

ALGORITHM COMPARISON AND FEATURE SELECTION FOR CLASSIFICATION OF BROILER CHICKEN HARVEST

Christian Cahyaningtyas^{*1}, Danny Manongga², Irwan Sembiring³

^{1,2,3}Magister Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Indonesia
Email: ¹972020017@student.uksw.edu, ²danny.manongga@uksw.edu, ³irwan@uksw.edu

(Naskah masuk: 23 Juli 2022, Revisi : 8 September 2022, diterbitkan: 26 Desember 2022)

Abstract

Broiler chickens are the result of superior breeds that produce a lot of meat. In practice, however, many breeders experience crop failure, which has a serious impact on the economy and can also affect farmer quality, resulting in sanctions. The value of the performance index produced at harvest indicates the success rate of harvesting broiler chickens. Broiler crop yield data can be used to help classify broiler crop yield data using an approach method. The CRISP-DM (Cross Industry Standard Process for Data Mining) method was used in this study's data mining technique. This study compares 3 classification algorithms to determine the best algorithm and 3 feature selection methods to determine the best method for improving algorithm performance. According to the findings of this study, the Random Forest algorithm is the best algorithm for classifying harvest data, with an accuracy rate of 89.14 percent. The best way to improve the algorithm's performance is to use the Backward Elimination method, which can increase the accuracy by 7.53 percent. As a result, the Random Forest + Backward Elimination algorithm yields an accuracy value of 96.67 percent. According to this study, the factors that influence crop yield increase are FCR, number of harvests, and body weight.

Keywords: Algorithm Classification, Broiler Chicken, CRISP-DM, Feature Selection, Rapid Miner.

ANALISIS PERBANDINGAN ALGORITMA DAN FEATURE SELECTION UNTUK KLASIFIKASI HASIL PANEN AYAM BROILER

Abstrak

Ayam broiler merupakan hasil ras unggulan yang mempunyai produktivitas yang tinggi dalam menghasilkan daging. Namun pada prakteknya tidak sedikit peternak yang mengalami gagal panen yang memberikan dampak cukup serius pada perekonomian dan hal ini juga dapat mempengaruhi kualitas peternak yang mengakibatkan peternak mendapatkan sanksi. Tingkat keberhasilan panen ayam broiler dapat dilihat dari nilai indek performa yang dihasilkan saat panen. *Data Mining* dapat dijadikan sebagai metode pendekatan untuk membantu mengklasifikasikan data hasil panen ayam broiler. Teknik *Data Mining* yang digunakan dalam penelitian ini dengan menggunakan metode CRISP-DM (*Cross Industry Standard Process For Data Mining*). Penelitian ini membandingkan 3 algoritma klasifikasi untuk mendapatkan algoritma terbaik dan membandingkan 3 *Feature Selection* untuk mendapatkan metode yang terbaik dalam meningkatkan performa algoritma. Hasil yang didapat dari penelitian ini adalah algoritma *Random Forest* menjadi algoritma yang terbaik dalam mengklasifikasi data hasil panen, dengan nilai akurasi sebesar 89,14%. Sedangkan metode yang terbaik untuk meningkatkan performa algoritma adalah dengan menggunakan metode *Backward Elimination* yang dapat meningkatkan akurasi sebesar 7,53%. Sehingga nilai akurasi yang didapat dari algoritma *Random Forest + Backward Elimination* menjadi 96,67%. Serta dari penelitian ini dapat diketahui atribut yang berpengaruh dalam peningkatan hasil panen yaitu FCR (*Feed Conversion Ratio*), jumlah panen dan berat badan.

Kata kunci: Algoritma Klasifikasi, Ayam Broiler, CRISP-DM, Feature Selection, Rapid Miner.

1. PENDAHULUAN

Ayam broiler atau ayam ras pedaging merupakan jenis ras unggulan yang dihasilkan dari persilangan ayam yang mempunyai produktivitas yang tinggi dalam menghasilkan daging. Ayam broiler dalam proses pertumbuhan atau produksi

daging ini relatif singkat, hanya dalam waktu 4-5 minggu ayam broiler sudah dapat dipanen dan dikonsumsi [1]. Pertumbuhan yang cepat dengan berat badan yang tinggi, konversi pakan yang rendah, siap panen di usia yang muda dan menghasilkan daging berserat dan lunak merupakan keunggulan

dari ayam broiler. Namun dalam prakteknya terdapat beberapa peternak yang mengalami kerugian atau gagal panen secara berulang dan berturut-turut. Hal ini sangat berdampak pada perekonomian para peternak dan mendapatkan sanksi dari perusahaan bagi peternak yang bermitra.

Tingkat keberhasilan produksi ayam broiler ini dapat dilihat dari hasil performa produksi atau sering disebut indeks performa (IP). Faktor yang mempengaruhi hasil indeks performa pada ayam broiler ini berdasarkan *Feed Conversion Ratio* (FCR), jumlah panen, berat badan dan umur panen. Tingkat kematian sekitar 5% dalam peternakan ayam broiler ini dianggap berhasil. Dengan kata lain jumlah panen yang dihasilkan sekitar 95% dapat dikatakan hasil panen bagus. Bobot badan saat panen ini menentukan harga jual ayam. Total konsumsi pakan harus seminimal mungkin untuk mendapatkan bobot badan tertentu. Semakin efisien ransum, maka performa usaha peternakan ayam menjadi lebih baik [2]. Indeks performa dapat dijadikan sebuah evaluasi bagi para peternak untuk memperoleh hasil panen yang lebih baik dan tentunya menghasilkan indeks performa yang lebih tinggi di periode berikutnya. *Data Mining* dapat dijadikan sebagai metode pendekatan untuk membantu mengklasifikasikan data hasil panen ayam broiler.

Data Mining merupakan teknik untuk menggali data guna mencari pola agar menghasilkan informasi yang berguna dari sebuah data yang besar, sehingga dapat digunakan pada dunia nyata [3]. Teknik dalam *Data Mining* terdapat dua metode yaitu metode prediktif dan metode deskriptif. Metode prediktif dilakukan dengan model klasifikasi. Klasifikasi adalah proses dimana mengubah suatu data kedalam sekumpulan kelas atau kategori yang sama [4].

Penelitian tentang *Data Mining* sudah pernah dilakukan oleh Sri Lestari dan Herta Apriani Silaban pada tahun 2018. Penelitian tersebut menggunakan 3 algoritma untuk perbandingan yaitu *Decision Tree*, *Naïve Bayes* dan *Logistic Regression*. Metode yang digunakan untuk meningkatkan performa pada penelitian tersebut menggunakan metode *Feature Selection Filter* yaitu *Information Gain* dan *Wrapper* yaitu *Backward Elimination*. Penelitian tersebut mengenai keputusan tentang penerbitan SPTNP. Hasil dari penelitian yang dilakukan, algoritma *Logistic Regression* merupakan algoritma terbaik dan metode *Feature Selection* yang paling baik untuk meningkatkan performa algoritma *Logistic Regression* adalah metode *Backward Elimination*. Serta faktor yang sangat berpengaruh terhadap penetapan SPTNP adalah faktor dari profil perusahaannya [5].

Penelitian mengenai *Data Mining* juga pernah dilakukan pada tahun 2021 yang dilakukan oleh Ahmad Fauzi. Penelitian tersebut menggunakan algoritma *Naïve Bayes*, *K-Nearest Neighbor* dan *Support Vector Machine* (SVM) dan menggunakan metode *Optimize Selection* untuk meningkatkan

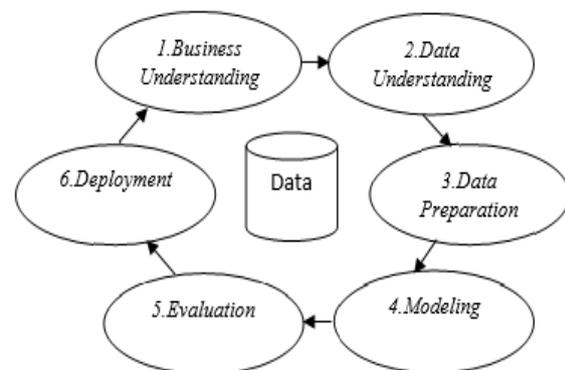
performa algoritma. Penelitian dilakukan dengan membandingkan algoritma tanpa metode optimasi dengan menggunakan metode optimasi. Hasil dari penelitian dapat disimpulkan bahwa metode optimasi *Optimize Selection* mampu meningkatkan performa ketiga algoritma. Peningkatan performa dapat dilihat dari tingkat akurasi yang dihasilkan. Algoritma dengan nilai akurasi tertinggi dihasilkan oleh algoritma *Naïve Bayes* dengan akurasi 90.18%. Serta algoritma dengan peningkatan nilai akurasi tertinggi setelah menggunakan metode *Optimize Selection* adalah algoritma *Support Vector Machine* dengan persentase kenaikan sebesar 8.47% [6].

Berdasarkan penelitian yang sudah dilakukan dan uraian yang sudah dijelaskan diatas maka dilakukan penelitian dengan menggunakan 3 algoritma klasifikasi yaitu *Naive Bayes*, *Random Forest* dan *K-Nearest Neighbor*. Serta untuk memperbaiki atau meningkatkan performa algoritma dengan menggunakan metode *Feature Selection* yaitu *Forward Selection*, *Backward Elimination* dan *Optimize Selection*. Tujuan dari penelitian ini adalah untuk mengklasifikasikan hasil panen ayam broiler dan mengetahui algoritma yang terbaik dan mengetahui metode terbaik untuk meningkatkan performa algoritma.

2. METODE PENELITIAN

2.1. Desain Penelitian

Metode yang digunakan dalam penelitian ini adalah dengan menggunakan metode CRISP-DM (*Cross Industry Standard Process for Data Mining*). Dengan menggunakan metode CRISP-DM ini peneliti mendapatkan strategi untuk memecahkan masalah dari sebuah industri dan sebagai sumber pengetahuan bagi peneliti [7]. Terdapat 6 fase yang ada pada metode CRISP-DM yaitu [8]:



Gambar 1. Tahapan Metode CRISP-DM

2.1.1. Business Understanding

Tahapan awal pada penelitian ini adalah dengan melakukan studi literatur dan wawancara mengenai ternak ayam broiler. Ayam broiler ini mempunyai produktivitas yang tinggi dalam menghasilkan daging

dalam waktu yang relatif singkat. Namun pada prakteknya peternak sering kali mengalami gagal panen secara berturut-turut. Tingkat keberhasilan panen dapat dilihat dari nilai indeks performa yang didapat. Data indeks performa akan dijadikan sebagai dataset untuk mengetahui pola tingkat keberhasilan panen dengan menggunakan teknik *Data Mining*:

2.1.2. Data Understanding

Pada penelitian ini dataset yang digunakan didapat dari hasil panen ayam PT. Cemerlang Unggas Lestari. Dataset yang dikumpulkan yaitu data dari tahun 2016-2021. Data yang didapat berupa data dalam bentuk *hardcopy* dan selanjutnya data diolah dan di-input kedalam *Microsoft Excel*. Data yang diinput sebanyak 451 data panen ayam.

2.1.3. Data Preparation

Data Preparation merupakan tahapan membuat dataset agar dapat dimodelkan menggunakan teknik *Data Mining*. Dataset yang dibuat berasal dari data mentah yang didapat sebelumnya. Hal yang diperlukan pada tahap ini adalah dengan melakukan *Data Selection* dan *Data Transformation*. Dimana *Data Selection* dilakukan untuk memilih atribut yang penting dan yang akan digunakan dalam proses pemodelan. Sedangkan *Data Transformation* dilakukan untuk mengubah sekumpulan data sesuai dengan proses *Data Mining*.

2.1.4. Modeling

Modeling merupakan tahapan memodelkan data. Pada penelitian ini peneliti menggunakan teknik *Data Mining*, yaitu dengan klasifikasi, pada klasifikasi membandingkan 3 algoritma klasifikasi yaitu algoritma *Naïve Bayes*, *Random Forest* dan *K-Nearest Neighbor*. Sehingga akan diketahui algoritma terbaik dengan melihat dari nilai akurasi yang didapat. Serta untuk memperbaiki performa algoritma menggunakan metode *Feature Selection* yaitu *Forward Selection*, *Backward Elimination* dan *Optimize Selection*. Dimana metode yang mampu meningkatkan performa tertinggi merupakan metode *Feature Selection* yang terbaik.

2.1.5. Evaluation

Evaluation merupakan tahapan evaluasi model yang sudah dilakukan. Pada tahap ini dilakukan validasi dan mengukur tingkat akurasi hasil yang sudah dicapai dengan model yang sudah dibuat. Untuk membandingkan performa 3 algoritma tersebut, maka menggunakan uji beda (*T-Test*). Dengan menggunakan *T-Test* dapat mengetahui perbedaan yang signifikan dari 3 algoritma tersebut. Uji beda (*T-Test*) merupakan metode pengujian hipotesis dengan menggunakan dua perlakuan yang berbeda. Walau dengan menggunakan objek yang sama namun sampel tetap dibagi menjadi 2 yaitu

dengan data perlakuan pertama dan data dengan perlakuan yang kedua. Uji beda (*T-Test*) bekerja dengan cara membandingkan 2 sampel yang berbeda [9]. Sedangkan untuk melihat nilai akurasi dapat dilihat dari hasil *Confusion Matrix*. *Confusion Matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur suatu kinerja pada metode klasifikasi. *Confusion Matrix* pada dasarnya membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang sebenarnya [10],[11]. Pada pengukuran kinerja menggunakan *Confusion Matrix* terdapat istilah sebagai representasi hasil proses klasifikasi yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP) dan *false negative* (FN) [12].

2.1.6. Deployment

Deployment merupakan tahapan implementasi atau penerapan yang perlu dilakukan setelah mengetahui pola atau pengetahuan untuk menyelesaikan masalah dengan mencapai tujuan yang sudah ditentukan. Hasil dari penelitian dapat direkomendasikan kepada instansi terkait agar dapat dijadikan masukan dalam memperbaiki masalah yang ada.

2.2. Feature Selection

Feature Selection adalah teknik pra pemrosesan data yang banyak digunakan dalam pengumpulan data yang pada dasarnya digunakan untuk mereduksi data dengan menghilangkan atribut yang tidak relevan dan menggunakan algoritma yang tepat dapat meningkatkan nilai akurasi [13]. Selain itu *Feature Selection* dapat meningkatkan pemahaman data, memfasilitasi visualisasi data yang lebih baik, mengurangi waktu pelatihan pada algoritma dan meningkatkan performa kinerja [14],[15].

2.3. Forward Selection

Forward Selection merupakan metode pencarian yang paling sederhana. *Forward Selection* merupakan teknik yang digunakan dalam mereduksi data dengan cara menghapus atribut yang tidak relevan. Pemodelannya akan dimulai dari nol peubah yang kemudian peubah satu persatu dimasukkan sampai pada kriteria tertentu dipenuhi [16]. *Forward Selection* dimulai dengan set kosong dan secara iteratif menambahkan fitur ke subset fitur target dari set fitur kandidat [17].

2.4. Backward Elimination

Backward Elimination merupakan metode yang mempunyai fungsi pengoptimalan performa untuk suatu model dengan sistem cara kerja pemilihan secara mundur, artinya semua variabel diuji setelah diuji variabel yang dianggap tidak relevan akan dihapus. Jika variabel yang akan diproses dianggap tidak berpengaruh maka akan dihapus juga dari

model. Variabel yang dimasukkan sudah selesai diproses maka proses eliminasi akan dihentikan [18].

2.5. Optimize Selection

Optimize Selection merupakan optimasi seleksi fitur yang mampu meningkatkan nilai akurasi secara optimal dengan memilih fitur terbaik [19]. *Optimize Selection* juga digunakan untuk mereduksi data yang tidak relevan. Namun untuk menentukan tujuannya metode ini menggunakan algoritma genetika.

2.6. Algoritma Klasifikasi

Klasifikasi merupakan proses yang dilakukan untuk menemukan pola karakteristik himpunan data dalam sebuah basis data dan mengklasifikasikan kedalam bentuk kelompok yang berbeda sesuai dengan model klasifikasi yang dihasilkan [20]. Proses klasifikasi bisa dilakukan setelah analisis relevansi yang bertujuan untuk menentukan atribut yang relevan dengan proses tersebut. Ketepatan prediksi dari suatu klasifikasi dapat dituangkan dalam tabel kontingensi (*Contingency Table*) atau *Confusion Matrix* (flach) [21].

2.7. Naïve Bayes

Naïve Bayes merupakan klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class* [22]. Kelebihan dari menggunakan algoritma *Naïve Bayes* adalah metode ini tidak menggunakan data *training* yang besar dalam menentukan estimasi atribut yang digunakan dalam proses klasifikasi dan mempunyai kinerja yang baik jika diimplementasikan pada dunia nyata [23]. Rumus dari algoritma *Naïve Bayes* dapat dilihat pada persamaan 1 [24].

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

Dimana :

B = Data *class* yang tidak diketahui

A = Hipotesis data B

P(A|B) = Prob. hipotesis A berdasarkan kondisi B

P(A) = Prob. hipotesis A

P(B|A) = Prob. hipotesis B berdasarkan kondisi A

P(B) = Prob. hipotesis B

2.8. Random Forest

Algoritma *Random Forest* merupakan salah satu metode yang dapat meningkatkan hasil akurasi karena dalam mengembangkan simpul untuk setiap *node* dilakukan secara acak [25]. Kelebihan dari *Random Forest* dapat menggabungkan banyak pohon dan untuk yang *single tree* yang terdiri hanya satu pohon dalam melakukan klasifikasi dan prediksi kelas [26]. Metode ini terdiri dari *root node*, *internal node* dan *leaf node*. *Root node* adalah simpul yang berada paling atas atau dengan kata lain disebut sebagai akar pohon. *Internal node* merupakan simpul

percabangan, dimana *node* ini mempunyai *output* minimal dua dan hanya ada satu *input*. Sedangkan *leaf node* adalah simpul akhir yang hanya mempunyai satu *input* dan tidak mempunyai *output*. Pohon keputusan dimulai dengan cara menghitung nilai *entropy* sebagai penentu tingkat ketidakmurnian atribut dan *information gain* [27]. Untuk menghitung nilai *entropy* dapat menggunakan persamaan 1, dan untuk menghitung nilai *information gain* dapat menggunakan persamaan 2 [28].

$$Entropy(Y) = -\sum_i p(p) \log_2 p(c|p) \quad (2)$$

Dimana:

Y = himpunan kasus

p(c|p) = proporsi nilai Y terhadap kelas c

$$Information\ Gain(Y, A) = Entropy(Y) -$$

$$\sum_{v \in values(a)} \frac{|Y_v|}{|Y_a|} Entropy(Y_v) \quad (3)$$

Dimana:

Values(a) = nilai yang mungkin dalam himpunan kasus a.

Y_v = subkelas dari Y dengan kelas v yang berhubungan dengan kelas a.

Y_a = semua nilai yang sesuai dengan a.

2.9. K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* adalah salah satu algoritma klasifikasi yang paling terkenal yang digunakan untuk memprediksi kelas dari sampel dengan kelas yang tidak ditentukan berdasarkan kelas dari sampel tetangganya [29]. *K-Nearest Neighbor* merupakan algoritma yang bekerja dengan mengelompokkan data berdasarkan nilai yang berdekatan dengan data tersebut [30]. Perhitungan jarak antara data dengan kelompok menggunakan perhitungan jarak *euclidean distance*. Rumus perhitungan *euclidean distance* dapat dilihat pada persamaan 3 [31].

$$Jarak = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} \quad (4)$$

Dimana : a1, a2, dan b1, b2 mewakili nilai atribut dari dua record.

Langkah-langkah dalam proses menggunakan algoritma *K-Nearest Neighbor* adalah sebagai berikut [32] :

1. Tentukan parameter dari k.
2. Hitung jarak menggunakan rumus *euclidean distance* pada persamaan 1.
3. Mengurutkan jarak data dari yang kecil ke yang besar.
4. Menentukan jarak terdekat sampai dengan urutan k.
5. Cari kelas yang sesuai.
Hitung jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

3. HASIL DAN PEMBAHASAN

3.1. Business Understanding

Business Understanding merupakan tahapan pertama untuk mengetahui suatu masalah. Permasalahan pada penelitian ini adalah gagal panen secara berulang dan berturut-turut memberikan dampak yang besar bagi para peternak. Selain mengganggu perekonomian, hal ini juga mempengaruhi kualitas peternak sehingga para peternak mendapatkan sanksi dari perusahaan. Pada penelitian ini tingkat keberhasilan panen diukur dari nilai indek performa yang didapat. Atribut yang digunakan dalam menghasilkan nilai indek performa yaitu FCR, jumlah panen, berat badan dan umur panen. Data IP ini yang akan dijadikan sebagai dataset yang akan diolah untuk menghasilkan pola dan pengetahuan mengenai keberhasilan panen ayam broiler.

3.2. Data Understanding

Data Understanding dimulai dengan mengumpulkan data hasil panen ayam broiler. Data yang digunakan didapat dari data dari hasil panen ayam pada perusahaan PT. Cemerlang Uggas Lestari. Berikut penjelasan dari masing-masing atribut yang ada pada data tersebut :

1. Kode peternak merupakan kode atau nomor peternak dari perusahaan.
2. DOC (*Day Old Chicken*) merupakan jumlah anak ayam yang masih berumur dibawah 10 hari yang akan digunakan sebagai bibit untuk peternak.
3. FCR (*Feed Conversion Ratio*) yaitu perbandingan dari total pakan yang akan digunakan dengan total bobot yang akan dihasilkan.
4. Jumlah panen merupakan jumlah ayam hidup (persen).
5. Berat panen merupakan rata-rata berat atau bobot panen (kg).
6. Umur merupakan rata-rata umur panen (hari).
7. Indek Performa merupakan nilai hasil panen ayam broiler.

Data hasil panen ayam broiler dapat dilihat pada tabel 1. Sedangkan kategori untuk hasil panen dapat dilihat pada tabel 2. Kategori hasil panen didapat dari hasil wawancara.

Tabel 1. Data Hasil Panen

Kode	DOC	FCR	Jumlah Panen	Berat Badan	Umur	IP
5901	14700	1,501	98,120	1,81	32	370
5581	20000	1,470	80,830	1,32	28	259
5611	15000	1,570	97,850	1,87	34	343
5902	14900	1,494	97,450	1,78	35	332
5591	20000	1,627	99,020	1,31	30	266
5711	12000	1,544	96,910	1,85	34	342
6591	18000	1,515	95,260	1,88	31	381
6601	17500	1,527	96,900	1,79	31	366
5831	6000	1,474	98,230	1,83	29	421
6701	12500	1,488	97,880	1,81	33	361

6702	12500	1,498	98,130	1,84	33	365
5551	4500	1,554	97,330	2,32	36	404
6611	7500	1,528	98,520	1,74	30	374
5061	14000	1,518	96,510	1,77	32	352
6581	24000	1,576	98,490	2,04	33	386
5361	6500	1,413	98,050	2,02	32	438
5612	15000	1,513	97,430	1,84	31	382
5582	20000	1,470	98,690	1,87	30	418
5541	7500	1,508	99,090	2,10	33	418
5821	21100	1,488	97,390	2,15	36	391

Tabel 2. Kategori Hasil Panen

No.	Indek Performa (IP)	Kategori
1	<299	Buruk
2	300-399	Bagus
3	>400	Sangat Bagus

3.3. Data Preparation

Pada *Data Preparation* atribut disesuaikan dengan metode *Data Mining* yang akan digunakan. Dataset yang digunakan dalam penelitian sebanyak 451 data. Pada tahapan *Data Selection* akan dilakukan analisis dan pemilihan atribut yang akan digunakan. Atribut yang akan digunakan yaitu atribut FCR, jumlah panen, berat badan, umur dan nilai indeks performa. Hasil dari pengolahan data pada tahap ini dapat dilihat pada tabel 3.

Tabel 3. Proses Data Selection

FCR	Jumlah Panen	Berat Badan	Umur	IP
1,501	98,120	1,81	32	370
1,470	80,830	1,32	28	259
1,570	97,850	1,87	34	343
1,494	97,450	1,78	35	332
1,627	99,020	1,31	30	266
1,544	96,910	1,85	34	342
1,515	95,260	1,88	31	381
1,527	96,900	1,79	31	366
1,474	98,230	1,83	29	421
1,488	97,880	1,81	33	361
1,498	98,130	1,84	33	365
1,554	97,330	2,32	36	404
1,528	98,520	1,74	30	374
1,518	96,510	1,77	32	352
1,576	98,490	2,04	33	386
1,413	98,050	2,02	32	438
1,513	97,430	1,84	31	382
1,470	98,690	1,87	30	418
1,508	99,090	2,10	33	418
1,488	97,390	2,15	36	391

Pada tahapan *Data Transformation*, sekumpulan data diubah kedalam bentuk data yang sesuai dengan proses pada *Data Mining*. Pada atribut IP akan diubah menjadi atribut kategori. Dimana nilai IP kurang dari 299 maka diberi label buruk, nilai IP 300-399 maka diberi label bagus dan sedangkan nilai IP diatas 400 maka diberi label sangat bagus. Tabel hasil dari proses *Data Transformation* dapat dilihat pada tabel 4.

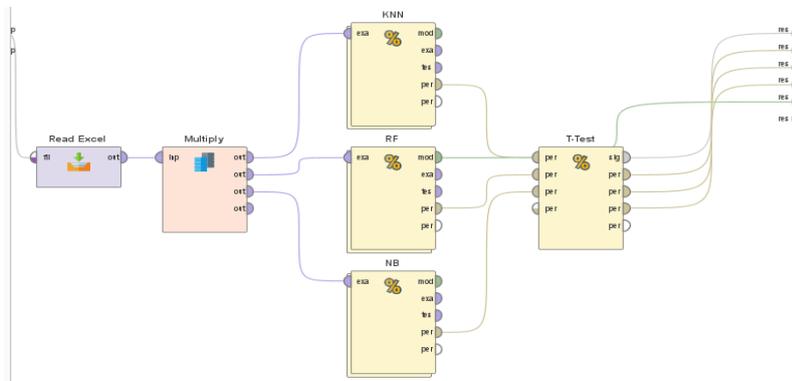
Tabel 4. Proses Data Transformasi

FCR	Jumlah Panen	Berat Badan	Umur	Kategori
1,501	98,120	1,81	32	Bagus
1,470	80,830	1,32	28	Buruk

1,570	97,850	1,87	34	Bagus	1,488	97,390	2,15	36	Bagus
1,494	97,450	1,78	35	Bagus				
1,627	99,020	1,31	30	Buruk					
1,544	96,910	1,85	34	Bagus					
1,515	95,260	1,88	31	Bagus					
1,527	96,900	1,79	31	Bagus					
1,474	98,230	1,83	29	Sangat Bagus					
1,488	97,880	1,81	33	Bagus					
1,498	98,130	1,84	33	Bagus					
1,554	97,330	2,32	36	Sangat Bagus					
1,528	98,520	1,74	30	Bagus					
1,518	96,510	1,77	32	Bagus					
1,576	98,490	2,04	33	Bagus					
1,413	98,050	2,02	32	Sangat Bagus					
1,513	97,430	1,84	31	Bagus					
1,470	98,690	1,87	30	Sangat Bagus					
1,508	99,090	2,10	33	Sangat Bagus					

3.4. Modeling

Modeling adalah suatu tahapan memilih teknik dalam *Data Mining* untuk menentukan algoritma yang akan digunakan. *Tools* yang akan digunakan dalam penelitian ini adalah dengan Rapid Miner versi 9.8. Data yang digunakan dalam penelitian ini dapat dilihat pada tabel 4. Algoritma klasifikasi yang terbaik didapat dengan cara membandingkan 3 algoritma yaitu *Naïve Bayes*, *Random Forest* dan *K-Nearest Neighbor*. Uji beda (*T-Test*) untuk membandingkan perbedaan signifikan antar 2 variabel. Hasil dari *T-Test* dapat menentukan algoritma klasifikasi yang terbaik. Model klasifikasi dengan 3 algoritma dapat dilihat pada gambar 2.

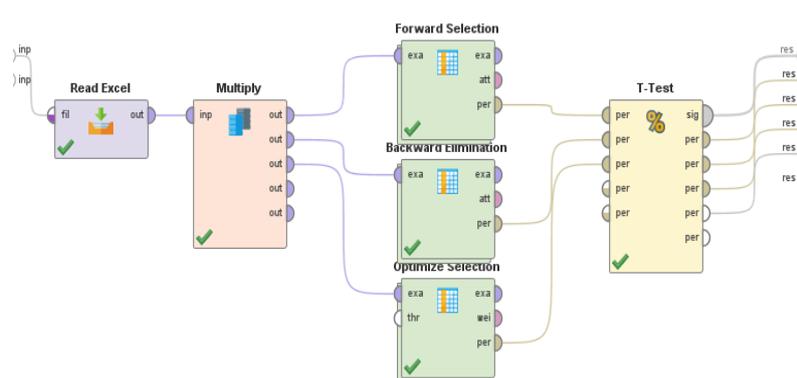


Gambar 2. Model Klasifikasi

Langkah awal dalam pemodelan algoritma pada Rapid Miner adalah dengan *drop operator read excel* pada halaman *process*, lalu *import* data yang digunakan, *setting* atribut dan label. Selanjutnya *drop operator multiply*, operator ini digunakan untuk menghubungkan dengan banyak operator agar dapat dijalankan dengan bersamaan. *Drop 3 cross validation* untuk masing-masing algoritma dengan *k-10 folds cross validation* dan *sampling type automatic*. Operator ini digunakan untuk menampilkan tingkat keakuratan algoritma. Didalam operator ini terdapat operator algoritma, *apply model* dan *performance*. Selanjutnya *drop operator T-Test*

guna membandingkan *performance* algoritma dengan alpha 0,05.

Metode *Feature Selection* akan digunakan untuk meningkatkan performa algoritma. Pada penelitian ini menggunakan *Forward Selection*, *Backward Elimination* dan *Optimize Selection*. *Feature Selection* dapat meningkatkan performa karena dapat menaikkan nilai akurasi. Membandingkan metode *Feature Selection* dapat menggunakan uji beda (*T-Test*). Uji beda (*T-Test*) digunakan untuk membandingkan perbedaan signifikan antar 2 variabel. Hasil dari *T-Test* dapat menentukan *Feature Selection* yang terbaik. Model dari *Feature Selection* dapat dilihat pada gambar 3.



Gambar 3. Model Feature Selection Random Forest

Langkah awal dalam pemodelan *Feature Selection* pada algoritma *Random Forest* adalah dengan *drop operator read excel* pada halaman *process*, lalu *import* data yang digunakan, *setting* atribut dan label. *Drop operator multiply*, operator ini untuk menghubungkan dengan banyak operator agar dapat dijalankan dengan bersamaan. *Drop operator Forward Selection, Backward Elimination* dan *Optimize Selection*. Didalam 3 operator ini terdapat

operator algoritma *Random Forest, apply model* dan *performance*. Terakhir *drop operator T-Test*.

3.5. Evaluation

Setelah melakukan *Modeling* dengan teknik *Data Mining*. Berikut merupakan hasil dari *modeling* dari masing-masing algoritma. Hasil *Confusion Matrix Naive Bayes* dapat dilihat pada gambar 4.

accuracy: 88.68% +/- 6.51% (micro average: 88.69%)

	true Bagus	true Buruk	true Sangat Bagus	class precision
pred. Bagus	242	27	15	85.21%
pred. Buruk	5	133	0	96.38%
pred. Sangat Bagus	4	0	25	86.21%
class recall	96.41%	83.12%	62.50%	

Gambar 4. *Confusion Matrix Naive Bayes*

$$\begin{aligned} \text{Accuracy} &= \frac{242 + 133 + 25}{451} \\ &= \frac{400}{451} \\ &= 88,69\% \end{aligned}$$

Berdasarkan pada hasil *Confusion Matrix* dan hasil perhitungan algoritma *Naive Bayes* didapat nilai akurasi *Naive Bayes* sebesar 88,69%. Hasil *Confusion Matrix* pada algoritma *Random Forest* dapat dilihat pada gambar 5.

accuracy: 89.13% +/- 3.71% (micro average: 89.14%)

	true Bagus	true Buruk	true Sangat Bagus	class precision
pred. Bagus	237	11	24	87.13%
pred. Buruk	7	149	0	95.51%
pred. Sangat Bagus	7	0	16	69.57%
class recall	94.42%	93.12%	40.00%	

Gambar 5. *Confusion Matrix Random Forest*

$$\begin{aligned} \text{Accuracy} &= \frac{237 + 149 + 16}{451} \\ &= \frac{402}{451} \\ &= 89,14\% \end{aligned}$$

Berdasarkan hasil dari *Confusion Matrix* yang dihasilkan dan perhitungan algoritma *Random Forest*, nilai akurasi yang didapat oleh algoritma *Random Forest* adalah sebesar 89,14%. Hasil dari *Confusion Matrix* algoritma *K-Nearest Neighbor* dapat dilihat pada gambar 6.

accuracy: 80.26% +/- 5.70% (micro average: 80.27%)

	true Bagus	true Buruk	true Sangat Bagus	class precision
pred. Bagus	224	31	31	78.32%
pred. Buruk	15	129	0	89.58%
pred. Sangat Bagus	12	0	9	42.86%
class recall	89.24%	80.62%	22.50%	

Gambar 6. *Confusion Matrix K-Nearest Neighbor*

$$\begin{aligned} \text{Accuracy} &= \frac{224 + 129 + 9}{451} \\ &= \frac{362}{451} \\ &= 80,27\% \end{aligned}$$

Dilihat dari hasil *Confusion Matrix* pada gambar 6 dan perhitungan dari algoritma *K-Nearest Neighbor*, nilai akurasi yang didapat yaitu sebesar 80,27%. Bentuk tabel dari *Confusion Matrix* dan

perhitungan dari masing-masing algoritma dapat dilihat pada tabel 5.

Tabel 5. Nilai Akurasi

Algoritma	Akurasi
<i>Naive Bayes</i>	88,69%
<i>Random Forest</i>	89,14%
<i>K-Nearest Neighbor</i>	80,27%

Dilihat dari tabel 5 nilai akurasi yang diperoleh dengan nilai akurasi tertinggi terdapat pada algoritma

Random Forest dengan nilai akurasi sebesar 89,14%. Hal ini menunjukkan bahwa algoritma Random Forest merupakan algoritma terbaik dalam

mengklasifikasikan hasil panen ayam broiler. Hasil dari uji beda T-Test dapat dilihat pada gambar 7 dibawah ini.

A	B	C	D
	0.803 +/- 0.057	0.891 +/- 0.037	0.887 +/- 0.065
0.803 +/- 0.057		0.001	0.006
0.891 +/- 0.037			0.852
0.887 +/- 0.065			

Gambar 7. Hasil Uji Beda (T-Test) Klasifikasi

Keterangan :

B = K-Nearest Neighbor

C = Random Forest

D = Naive Bayes

Berdasarkan hasil uji beda T-test, didapat algoritma dengan nilai alpha kurang dari 0,05 maka terdapat perbedaan yang signifikan. Pada gambar 7 algoritma yang menunjukkan nilai alpha kurang dari 0,05 merupakan algoritma K-Nearest Neighbor dengan Random Forest dan Naive Bayes. Sedangkan algoritma Random Forest dengan Naive Bayes menunjukkan nilai alpha lebih dari 0,05 yaitu 0,852. Hal ini menunjukkan bahwa algoritma Random Forest

dengan Naive Bayes tidak ada perbedaan yang signifikan. Sehingga dapat disimpulkan bahwa algoritma terbaik adalah algoritma Random Forest.

Berdasarkan perbandingan 3 algoritma tersebut telah didapat algoritma terbaik yaitu algoritma Random Forest. Langkah selanjutnya untuk mengetahui metode Feature Selection terbaik dalam meningkatkan performa algoritma adalah dengan membandingkan 3 Feature Selection yaitu Forward Selection, Backward Elimination dan Optimize Selection. Hasil dari Confusion Matrix Random Forest + Forward Selection dapat dilihat pada gambar 8.

accuracy: 96.45%

	true Bagus	true Buruk	true Sangat Bagus	class precision
pred. Bagus	250	2	13	94.34%
pred. Buruk	1	158	0	99.37%
pred. Sangat Bagus	0	0	27	100.00%
class recall	99.60%	98.75%	67.50%	

Gambar 8. Confusion Matrix Random Forest + Forward Selection

$$\begin{aligned}
 \text{Accuracy} &= \frac{250 + 158 + 27}{451} \\
 &= \frac{435}{451} \\
 &= 96,45\%
 \end{aligned}$$

Berdasarkan pada hasil Confusion Matrix dan hasil perhitungan algoritma Random Forest + Forward Selection didapat nilai akurasi sebesar 96,45%. Hasil Confusion Matrix pada algoritma Random Forest + Backward Elimination dapat dilihat pada gambar 9.

accuracy: 96.67%

	true Bagus	true Buruk	true Sangat Bagus	class precision
pred. Bagus	251	2	13	94.36%
pred. Buruk	0	158	0	100.00%
pred. Sangat Bagus	0	0	27	100.00%
class recall	100.00%	98.75%	67.50%	

Gambar 9. Confusion Matrix Random Forest + Backward Elimination

$$\begin{aligned}
 \text{Accuracy} &= \frac{251 + 158 + 27}{451} \\
 &= \frac{436}{451} \\
 &= 96,67\%
 \end{aligned}$$

Berdasarkan hasil dari Confusion Matrix yang dihasilkan dan perhitungan algoritma Random Forest + Backward Elimination, nilai akurasi yang didapat oleh algoritma Random Forest + Backward Elimination adalah sebesar 96,67%. Hasil dari Confusion Matrix algoritma Random Forest + Optimize Selection dapat dilihat pada gambar 10.

accuracy: 96.23%

	true Bagus	true Buruk	true Sangat Bagus	class precision
pred. Bagus	250	3	13	93.98%
pred. Buruk	1	157	0	99.37%
pred. Sangat Bagus	0	0	27	100.00%
class recall	99.60%	98.12%	67.50%	

Gambar 10. Confusion Matrix Random Forest + Optimize Selection

Dilihat dari hasil *Confusion Matrix* pada gambar 10 dan perhitungan dari algoritma *Random Forest + Optimize Selection*, nilai akurasi yang didapat yaitu sebesar 96,23%. Bentuk tabel dari *Confusion Matrix* dan perhitungan dari masing-masing algoritma dapat dilihat pada tabel 6.

Tabel 6. Hasil Algoritma *Random Forest + Feature Selection*

Algoritma	Akurasi	Selisih
<i>Random Forest + Forward Selection</i>	96,45%	7,31%
<i>Random Forest + Backward Elimination</i>	96,67%	7,53%

<i>Random Forest + Optimize Selection</i>	96,23%	7,09%
-------------------------------------------	--------	-------

Dilihat dari tabel hasil nilai akurasi yang diperoleh, nilai akurasi tertinggi terdapat pada algoritma *Random Forest + Backward Elimination* dengan nilai akurasi sebesar 96,67% dengan selisih 7,53%. Hal ini menunjukkan bahwa algoritma *Random Forest + Backward Elimination* merupakan algoritma terbaik dalam mengklasifikasikan hasil panen ayam broiler. Hasil dari uji beda *T-Test* dapat dilihat pada gambar 11 dibawah ini.

A	B	C	D
	0.965 +/- ?	0.967 +/- ?	0.962 +/- ?
0.965 +/- ?		1.000	1.000
0.967 +/- ?			1.000
0.962 +/- ?			

Gambar 11. Hasil Uji Beda (*T-Test*) *Feature Selection*

Keterangan :

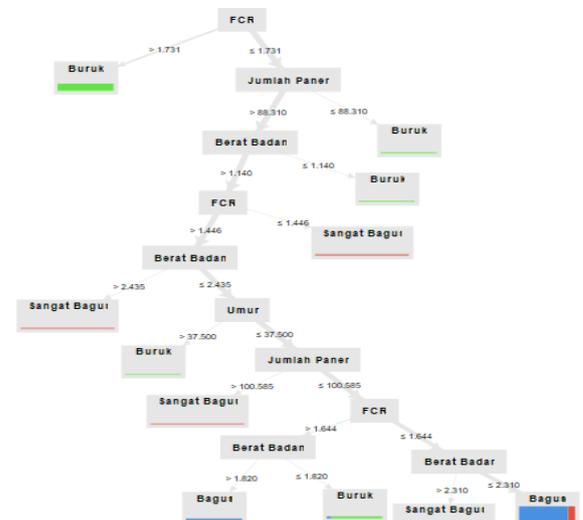
- B = *Forward Selection*
- C = *Backward Elimination*
- D = *Optimize Selection*

Berdasarkan hasil uji beda *T-test*, didapat algoritma dengan nilai alpha lebih dari 0,05 maka tidak ada perbedaan yang signifikan. Pada gambar 11 *Forward Selection*, *Backward Elimination* dan *Optimize Selection* mempunyai nilai alpha lebih dari 0,05, hal ini menunjukkan bahwa ketiga metode *Feature Selection* tersebut tidak ada perbedaan yang signifikan. Sehingga ketiga algoritma tersebut dapat dikatakan baik dalam meningkatkan performa algoritma *Random Forest*.

Berdasarkan hasil *confusion matrix* dan hasil uji beda *T-Test* algoritma terbaik dalam mengklasifikasikan data hasil panen ayam broiler adalah algoritma *Random Forest*. Sedangkan metode *Feature Selection* yang baik untuk meningkatkan performa algoritma *Random Forest* adalah metode *Forward Selection*, *Backward Elimination* dan *Optimize Selection*. Namun berdasarkan nilai akurasi tertinggi dan selisih tertinggi dalam penambahan *Feature Selection* dengan tidak menggunakan *Feature Selection* adalah *Random Forest + Backward Elimination* dengan nilai akurasi yang didapat 96,67% dengan selisih 7,53%. Sehingga dapat dikatakan algoritma *Random Forest + Backward Elimination* merupakan algoritma dan *Feature Selection* terbaik.

Pada gambar 12 merupakan hasil pohon keputusan dari hasil panen yang telah dimodelkan

sebelumnya. untuk mempermudah analisis berikut deskripsi pohon keputusan tersebut :



Gambar 12. Pohon Keputusan

Tree

```

FCR > 1.731: Buruk (Bagus=0, Buruk=116, Sangat Bagus=0)
FCR ≤ 1.731
| Jumlah Panen > 88.310
| | Berat Badan > 1.140
| | | FCR > 1.446
| | | | Berat Badan > 2.435: Sangat Bagus (Bagus=0, Buruk=0, Sangat Bagus=3)
| | | | Berat Badan ≤ 2.435
| | | | | Umur > 37.500: Buruk (Bagus=0, Buruk=1, Sangat Bagus=0)
| | | | | Umur ≤ 37.500
| | | | | | Jumlah Panen > 100.585: Sangat Bagus (Bagus=0, Buruk=0, Sangat Bagus=1)
| | | | | | Jumlah Panen ≤ 100.585
| | | | | | | FCR > 1.644
| | | | | | | | Berat Badan > 1.820: Bagus (Bagus=33, Buruk=0, Sangat Bagus=0)
| | | | | | | | Berat Badan ≤ 1.820: Buruk (Bagus=2, Buruk=23, Sangat Bagus=0)
| | | | | | | | FCR ≤ 1.644
| | | | | | | | | Berat Badan > 2.310: Sangat Bagus (Bagus=0, Buruk=0, Sangat Bagus=3)
| | | | | | | | | Berat Badan ≤ 2.310: Bagus (Bagus=221, Buruk=1, Sangat Bagus=30)
| | | | | | | | | FCR ≤ 1.446: Sangat Bagus (Bagus=0, Buruk=0, Sangat Bagus=9)
| | | | | | | | | Berat Badan ≤ 1.140: Buruk (Bagus=0, Buruk=3, Sangat Bagus=0)
| | | | | | | | | Jumlah Panen ≤ 88.310: Buruk (Bagus=0, Buruk=5, Sangat Bagus=0)
    
```

Gambar 13. Deskripsi Pohon Keputusan

Berdasarkan dari pohon keputusan diatas, dapat dilihat dan disimpulkan bahwa faktor atau atribut yang paling berpengaruh dalam meningkatkan hasil panen ayam broiler yang dilihat dari hasil indek performanya yaitu dari faktor FCR, jumlah panen dan berat badan. Sedangkan faktor yang tidak berpengaruh adalah pada faktor umur panen.

3.6. Deployment

Berdasarkan penelitian yang sudah dilakukan, telah didapat pola yang ada dalam proses *Data Mining*. Algoritma yang paling akurat dalam klasifikasi hasil panen adalah algoritma *Random Forest*. Serta untuk meningkatkan performa algoritma dapat menggunakan *Backward Elimination*. Serta diketahui atribut yang berpengaruh dalam peningkatan hasil panen ayam broiler adalah FCR, jumlah panen dan berat badan. Sehingga hal ini dapat menjadikan perhatian bagi para peternak agar lebih memperhatikan atribut yang paling berpengaruh dalam keberhasilan panen dan dapat menjadikan evaluasi bagi para peternak agar hasil panen menjadi baik.

4. KESIMPULAN

Pada penelitian ini telah dilakukan klasifikasi data hasil panen ayam broiler dengan menggunakan teknik *Data Mining*. Hasil dari membandingkan 3 algoritma klasifikasi didapat algoritma terbaik untuk mengklasifikasikan data hasil panen ayam broiler adalah algoritma *Random Forest* dengan nilai akurasi sebesar 89,14%. Sedangkan metode terbaik untuk meningkatkan performa algoritma adalah metode *Backward Elimination* dengan peningkatan performa metode tersebut sebesar 7,53%. Sehingga nilai akurasi yang didapat dari algoritma *Random Forest* + *Backward Elimination* menjadi 96,67%. Selain itu dapat diketahui atribut yang berpengaruh dalam peningkatan hasil panen yaitu FCR, jumlah panen dan berat badan. Hal ini dapat dijadikan perhatian lebih bagi para peternak agar hasil panen menjadi lebih baik.

DAFTAR PUSTAKA

- [1] T. Nuryati, "Analisis Performans Ayam Broiler Pada Kandang Tertutup Dan Kandang Terbuka," *J. Peternak. Nusant.*, vol. 5, no. 2, pp. 77–86, 2019.
- [2] S. Pakage *et al.*, "Pengukuran Performa Produksi Ayam Pedaging pada Closed House System dan Open House System di Kabupaten Malang Jawa Timur Indonesia," *J. Sain Peternak. Indones.*, vol. 15, no. 4, pp. 383–389, 2020.
- [3] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, "Comparison Of Data Mining Algorithm For Forecasting Bitcoin Crypto Currency Trends," *J. Tek. Inform.*, vol. 3, no. 2, pp. 245–248, 2022.
- [4] Y. Religia, A. Nugroho, and H. Wahyu, "Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 187–192, 2021.
- [5] S. Lestari and H. A. Silaban, "Implementasi Data Mining Dalam Penerbitan Surat Penetapan Tarif Dan Nilai Pabean Menggunakan Metode Classification Pada Direktorat Jenderal Bea Dan Cukai," *CKI SPOT*, vol. 11, no. 2, pp. 138–149, 2018.
- [6] A. Fauzi, "Analisis Data Bank Direct Marketing dengan Perbandingan Klasifikasi Data Mining Berbasis Optimize Selection (Evolutionary)," *J. Inform. Univ. Pamulang*, vol. 6, no. 1, pp. 102–111, 2021.
- [7] S. Amri, "Perbandingan Kerangka Model Klasifikasi untuk Pemilihan Metode Kontrasepsi dengan Pendekatan CRIPS-DM," *Inf. Sci. Libr.*, vol. 1, no. 1, pp. 14–23, 2020.
- [8] A. Yumalia and R. E. Indrajit, "Penerapan Konsep Business Intelligence Untuk Percepatan Penyelesaian Perkara Pada Panmud Perdata Khusus Mahkamah Agung RI," *IKRAITH-INFORMATIKA*, vol. 1, no. 2, pp. 61–69, 2017.
- [9] P. A. Rahayuningsih, "Komparasi Algoritma Klasifikasi Data Mining untuk Memprediksi Tingkat Kematian Dini Kanker dengan Dataset Early Death Cancer," *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 2, pp. 63–68, 2019.
- [10] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J-SAKTI*, vol. 5, no. 2, pp. 697–711, 2021.
- [11] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Citec J.*, vol. 6, no. 1, pp. 1–11, 2019.
- [12] Karsito and S. Susanti, "Pengajuan Kredit Rumah Dengan ALgoritma Naïve Bayes Di Perumahan Azzura Residencia," *SIGMA – J. Teknol. Pelita Bangsa*, vol. 9, no. 3, pp. 43–48, 2019.
- [13] O. Somantri and M. Khambali, "Feature Selection Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes dan Algoritme Genetika," *JNTETI*, vol. 6, no. 3, pp. 301–306, 2017.
- [14] O. Cigdem and H. Demirel, "Performance Analysis of Different Classification Algorithms Using Different Feature Selection Methods on Parkinson's Disease Detection," *J. Neurosci. Methods*, pp. 1–13,

- 2018, doi: 10.1016/j.jneumeth.2018.08.017.
- [15] D. Jain and V. Singh, "Feature Selection and Classification Systems for Chronic Disease Prediction: A Review," *Egypt. Informatics J.*, 2018, doi: 10.1016/j.eij.2018.03.002.
- [16] M. F. Nugroho and S. Wibowo, "Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes," *J. Inform. UPGRIS*, vol. 3, no. 1, pp. 63–70, 2017.
- [17] Z. He, L. Li, Z. Huang, and H. Situ, "Quantum-Enhanced Feature Selection With Forward Selection and Backward Elimination," *Quantum Inf Process*, pp. 1–11, 2018, doi: 10.1007/s11128-018-1924-8.
- [18] I. C. R. Drajana, "Metode Support Vector Machine Dan Forward Selection Prediksi Pembayaran Pembelian Bahan Baku Kopra," *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 116–123, 2017.
- [19] R. T. Prasetyo and E. Ripandi, "Optimasi Klasifikasi Jenis Hutan Menggunakan Deep Learning Berbasis Optimize Selection," *J. Inform.*, vol. 6, no. 1, pp. 100–106, 2019.
- [20] A. R. Kadafi, "Perbandingan Algoritma Untuk Klasifikasi Nilai Pada Penjurusan Siswa SMA," *J. ELTIKOM*, vol. 2, no. 2, pp. 67–77, 2018.
- [21] Ardiyansyah, P. A. Rahayuningsih, and R. Maulana, "Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner," *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 20–28, 2018.
- [22] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 219–225, 2021, doi: 10.30865/jurikom.v8i6.3655.
- [23] T. Imandasari, E. Irawan, A. P. Windarto, and A. Wanto, "Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air," *Pros. Semin. Nas. Ris. Inf.*, pp. 750–761, 2019.
- [24] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018.
- [25] R. G. Ramli and Y. Sibaroni, "Klasifikasi Topik Twitter menggunakan Metode Random Forest dan Fitur Ekspansi Word2Vec," *e-Proceeding Eng.*, vol. 9, no. 1, pp. 79–92, 2022.
- [26] B. Prasajo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest," *J. Nas. Teknol. dan Sist. Inf.*, vol. 07, no. 02, pp. 79–89, 2021.
- [27] G. A. Sandag, "Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest," *Cogito Smart J.*, vol. 6, no. 2, pp. 167–178, 2020.
- [28] V. W. Siburian and I. E. Mulyana, "Prediksi Harga Ponsel Menggunakan Metode Random Forest," *Pros. Annu. Res. Semin.*, vol. 4, no. 1, pp. 144–147, 2018.
- [29] M. Faid, M. Jasri, and T. Rahmawati, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," *TEKNIKA*, vol. 8, no. 1, pp. 11–16, 2019, doi: 10.34148/teknika.v8i1.95.
- [30] R. R. Baharuddin, M. Niswar, and A. A. Ilham, "Deteksi Kepiting Molting Menggunakan Teknik Klasifikasi Machine Learning," *J. J-Ensatec*, vol. 08, no. 01, pp. 599–610, 2021.
- [31] Rudiono and D. Avianto, "Implementasi Ekstraksi Ciri Histogram dan K-Nearest Neighbor untuk Klasifikasi Jenis Tanah Di Kota Banjar, Jawa Barat," *J. Buana Inform.*, vol. 10, no. 2, pp. 85–98, 2019.
- [32] L. Andiani, Sukemi, and D. P. Rini, "Analisis Penyakit Jantung Menggunakan Metode KNN Dan Random Forest," *Pros. Annu. Res. Semin.*, vol. 5, no. 1, pp. 165–169, 2019.