihood* untuk data uji ke-2

| **P(A|H)** | **Hari hujan** | **Kecepatan Angin** | **Kelembapan** | **Suhu** | **Tekanan Udara** |
| --- | --- | --- | --- | --- | --- |
| **Rendah** | 0,000537063 | 0,014803734 | 0,06441213 | 0,228865 | 0,041127 |
| **Sedang** | 0,054316914 | 0,010191571 | 0,073503363 | 0,152459 | 0,054005 |
| **Tinggi** | 0,051526372 | 0,011772692 | 0,039074439 | 0,149833 | 0,056422 |
| **Ekstrem** | 0,000363047 | 3,26278E-39 | 7,32847E-05 | 0,032626 | 0,705237 |

Sumber: Hasil Penelitian (2024)

Tabel III. 6. Hasil perhitungan *likelihood* pada data uji ke-3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **P(A|H)** | **Hari hujan** | **Kecepatan Angin** | **Kelembapan** | **Suhu** | **Tekanan Udara** |
| **Rendah** | 0,011145549 | 0,061206069 | 0,002950482 | 0,23972 | 0,039902 |
| **Sedang** | 0,060349175 | 0,032435843 | 0,071135571 | 0,164119 | 0,052408 |
| **Tinggi** | 0,04710871 | 0,044245328 | 0,109623029 | 0,156573 | 0,054574 |
| **Ekstrem** | 7,80886E-08 | 0,032288188 | 0,014866286 | 0,004062 | 0,612721 |

Sumber: Hasil Penelitian (2024)

Tabel III. 7. Hasil perhitungan *likelihood* pada data uji ke-4

| **P(A|H)** | **Hari hujan** | **Kecepatan Angin** | **Kelembapan** | **Suhu** | **Tekanan Udara** |
| --- | --- | --- | --- | --- | --- |
| **Rendah** | 0,031715066 | 0,002763371 | 0,000139732 | 0,073927 | 0,049667 |
| **Sedang** | 0,055316026 | 0,022913673 | 0,041617758 | 0,130996 | 0,0633 |
| **Tinggi** | 0,040723269 | 0,012895237 | 0,081983758 | 0,113796 | 0,06718 |
| **Ekstrem** | 3,01884E-10 | 2,0004E-185 | 0,080366384 | 4,93E-22 | 4,05E-05 |

Sumber: Hasil Penelitian (2024)

Tabel III. 8. Hasil perhitungan *likelihood* pada data uji ke-5



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **P(A|H)** | **Hari hujan** | **Kecepatan Angin** | **Kelembapan** | **Suhu** | **Tekanan Udara** |
| **Rendah** | 1,2994E-07 | 0,055838731 | 1,17231E-07 | 0,145151 | 0,044957 |
| **Sedang** | 0,018428028 | 0,02789486 | 0,007202312 | 0,098955 | 0,058685 |
| **Tinggi** | 0,031560227 | 0,038336034 | 0,011095295 | 0,112024 | 0,061851 |
| **Ekstrem** | 0,182911105 | 0,488212544 | 0,109847822 | 0,940316 | 0,147826 |

Sumber: Hasil Penelitian (2024)

Probabilitas kelas setiap data uji

Tabel III. 9. Hasil perhitungan probabilitas posterior dari tiap data uji

| **Data uji** | **Kelas** | **Nilai Probabilitas Posterior** | **Nilai Terbesar** |
| --- | --- | --- | --- |
| **ke-1** | Rendah | 7,62663E-08 | 7,62663E-08 |
| Sedang | 2,00043E-09 |
| Tinggi | 9,4891E-12 |
| Ekstrem | 1,0192E-124 |
| **ke-2** | Rendah | 1,0089E-09 | 1,79198E-07 |
| Sedang | 1,79198E-07 |
| Tinggi | 4,19397E-08 |
| Ekstrem | 9,29015E-50 |
| **ke-3** | Rendah | 4,02963E-09 | 6,40615E-07 |
| Sedang | 6,40615E-07 |
| Tinggi | 4,08643E-07 |
| Ekstrem | 4,3395E-15 |
| **ke-4** | Rendah | 9,41116E-12 | 2,33962E-07 |
| Sedang | 2,33962E-07 |
| Tinggi | 6,88871E-08 |
| Ekstrem | 4,5041E-223 |
| **ke-5** | Rendah | 1,16175E-18 | 6,34199E-05 |
| Sedang | 1,15001E-08 |
| Tinggi | 1,94678E-08 |
| Ekstrem | 6,34199E-05 |

Sumber: Hasil Penelitian (2024)

Evaluasi hasil

Berdasarkan hasil perhitungan di atas diperoleh *confusion* *matrix* sebagai berikut:

Tabel III. 10. Confusion matrix data uji curah hujan

|  |  |  |
| --- | --- | --- |
| **Aktual**  | **Prediksi** | **Recall** |
| **Rendah** | **Sedang** | **Tinggi** | **Ekstrem** |
| **Rendah** | 1 | 0 | 0 | 0 | **1** |
| **Sedang** | 0 | 2 | 0 | 0 | **1** |
| **Tinggi** | 0 | 1 | 0 | 0 | **0** |
| **Ekstrem** | 0 | 0 | 0 | 1 | **1** |
| **Precision** | **1** | **0,67** | **0** | **1** |  |

Sumber: Hasil Penelitian (2024)

Dengan tingkat akurasi sebesar 80%, yang diperoleh dari $Accuracy=\frac{Jumlah prediksi benar}{Jumlah total sampel}$

**3.2 Menggunakan *Tools* RapidMiner**

Berikut langkah-langkah yang dilakukan dalam menghitung klasifikasi *Naive* *Bayes* menggunakan *tools* RapidMiner:

Upload data latih dan data uji yang telah disiapkan ke database RapidMiner

Lakukan pengolahan data dengan memastikan jenis data yang di input sesuai dengan jenis data dalam klasifikasi Naive Bayes, serta tetapkan label pada data yang menjadi kelas. Untuk parameter kecepatan angin, kelembapan udara, suhu, serta tekanan udara ubah tipe data dari polynominal menjadi real, sedangkan jumlah hari hujan diubah menjadi integer.Dapat dilihat pada Gambar III.3:

Sumber: Hasil Penelitian (2024)

Gambar III.3. Transformasi data curah hujan pada RapidMiner

Membuat modeling *Naive Bayes Clasification* sesuai dengan kebutuhan. Dalam modeling ini diperlukan beberapa operator yaitu: *operator Naive Bayes* sebagai jenis klasifikasi yang digunakan, operator *apply model* digunakan untuk menerapkan model yang dilatih sebelumnya menggunakan data latih dan yang terakhir *operator performance* yang digunakan untuk mengevaluasi kinerja model yang memberikan daftar nilai kinerja secara otomatis sesuai dengan tugas yang diberikan (contoh kriteria: *accuracy, precision,* dan *recall*). Dapat dilihat pada Gambar III.4:

Sumber: Hasil Penelitian (2024)

Gambar III. 1. Modeling Naive Bayes pada RapidMiner

Pengujian menggunakan RapidMiner pada data curah hujan menghasilkan beberapa hal, yaitu: a. Distribusi model (probabilitas kelas), b. Nilai *mean* dan *standard deviation* dari tiap-tiap parameter, c. Prediksi kelas tiap data uji. Untuk distribusi model dapat dilihat pada Gambar III.5, untuk nilai *mean* dan *standard deviation* dapat dilihat pada Gambar III.6 dan untuk prediksi dapat dilihat pada Gambar III.7:

Sumber: Hasil Penelitian (2024)

Gambar III. 5. Hasil Perhitungan Probabilitas tiap kelas pada RapidMiner

Sumber: Hasil Penelitian (2024)

Gambar III.6. Hasil perhitungan nilai mean dan standard deviation

Sumber: Hasil Penelitian (2024)

Gambar III. 7. Hasil prediksi curah hujan tiap data uji pada RapidMiner

Hasil evaluasi performance menggunakan *tools RapidMiner* menggunakan algoritma *Naive Bayes* diperoleh akurasi sebesar 80%. Dapat dilihat pada Gambar III.8:

Sumber: Hasil Penelitian (2024)

Gambar III.8. Hasil Evaluasi menggunakan RapidMiner

**VI. PENUTUP**

**4.1 Kesimpulan**

Setelah penelitian dilakukan maka dapat disimpulkan beberapa hal sebagai berikut:

1. Hasil penelitian curah hujan menggunakan perhitungan manual dan menggunakan aplikasi *RapidMiner* mendapatkan hasil posterior maksimal(terbesar) tiap data uji pada prediksi kelas yang sama.
2. Dalam penerapan naive bayes secara manual dan menggunakan aplikasi *RapidMiner* terdapat perbedaan pada jumlah nilai *posterior*/*confidence* yang mana pada perhitungan manual mendapatkan hasil berturut-turut: 7,62663E-08 (rendah), 2,98304E-09 (sedang), 1,65532E-09 (sedang), 6,58136E-07 (sedang), 6,34199E-05 (ekstrem), sedangkan pada *RapidMiner* mendapatkan hasil: 0,974 (rendah), 0,807 (sedang), 0,608 (sedang), 0,773 (sedang), 1,000 (ekstrem).
3. Tingkat akurasi perhitungan curah hujan menggunakan algoritma *Naive Bayes* sebesar 80% baik menggunakan perhitungan secara manual maupun menggunakan aplikasi *RapidMiner*.
4. Hasil yang masih rendah pada katagori curah hujan tinggi terletak pada data, seharusnya tidak ada perbedaan yang signifikan di antara keempat kelas. Perlu dilakukan pengujian ulang dengan data yang lebih banyak.

**4.2 Saran**

Berdasarkan hasil penelitian pembahasan dan analisa pada penelitian, maka peneliti memberikan saran untuk penelitian selanjutnya yang diharapkan bias dijadikan dasar untuk mengembangkan penelitian ini yaitu:

1. Perlu dilakukan pengujian ulang dengan metode yang sama dengan jumlah data yang lebih banyak
2. Perlu adanya perbandingan menggunakan metode lain atau gabungan dari beberapa metode yang dapat meningkatkan akurasi

**DAFTAR REFERENSI**

Ismail, M., Hassan, N., & Bafjaish, S. S. (2020). Comparative Analysis of Naive Bayesian Techniques in Health-Related for Classification Task. *Journal of Soft Computing and Data Mining*, *1*(2), 1–10. https://doi.org/10.30880/jscdm.2020.01.02.001

Kirana, A. N., Nurhakim, B., Permana, S. E., Prihartono, W., & Dwilestari, G. (2024). Implementasi Algoritma Naive Bayes Untuk Memprediksi Cuaca Menggunakan Rapidminer. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 1637–1642. https://doi.org/10.36040/jati.v8i2.8967

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, *5*(2), 697–711.

Rahmaulidyah, F. N., Hayati, M. N., & Goejantoro, R. (2021). Perbandingan Metode Klasifikasi Naive Bayes dan K-Nearest Neighbor pada Data Status Pembayaran Pajak Pertambahan Nilai di Kantor Pelayanan Pajak Pratama Samarinda Ulu. *Eksponensial*, *12*(2), 161. https://doi.org/10.30872/eksponensial.v12i2.809

Risnawati, I., Rahma, S., Kusuma, F., Salsabila, N., Nourmansyah, A., Farhan, A., & Irfiani, E. (2023). Klasifikasi Data Mining Untuk Mengestimasi Potensi Curah Hujan Berdampak Banjir Daerah Menggunakan Algoritma C4.5. *Jurnal INSAN Journal of Information System Management Innovation*, *3*(2), 78–84. https://doi.org/10.31294/jinsan.v3i2.3050

Rizki, F., Kharisma Putra, M. P., Assuja, M. A., & Ariany, F. (2023). Implementasi Deep Leraning Lenet Dengan Augmentasi Data Pada Identifikasi Anggrek. *Jurnal Informatika Dan Rekayasa Perangkat Lunak (JATIKA)*, *4*(3), 357–366. https://doi.org/10.33365/jatika.v4i3.3652