

Studi Komparasi Algoritma Klasifikasi C5.0, SVM dan Naive Bayes dengan Studi Kasus Prediksi Banjir

Comparative Study of Classification Algorithm between C5.0, SVM and Naive Bayes with Case Study of Flood Prediction

Devi Fitriana¹, Wawan Gunawan², Anggi Puspita Sari³

^{1,2} Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana

³ Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

E-mail: ¹devi.fitriana@mercubuana.ac.id, ²wawan.gunawan@mercubuana.ac.id,

³anggi.apr@bsi.ac.id

Abstrak

Indonesia merupakan negara tropis yang memiliki jumlah penduduk yang banyak sehingga mengakibatkan banyak sekali bencana alam yang harus diterima oleh Indonesia. Penelitian ini difokuskan pada bencana banjir yang nantinya dapat dimanfaatkan untuk mengatasi bencana kekeringan dengan cara penampungan air hujan. Selanjutnya berdasarkan luas wilayah dan jumlah penduduk yang ada, jika kita bandingkan dengan bencana banjir yang terjadi maka provinsi Jawa Barat yang seharusnya dapat perhatian lebih besar karena luas wilayah untuk masing-masing penduduk paling kecil jika dibandingkan dengan provinsi yang lain. Penelitian ini yang dilakukan menggunakan algoritma *Support Vector Machine* (SVM), C5.0 dan Naive Bayes yang digunakan untuk melakukan prediksi banjir untuk membantu pencegahan kebencanaan agar tidak terjadi korban yang lebih banyak. algoritma SVM dan C5.0 memiliki nilai akurasi yang sama yaitu sebesar 93.75% sedangkan algoritma Naive Bayes memiliki nilai akurasi sebesar 81,25. Sehingga dapat disimpulkan bahwa algoritma ini lebih akurat digunakan untuk melakukan prediksi. Sedangkan untuk waktu pemrosesannya maka algoritma Naive Bayes bisa dikatakan lebih cepat jika dibandingkan algoritma SVM dan juga algoritma C5.0.

Kata kunci: Kebencanaan, SVM, C5.0, Naive Bayes

Abstract

Indonesia is a tropical country that has a large population, resulting in many natural disasters that must be accepted by Indonesia. This research is focused on flood disasters which can later be used to overcome drought disasters by storing rainwater. Furthermore, based on the size of the area and the existing population, if we compare it with the flood disaster that occurred, the province of West Java should receive greater attention because the area for each population is the smallest when compared to other provinces. This research was conducted using the Support Vector Machine (SVM), C5.0 and Naive Bayes algorithms which were used to predict floods to help prevent disasters so that more victims would not occur. SVM and C5.0 algorithms have the same accuracy value of 93.75% while the Naive Bayes algorithm has an accuracy value of 81.25. So it can be concluded that this algorithm is more accurate to use for making predictions. As for the processing time, the Naive Bayes algorithm can be said to be faster than the SVM algorithm and the C5.0 algorithm.

Keywords: Disaster, SVM, C5.0, Naive Bayes

1. PENDAHULUAN

Indonesia merupakan negara tropis terbesar di dunia dengan memiliki daerah yang dilalui oleh garis khatulistiwa. Iklim tropis yang ada di Indonesia mengakibatkan banyak sekali bencana alam yang harus diterima oleh Indonesia. Karena banyaknya jumlah penduduk yang terdampak akibat bencana alam terdapat pada bencana kekeringan dan juga bencana banjir, maka penelitian ini akan difokuskan pada bencana banjir yang nantinya dapat dimanfaatkan untuk mengatasi bencana kekeringan dengan cara penampungan air hujan[1]. Selanjutnya berdasarkan luas wilayah dan jumlah penduduk yang ada, jika dibandingkan dengan bencana banjir yang terjadi.

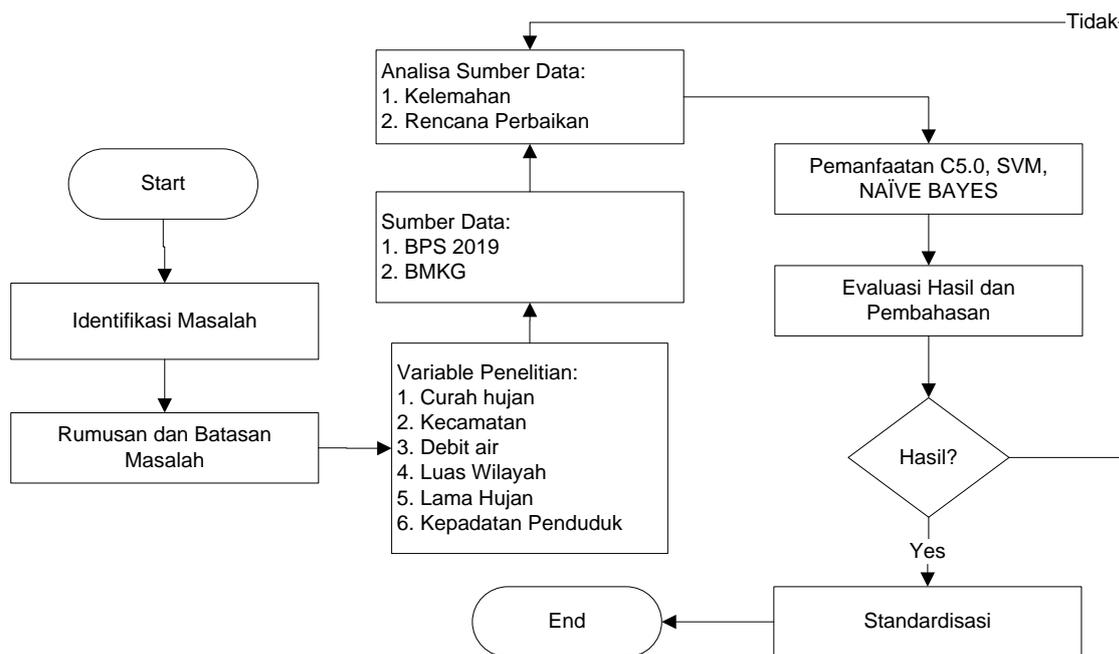
Penelitian yang dilakukan sebelumnya menggunakan software R dengan menggunakan kombinasi data training 60% dan data testing 40% dengan menggunakan algoritma SVM secara akurat 100% menyatakan banjir dalam 6 hari ke depan[2]. Sedangkan penggunaan algoritma C4.5 dapat memberikan nilai akurasi sebesar 84,385% untuk prediksi banjir di kabupaten Karawang [3]. Selanjutnya berdasarkan hasil perbandingan antara algoritma SVM dengan Naive Bayes pada bencana banjir di Indonesia didapatkan hasil akurasi penggunaan SVM sebesar 48,90 dan Naive Bayes sebesar 64,70% [4]. Pada penelitian lain yang dilakukan untuk bencana erupsi menggunakan algoritma Naive Bayes menyatakan bahwa nilai akurasi yang dihasilkan mencapai sebesar 89,59% [5].

Penelitian lain pun menjelaskan pada bencana banjir menggunakan algoritma C5.0 dan SVM dihasilkan bahwa nilai AUC pada penggunaan algoritma SVM lebih tinggi yaitu sebesar 0,921 daripada penggunaan algoritma C5.0 yang sebesar 0,919 [6]. Pada penelitian sebelumnya dihasilkan tingkat akurasi yang berbeda ketika menggunakan algoritma SVM, akurasi saat banjir 6 hari ke depan bernilai 100% namun pada kondisi banjir lain bernilai 48,9%. Selain itu pada penggunaan algoritma Naive Bayes pun memiliki tingkat akurasi yang berbeda yaitu sebesar 64,7% dan 89,59%. Penelitian ini membuktikan dan melakukan komparasi atas penggunaan algoritma SVM, C5.0 dan Naive Bayes dengan menggunakan data yang sama untuk mengetahui tingkat keakurasian dari ketiga algoritma tersebut.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode eksperimen. Metode eksperimen dibedakan menjadi dua, yaitu eksperimen murni (true eksperimen) dan eksperimen semu (quasi eksperimen). Pada penelitian ini peneliti menggunakan metode penelitian eksperimen semu (quasi eksperimen), karena peneliti tidak dapat melakukan kontrol atas variabel yang digunakan atau melakukan manipulasi atas seluruh variabel yang digunakan. Tahapan yang dilakukan pada penelitian ini dapat terlihat seperti pada gambar 1 di bawah ini.

Pada tahapan awal, peneliti melakukan identifikasi terlebih dahulu terhadap masalah yang ditemukan mengenai seringnya terjadi bencana banjir di Indonesia sehingga menimbulkan banyak korban jiwa. Setelah itu peneliti mencari tahu variabel yang dibutuhkan guna menjadi penyebab terjadinya bencana banjir, dimulai dari curah hujan, debit air, kecamatan yang terkena dampak bencana banjir, luas wilayah yang mempengaruhi letak geografis dan juga demografi, kepadatan penduduk pada suatu wilayah dan juga lamanya curah hujan yang turun pada wilayah tertentu. Selanjutnya setelah menentukan variabel, peneliti mencoba mencari data pendukung yang didapat berdasarkan data BPS di tahun 2019 dan juga data BMKG. Selanjutnya dilakukan olah data menggunakan algoritma SVM, C5.0 dan juga Naive Bayes untuk mengetahui akurasi dari masing-masing algoritma yang ditentukan.



Gambar 1. Tahapan penelitian

2.1. Support Vector Machines (SVM)

SVM adalah suatu teknik yang telah diterapkan sebagian besar dalam pengenalan pola di bidang jaringan karena teknik ini telah terbukti efisien jika dibandingkan teknik pengenalan pola lainnya [7]. SVM merupakan suatu model klasifikasi berdasarkan atas kecerdasan buatan yang telah mempengaruhi berbagai macam proses pengambilan keputusan di berbagai bidang termasuk pada bidang kedokteran. Algoritma SVM memiliki kinerja yang sesuai dan ini merupakan kriteria yang sangat penting dalam pengambilan keputusan. Algoritma ini jika dibandingkan dengan Logistic Regression memiliki hasil yang lebih baik saat dilakukan untuk mendiagnosis OSA pada pasien sehingga para dokter disarankan menggunakan alat ini sebagai alat skrining tambahan[8].

2.2. Naive Bayes

Teorema Bayes yang ditemukan oleh Thomas Bayes pada abad ke-18 merupakan pola dalam pembelajaran Bayesian yang sangat familiar dan memiliki sudut pandang terarah untuk menghasilkan hasil probabilitas yang diinginkan. Model Naïve Bayes Classifier sebagai metode klasifikasi untuk menunjukkan bahwa model tersebut memiliki akurasi, presisi, recall yang baik, skor f 1 yang baik dalam pengklasifikasian. Naïve Bayes Classifier adalah teknik prediksi yang didasarkan pada probabilitas sederhana dan pada penerapan teorema Bayes (atau aturan Bayes) dengan asumsi independensi yang kuat. Oleh karena itu, model ini mampu menghasilkan akurasi klasifikasi yang lebih tinggi dengan kompleksitas yang lebih sedikit. Secara khusus, model ini mencapai akurasi klasifikasi hingga 95,24%, sehingga model ini dapat menjadi alat analisis yang efisien [9]. Naive Bayes dapat memberikan nilai akurasi yang lebih baik dalam melakukan klasifikasi teks dengan menggunakan particle swarm optimization dengan menggunakan parameter [10].

2.3. C5.0

C5.0 adalah versi komersial dari C4.5 yang secara luas digunakan di banyak pemaketan data mining seperti Clementine and RuleQuest. Tidak seperti C4.5, penggunaan algoritma yang tepat untuk C5.0 belum terungkap. Hasil menunjukkan bahwa C5.0 meningkatkan pada penggunaan memori sekitar 90%, lebih cepat daripada C4.5 [11]. Algoritma C5.0 adalah salah satu algoritma klasifikasi data mining yang khususnya diterapkan pada teknik decision tree. C5.0 merupakan penyempurnaan dari algoritma sebelumnya yang dibentuk oleh Ross Quinlan pada

tahun 1987, yaitu ID3 dan C4.5. Dalam algoritma ini pemilihan atribut yang akan diproses menggunakan information gain. Dalam memilih atribut untuk pemecah obyek dalam beberapa kelas harus dipilih atribut yang menghasilkan information gain paling besar. Atribut dengan nilai information gain tertinggi akan dipilih sebagai parent bagi node selanjutnya.

2.4. Accuracy, Precision dan Recall

Setelah proses perhitungan dengan ke-3 algoritma, selanjutnya dilakukan perhitungan pada algoritma decision tree dengan menghitung Accuracy, precision dan Recall dengan membandingkan nilai prediksi dengan nilai sebenarnya.

		Nilai Sebenarnya	
		True	False
Nilai Prediksi	True	TP	FP
	False	FN	TN

Dimana

- TP : True Positive
- TN : True Negative
- FP : False Positive
- FN : False Negatif

- a. *Accuracy* merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. Akurasi ini menjawab pertanyaan mengenai bencana banjir

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- b. *Precision* merupakan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model atau rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. *Precision* menjawab pertanyaan “Berapa persen konsumen yang tidak klaim dari keseluruhan konsumen yang sebenarnya tidak klaim?”

$$precision = \frac{TP}{TP + FP} \tag{2}$$

- c. *Recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Recall menjawab pertanyaan “Berapa persen daerah banjir dari keseluruhan data yang ada?”

$$precision = \frac{TP}{TP + FN} \tag{3}$$

3. HASIL DAN PEMBAHASAN

A. Algoritma C5.0

Merupakan perhitungan accuracy data *training* menggunakan algoritma C5.0 yang menghasilkan nilai AUC nya 0.996. Diketahui data *training* terdiri dari 200 *record* data, selanjutnya dilakukan pengolahan data seperti yang ditampilkan pada gambar 2

```
1. === Run information ===
2.
3. Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
4. Relation:    Data Olah
```

```

5. Instances:      48
6. Attributes:    7
7.               Kecamatan
8.               Intensitas Curah Hujan
9.               Debit Air
10.              Luas Wilayah
11.              Lamanya Hujan
12.              Kepadatan Penduduk
13.              target
14. Test mode:    split 66.0% train, remainder test
15.
16. === Classifier model (full training set) ===
17.
18. J48 pruned tree
19. -----
20.
21. Intensitas Curah Hujan <= 393.4: Tidak Banjir (42.0/2.0)
22. Intensitas Curah Hujan > 393.4: Banjir (6.0/1.0)
23.
24. Number of Leaves :      2
25.
26. Size of the tree :      3
27.
28.
29. Time taken to build model: 0 seconds
30.
31. === Evaluation on test split ===
32.
33. Time taken to test model on test split: 0 seconds
34. === Detailed Accuracy By Class ===
35.
36.              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
37.              0,933   0,000   1,000     0,933   0,966     0,683   0,967    0,996    Tidak Banjir
38.              1,000   0,067   0,500     1,000   0,667     0,683   0,967    0,500    Banjir
39. Weighted Avg.   0,938   0,004   0,969     0,938   0,947     0,683   0,967    0,965

```

Gambar2. Pengolahan data menggunakan algoritma C5.0

Pada gambar 2 Klasifikasi dengan menggunakan tree J.48, total Instances : 48 dengan 7 atribut yaitu kecamatan, intensitas curah hujan, debit air, luas wilayah, lamanya hujan, kepadatan penduduk . Data di lakukan training dengan menggunakan metode : split dengan ketentuan 66% sehingga Menghasilkan pohon dengan ketentuan: 2 daun dan size pohon 3 yang dinyatakan dengan 42 instances yg tidak banjir karena intensitas curah hujan <= 393,4 dan 6 instances banjir.

Hasil pengujian data klaim produksi dilakukan pada gambar 3 menyebutkan bahwa nilai akurasi yang dihasilkan sebesar 93,75%

```

40.
41. === Summary ===
42.
43. Correctly Classified Instances      15          93.75  %
44. Incorrectly Classified Instances    1           6.25  %
45. Kappa statistic                     0.6364
46. Mean absolute error                 0.1696
47. Root mean squared error            0.2704
48. Relative absolute error            69.9134 %
49. Root relative squared error        96.1089 %
50. Total Number of Instances         16
51.

```

Gambar3. Pengujian data produksi menggunakan algoritma C5.0

Dari proses yang terjadi pada gambar 3, selanjutnya dilakukan perhitungan pada algoritma decision tree dengan menghitung Accuracy, precision dan Recall.

a. *Accuracy*

Perhitungan *accuracy*nya seperti pada persamaan 1 adalah sebagai berikut:

$$Accuracy = \left(\left(\frac{14+1}{14+1+1+0} \right) \times 100\% \right) = 93,75 \%$$

Jadi, hasil *Accuracy* dalam menghitung prediksi banjir hasilnya adalah 93,75%.

b. *Precision*

Perhitungan *Precision*nya seperti pada persamaan 2 adalah sebagai berikut:

$$Precision = \left(\left(\frac{14}{14+1} \right) \times 100\% \right) = 93,33 \%$$

Jadi, hasil *Precision* dalam menghitung hasil tidak banjir adalah 93,33%.

c. *Recall*

Perhitungan *Recall*nya seperti pada persamaan 3 adalah sebagai berikut:

$$Recall = \left(\left(\frac{14}{14+0} \right) \times 100\% \right) = 100 \%$$

Aplikasi Weka juga memiliki hasil confusion matrix seperti pada gambar 4

```
52.
53. === Confusion Matrix ===
54.
55.  a  b  <-- classified as
56. 14  1 | a = Tidak Banjir
57.  0  1 | b = Banjir
```

Gambar 4. Confusion Matrix menggunakan algoritma C5.0

B. Algoritma SVM

Merupakan perhitungan accuracy data *training* menggunakan algoritma SVM yang menghasilkan nilai AUC nya 0.938. Diketahui data *training* terdiri dari 200 *record* data seperti tampak pada gambar 5

```
1. === Run information ===
2.
3. Scheme:      weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -
  1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 2.0 -C 250007" -
  calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-
  places 4"
4. Relation:    Data Olahan
5. Instances:   48
6. Attributes:  7
7.             Kecamatan
8.             Intensitas Curah Hujan
9.             Debit Air
10.            Luas Wilayah
11.            Lamanya Hujan
12.            Kepadatan Penduduk
13.            target
14. Test mode:  split 66.0% train, remainder test
15.
16. === Classifier model (full training set) ===
17.
```

```

18. SMO
19.
20. Kernel used:
21. Poly Kernel:  $K(x,y) = \langle x,y \rangle^2.0$ 
22.
23. Classifier for classes: Tidak Banjir, Banjir
24.
25. BinarySMO
26.
27. - 0.6772 * <1 0 0 0 0.310936 0.586789 1 0.222222 0 > * X]
28. - 0.0545 * <0 0 0 1 0.70193 0.008643 0 0.111111 0.154357 > * X]
29. - 1 * <0 1 0 0 0.33238 0.00774 0.38966 0.333333 0.291075 > * X]
30. + 1 * <0 0 0 1 0.790386 0.040012 0 0.111111 0.154357 > * X]
31. - 0.8552 * <0 0 0 1 0.703002 0.004598 0 0.111111 0.154357 > * X]
32. - 0.0222 * <0 0 0 1 0.584703 0.018595 0 0.555556 0.154357 > * X]
33. + 0.6388 * <1 0 0 0 0.805933 0.612978 1 0.444444 0 > * X]
34. - 1 * <0 0 1 0 0.514653 0.015044 0.501034 0.222222 1 > * X]
35. - 1 * <0 0 1 0 0.53431 0.010563 0.501034 0.111111 1 > * X]
36. + 1 * <0 0 1 0 0.335954 0.004336 0.501034 0.111111 1 > * X]
37. - 0.0071 * <1 0 0 0 0.35025 1 1 0 0 > * X]
38. - 0.0275 * <0 0 1 0 0.40386 0.015015 0.501034 0.111111 1 > * X]
39. + 1 * <0 0 0 1 0.791637 0.040594 0 0.111111 0.154357 > * X]
40. - 0.9953 * <0 1 0 0 0.31451 0.006431 0.38966 0.222222 0.291075 > * X]
41. + 1 * <0 1 0 0 0.353824 0.007624 0.38966 0 0.291075 > * X]
42. - 1 * <0 0 0 1 1 0.041758 0 0.111111 0.154357 > * X]
43. + 1 * <0 0 1 0 0.897069 0.007071 0.501034 0.111111 1 > * X]
44. + 1 * <0 1 0 0 0.875625 0.006664 0.38966 0.444444 0.291075 > * X]
45. - 1.6815
46.
47. Number of support vectors: 18
48.
49. Number of kernel evaluations: 892 (82.117% cached)
50.
51.
52.
53. Time taken to build model: 0.13 seconds
54.
55. === Evaluation on test split ===
56.
57. Time taken to test model on test split: 0.08 seconds
58. === Detailed Accuracy By Class ===
59.
60.
61. TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
62. 1,000 1,000 0,938 1,000 0,968 ? 0,500 0,938 Tidak Banjir
63. 0,000 0,000 ? 0,000 ? ? 0,500 0,063 Banjir
63. Weighted Avg. 0,938 0,938 ? 0,938 ? ? 0,500 0,883

```

Gambar 5. Pengolahan data menggunakan algoritma SVM

Hasil pengujian data klaim produksi dilakukan pada gambar 6

```

64.
65. === Summary ===
66.
67. Correctly Classified Instances 15 93.75 %
68. Incorrectly Classified Instances 1 6.25 %
69. Kappa statistic 0
70. Mean absolute error 0.0625
71. Root mean squared error 0.25
72. Relative absolute error 25.7576 %
73. Root relative squared error 88.8604 %

```

```
74. Total Number of Instances      16
75.
```

Gambar 6. Pengujian data produksi menggunakan algoritma SVM

Aplikasi Weka juga memiliki hasil confusion matrix seperti pada gambar 7

```
76.
77. === Confusion Matrix ===
78.
79.  a  b  <-- classified as
80. 15  0 | a = Tidak Banjir
81.  1  0 | b = Banjir
```

Gambar 7. Confusion Matrix menggunakan algoritma SVM

Selanjutnya dilakukan perhitungan pada algoritma decision tree dengan menghitung Accuracy, precision dan Recall.

d. *Accuracy*

Perhitungan *accuracy* seperti pada persamaan 1 adalah sebagai berikut:

$$Accuracy = \left(\left(\frac{15+0}{15+0+0+1} \right) \times 100\% \right) = 93,75 \%$$

Jadi, hasil *Accuracy* dalam menghitung prediksi banjir hasilnya adalah 93,75%.

e. *Precision*

Perhitungan *Precisionnya* seperti pada persamaan 2 adalah sebagai berikut:

$$Precision = \left(\left(\frac{15}{15+0} \right) \times 100\% \right) = 100 \%$$

Jadi, hasil *Precision* dalam menghitung hasil tidak banjir adalah 100%.

f. *Recall*

Perhitungan *Recall* seperti pada persamaan 3 adalah sebagai berikut:

$$Recall = \left(\left(\frac{15}{15+1} \right) \times 100\% \right) = 93,75 \%$$

C. *Algoritma Naive Bayes*

```
1.  === Run information ===
2.
3.  Scheme:      weka.classifiers.bayes.NaiveBayes
4.  Relation:    Data Olahan
5.  Instances:   48
6.  Attributes:  7
7.              Kecamatan
8.              Intensitas Curah Hujan
9.              Debit Air
10.             Luas Wilayah
11.             Lamanya Hujan
12.             Kepadatan Penduduk
13.             target
14. Test mode:   split 66.0% train, remainder test
15.
16. === Classifier model (full training set) ===
17.
18. Naive Bayes Classifier
19.
20.              Class
```

21. Attribute	Tidak Banjir	Banjir							
22.	(0.84)	(0.16)							
23. =====									
24. Kecamatan									
25. Banjaran	12.0	2.0							
26. Bojongsoang	11.0	3.0							
27. Cileuyi	11.0	3.0							
28. Ciwidey	11.0	3.0							
29. [total]	45.0	11.0							
30.									
31. Intensitas Curah Hujan									
32. mean	147.1855	384.59							
33. std. dev.	123.3249	125.4647							
34. weight sum	41	7							
35. precision	15.1243	15.1243							
36.									
37. Debit Air									
38. mean	52.8412	36.286							
39. std. dev.	89.3343	70.8085							
40. weight sum	41	7							
41. precision	7.4707	7.4707							
42.									
43. Luas Wilayah									
44. mean	28.6211	25.7905							
45. std. dev.	12.7309	11.6236							
46. weight sum	41	7							
47. precision	11.2833	11.2833							
48.									
49. Lamanya Hujan									
50. mean	2.9268	2.3571							
51. std. dev.	1.8431	1.3553							
52. weight sum	41	7							
53. precision	1.5	1.5							
54.									
55. Kepadatan Penduduk									
56. mean	99288.6829	105736							
57. std. dev.	33722.8171	32374.9059							
58. weight sum	41	7							
59. precision	30839.6667	30839.6667							
60.									
61.									
62.									
63. Time taken to build model: 0 seconds									
64.									
65. === Evaluation on test split ===									
66.									
67. Time taken to test model on test split: 0.01 seconds									
68. === Detailed Accuracy By Class ===									
69.									
70.	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
71.	0,800	0,000	1,000	0,800	0,889	0,447	0,867	0,991	Tidak Banjir
72.	1,000	0,200	0,250	1,000	0,400	0,447	0,867	0,333	Banjir
73. Weighted Avg.	0,813	0,013	0,953	0,813	0,858	0,447	0,867	0,950	

Gambar 8. Pengolahan data menggunakan algoritma Naïve Bayes

Hasil pengujian data klaim produksi dilakukan pada gambar 9

74.									
75. === Summary ===									
76.									
77. Correctly Classified Instances				13			81.25	%	

78. Incorrectly Classified Instances	3	18.75 %
79. Kappa statistic	0.3333	
80. Mean absolute error	0.2238	
81. Root mean squared error	0.3646	
82. Relative absolute error	92.2249 %	
83. Root relative squared error	129.61 %	
84. Total Number of Instances	16	
85.		

Gambar 9. Pengujian klaim produksi menggunakan algoritma Naïve Bayes

Aplikasi Weka juga memiliki hasil confusion matrix seperti tabel di bawah ini:

```

86.
87. === Confusion Matrix ===
88.
89.  a  b  <-- classified as
90. 12  3 | a = Tidak Banjir
91.  0  1 | b = Banjir
    
```

Gambar 10. Confusion matrix menggunakan algoritma Naïve Bayes

Selanjutnya dilakukan perhitungan pada algoritma decision tree dengan menghitung Accuracy, precision dan Recall.

a. *Accuracy*

Perhitungan *accuracy* seperti pada persamaan 1 adalah sebagai berikut:

$$Accuracy = \left(\left(\frac{12+1}{12+1+3+0} \right) \times 100\% \right) = 81,25 \%$$

Jadi, hasil *Accuracy* dalam menghitung prediksi banjir hasilnya adalah 81,25%.

b. *Precision*

Perhitungan *Precision* seperti pada persamaan 2 adalah sebagai berikut:

$$Precision = \left(\left(\frac{12}{12+3} \right) \times 100\% \right) = 80 \%$$

Jadi, hasil *Precision* dalam menghitung hasil tidak banjir adalah 80%.

c. *Recall*

Perhitungan *Recall* seperti pada persamaan 3 adalah sebagai berikut:

$$Recall = \left(\left(\frac{12}{12+0} \right) \times 100\% \right) = 100 \%$$

4. KESIMPULAN DAN SARAN

Berdasarkan perhitungan yang telah dilakukan, maka algoritma SVM dan C5.0 memiliki nilai akurasi yang sama yaitu sebesar 93.75% sedangkan algoritma Naive Bayes memiliki nilai akurasi sebesar 81,25. Sehingga dapat disimpulkan bahwa algoritma ini lebih akurat dan efisien untuk digunakan untuk melakukan prediksi. Sedangkan untuk waktu pemrosesannya maka algoritma Naive Bayes bisa dikatakan lebih cepat jika dibandingkan algoritma SVM dan juga algoritma C5.0.

DAFTAR PUSTAKA

- [1] B. Harsoyo, "Teknik Pemanenan Air Hujan (Rain Water Harvesting) Sebagai Alternatif Upaya Penyelamatan Sumberdaya Air Di Wilayah Dki Jakarta," *J. Sains Teknol. Modif. Cuaca*, vol. 11, no. 2, p. 29, 2010, doi: 10.29122/jstmc.v11i2.2183.
- [2] I. Fitriyaningsih and Y. Basani, "Prediksi Kejadian Banjir dengan Ensemble Machine Learning Menggunakan BP-NN dan SVM," *J. Teknol. dan Sist. Komput.*, vol. 7, no. 3, pp. 93-97, 2019, doi: 10.14710/jtsiskom.7.3.2019.93-97.
- [3] A. Khusaeri *et al.*, "Algoritma C4.5 Untuk Pemodelan Daerah Rawan Banjir Studi Kasus

- Kabupaten Karawang Jawa Barat,” *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 132–136, 2017, doi: 10.33096/ilkom.v9i2.128.132-136.
- [4] R. K. Abdullah and E. Utami, “Studi Komparasi Metode SVM dan Naive Bayes pada Data Bencana Banjir di Indonesia pembaca ataupun peneliti bisa melihat pola yang tersembunyi di Indonesia,” *TECNOSCIENZA*, vol. 3, no. 1, pp. 103–122, 2018.
- [5] T. W. Bank, D. Mining, N. Bayes, and R. Rekon, “Penerapan Algoritma Klasifikasi Naive Bayes Untuk Data Status Huni Rumah Bantuan Dana Rehabilitasi Dan Rekonstruksi Pasca Bencana Erupsi Gunung Merapi 2010 Implementation of Naive Bayes Classification Algorithm for Occupancy House Data Status Fund Assista,” *Semin. Nas. UNRIYO*, pp. 1–10, 2019.
- [6] S. Dwiasnati and Y. Devianto, “Classification of Flood Disaster Predictions using the C5.0 and SVM Algorithms based on Flood Disaster Prone Areas,” *Int. J. Comput. Trends Technol.*, vol. 67, no. 07, pp. 49–53, 2019, doi: 10.14445/22312803/ijctt-v67i7p107.
- [7] A. Pradhan, “SUPPORT VECTOR MACHINE-A Survey,” no. April 2015, 2017.
- [8] M. A. Hammad, S. A. S. Sulaiman, N. A. Aziz, and D. A. M. Noor, “Comparison of support vector machine based on genetic algorithm with logistic regression to diagnose obstructive sleep apnea,” *J. Res. Med. Sci.*, vol. 24, no. 1, pp. 1–7, 2019, doi: 10.4103/jrms.JRMS.
- [9] N. Salmi and Z. Rustam, “Naïve Bayes Classifier Models for Predicting the Colon Cancer,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 546, no. 5, 2019.
- [10] K. S. Nugroho, I. Istiadi, and F. Marisa, “Naive Bayes classifier optimization for text classification on e-government using particle swarm optimization,” *J. Teknol. dan Sist. Komput.*, vol. 8, no. 1, pp. 21–26, 2020, doi: 10.14710/jtsiskom.8.1.2020.21-26.
- [11] M. H. Dunham, *Data Mining Introductory and Advanced Topics*. Prentice Hall: Pearson Education, 2013.