

Perbandingan Kinerja Algoritma Naïve Bayes Dan C.45 Dalam Klasifikasi Spam Email

Sulaeman¹, Nana Suarna², Abdul Ajiz³ Agus Bahtiar⁴, Fathurrohman⁵

Program Studi Teknik Informatika, STMIK IKMI Cirebon, Kota Cirebon Indonesia¹²

Program Studi Komputerisasi Akuntansi, STMIK IKMI Cirebon, Kota Cirebon Indonesia³

Program Studi Sistem Informasi, STMIK IKMI Cirebon, Kota Cirebon Indonesia⁴

Program Studi Rekayasa Perangkat Lunak, STMIK IKMI Cirebon, Kota Cirebon Indonesia⁵

Email : lemansulaeman11@gmail.com¹, st_nana@yahoo.co.id², abdulajiz00@gmail.com³, agusbahtia00⁴, fathur00@gmail.com⁵

Email Penulis Korespondensi : lemansulaeman11@gmail.com

Submitted : 09-06-2022; Accepted 23-06-2022; Published 24-06-2022

Abstract-Antispam with a certain algorithm that can separate spam-mail from non-spam mail. The performance comparison between the naïve bayes algorithm and the decision tree using the C4.5 algorithm proves that the decision tree with the C4.5 algorithm is more efficient and the simplest when compared to the other three algorithms. With its simplicity, the C4.5 algorithm gives better results for spam-mail classification. The Naïve Bayes and C4.5 algorithms perform well in identifying whether an email is spam or non-spam. However, it is not yet known which of the two is superior in performance. Therefore, these two algorithms need to be compared based on the results of the accuracy of yahoo Naïve Bayes accuracy of 96.70% with details, namely Prediction of Ham and true Ham of 3385 Data, Prediction of Ham and true Spam of 165 Data, Prediction of Spam and true Ham of 0 Data , Predicted Spam and True Spam as much as 1448 Data. Based on the results of yahoo C.45 accuracy, the quality of accuracy is 96.68% as many as details, namely Ham Prediction and true Ham 3385 Data, Ham Prediction and true Spam 166 Data, Spam Prediction and true Ham as much as 0 Spam Prediction Data and True Spam as much as 1447 Data. Based on the results of the comparison test, the best Yahoo results obtained by measuring the level of accuracy so that the C.45 algorithm can be obtained which has a value of 96.68%. Then on the application of the Nave Bayes algorithm model explains that the level of accuracy can be obtained from the Nave Bayes algorithm with a value of 96.79 %. It can be interpreted that the Nave Bayes data algorithm is categorized as a decision-making guide

Keyword : Comparison, Spam, Ham.

Abstrak-Antispam dengan algoritma tertentu yang dapat memisahkan antara spam-mail dengan non spam mail. Perbandingan kinerja antara algoritma naïve bayes, dan decision tree yang memakai algoritma C4.5 membuktikan bahwa decision tree dengan algoritma C4.5 lebih efisien dan paling sederhana jika dibandingkan dengan algoritma yang lain. Dengan kesederhanaannya, algoritma C4.5 memberikan hasil yang lebih baik untuk klasifikasi spam-mail. Algoritma Naïve Bayes dan C4.5 mempunyai kinerja yang baik dalam mengidentifikasi apakah suatu email adalah spam atau non-spam. Namun, belum diketahui algoritma mana diantara keduanya yang lebih unggul kinerjanya. Oleh karena itu kedua algoritma ini perlu dibandingkan Berdasarkan hasil akurasi algoritma Naïve Bayes menghasilkan akurasi sebesar 96,70% dengan rincian yaitu Prediksi Ham dan true Ham sebanyak 3385 Data, Prediksi Ham dan true Spam sebanyak 165 Data, Prediksi Spam dan true Ham sebanyak 0 Data, Prediksi Spam dan true Spam sebanyak 1448 Data. Berdasarkan hasil akurasi algoritma C.45 menghasilkan akurasi sebesar 96,68% dengan rincian yaitu Prediksi Ham dan true Ham sebanyak 3385 Data, Prediksi Ham dan true Spam sebanyak 166

Data, Prediksi Spam dan true Ham sebanyak 0 Data Prediksi Spam dan true Spam sebanyak 1447 Data. Berdasarkan hasil uji komparasi diperoleh hasil algoritma terbaik dengan mengukur tingkat hasil akurasi sehingga dapat diperoleh algoritma C4.5 memiliki nilai sebesar 96.68% Kemudian pada penerapan model algoritma naïve bayes menjelaskan bahwa tingkat hasil akurasi dapat diperoleh dari algoritma naïve bayes dengan nilai sebesar 96.79%. bisa di artikan bahwa algoritma naïve bayes data dikategorikan sebagai pedoman pengambilan keputusan.

Kata Kunci : Komparasi, Spam, Ham.

I. PENDAHULUAN

Perkembangan teknologi internet sekarang berkembang secara pesat, banyak manfaat yang dirasakan masyarakat namun banyak bentuk serangan terhadap sebuah webset yang dilakukan oleh hacker, cracker dan virus, salah satu fasilitas internet yang sering digunakan adalah e-mail. Bentuk penyerangan terhadap e-mail sering kita temui yaitu adanya spam yang sering masuk ke email, pengguna email tentunya terganggu sehingga merugikan pengguna email atau perusahaan. Penggunaan alat bantu untuk memudahkan kinerja saat menggunakan email tentunya sangat penting agar kita lebih efisien dalam menentukan mana spam email dan mana non-spam email. Penggunaan alat bantu yaitu dengan pemasangan sistem Algoritma Naive Bayes dan C.45, kedua alat ini mana yang lebih baik kinerjanya.

Spam messages membanjiri internet dengan mengirimkan salinan pesan-pesan yang sama untuk memaksa agar pesan-pesan tersebut sampai kepada pemakai yang tidak memilih untuk menerimanya. Akibatnya banyak pemakai yang merasa terganggu oleh banyaknya waktu yang dihabiskan untuk menghapus pesan spam, besarnya biaya yang harus dikeluarkan, dan besarnya bandwidth jaringan. Dalam klasifikasi spam email atau bukan perlu dilakukan kajian yang sangat mendalam, artinya perlu dilakukan uji kinerja beberapa metode yang tepat dalam menganalisis spam seperti ini.

Jurnal Informatika dan Komputer dengan judul Klasifikasi Algoritma Naïve Bayes dan SVM Berbasis PSO Dalam Memprediksi Spam Email Pada Hotline-Sapto. Hasil Penelitian tersebut hasil pengujian diperoleh perhitungan SVM metode dengan PSO mendapatkan nilai akurasi sebesar

85,25% dengan AUC sebesar 0,892 (Hengki & Wahyudi, 2020).

jurnal Jurnal Teknovasi Volume 03, Nomor 2, 2016, 5157 dengan judul Analisis Spam Dengan Menggunakan Naïve Bayes. Hasil pengklasifikasian yang disajikan dalam bentuk tabel dapat terlihat dengan jelas informasi kategori spam atau bukan spam. Diharapkan untuk penelitian selanjutnya dapat dilakukan perbandingan dengan metode algoritma yang lain dan data training yang lebih banyak lagi (Juang, 2016).

Dalam melakukan kinerja model klasifikasi dibutuhkan analisa yang tepat untuk pemilihan metode, penulis mencoba melakukan penelitian eksperimen yaitu dengan membandingkan kinerja dari algoritma naïve bayes dan algoritma c.45 adapun data yang akan digunakan diambil dari <https://www.kaggle.com/> yang diakses pada tanggal 20 September 2021. Masalah yang akan di uji untuk klarifikasi ham dan spam sebagai berikut : Dalam klasifikasi email belum adanya filter kategori spam dan tidak spam. Tidak tersedianya alat atau tool yang dapat memfilter spam email. Peningkatan spam email yang tinggi mengakibatkan penumpukan jumlah email.

Permasalahan yang akan dicari solusinya dengan penelitian yang akan dilakukan: Penelitian ini menggunakan dataset spam dan not spam berasal dari <https://www.kaggle.com/> yang diakses pada tanggal 20 September 2021 Penelitian ini menggunakan Algoritma Naïve Bayes Dan C.45, data set menggunakan aplikasi excel dan tool RapidMiner . Hasil penelitian ini menghasilkan kinerja terbaik pada Algoritma Naïve Bayes Dan C.45.

II. METODOLOGI PENELITIAN

Penelitian yang dilakukan oleh Meriohengki, Mochamad Wahyudi, pada jurnal Paradigma – Jurnal Informatika dan Komputer dengan judul Klasifikasi Algoritma Naïve Bayes dan SVM Berbasis PSO Dalam Memprediksi Spam Email Pada Hotline-Sapto Menjelaskan bahwa Akreditasi dapat diartikan sebagai upaya pemerintah untuk membakukan dan menjamin kualitas alumni perguruan tinggi sehingga kualitas verifikasi antar perguruan tinggi tidak terlalu variatif dan dalam sesuai dengan kebutuhan kerja. Pada kesempatan kali ini penelitian yang akan dilakukan adalah klasifikasi email spam dari akun hotline-sapto dan preprocessing serta perhitungan akurasinya, AUC dengan berbagai metode klasifikasi data mining, termasuk algoritma Naïve Bayes, Support Vector Machine (SVM), metode ini digunakan untuk memprediksi email spam dengan tujuan algoritma yang paling banyak dipilih algoritma akurat yang dapat memprediksi email spam. Dari hasil pengujian diperoleh perhitungan SVM metode dengan PSO mendapatkan nilai akurasi sebesar 85,25% dengan AUC sebesar 0,892 (Hengki & Wahyudi, 2020) Penelitian yang dilakukan Darma Juang pada Jurnal Teknovasi Volume 03, Nomor 2, 2016, 5157 dengan judul Analisis Spam Dengan Menggunakan Naïve Bayes, menjelaskan bahwa Perkembangan teknologi internet sekarang berkembang secara pesat. Banyak bentuk-bentuk penyerangan terhadap sebuah situs web site. Dengan makin banyaknya virus, kenyamanan saat berinternet-an pun ikut berkurang salah satunya adalah e-mail. SPAM (Stupid pointless Annoying Message) yang berarti e-mail sampah.

Tujuan penelitian ini dilakukan untuk memahami cara kerja metode naïve bayes di dalam menganalisis spam berdasarkan frekuensi kata. hasil penelitian yang sudah dilakukan bahwa algoritma naïve bayes dapat mengklasifikasikan suatu pesan ke dalam dua kelas yaitu spam dan non spam. Dari pengklasifikasian tersebut sangat dipengaruhi oleh proses training sehingga dapat disimpulkan hasil pengklasifikasian yang disajikan dalam bentuk tabel dapat terlihat dengan jelas informasi kategori spam atau bukan spam. Diharapkan untuk penelitian selanjutnya dapat dilakukan perbandingan dengan metode algoritma yang lain dan data training yang lebih banyak lagi[2][4]

Penelitian yang dilakukan oleh antonius rachmat chrismanto, yuan lukito, anton susilo pada jurnal jurnal edukasi dan penelitian informatika dengan judul implementasi distance weighted k-nearest neighbor untuk klasifikasi spam & non-spam pada komentar instagram menjelaskan instagram (ig) menjadi salah satu sosial media yang sering dipakai untuk membagikan momen dari para penggunanya. Tujuan penelitian ini adalah mengimplementasikan dan mengetahui akurasi algoritma dwknn untuk deteksi komentar spam pada ig. Metode dwknn digunakan sebagai perbaikan dari metode knn melalui pelatihan sistem dengan data latih acak. Setelah proses pelatihan, dilakukan pengujian berdasarkan data uji dan latih dengan parameter nilai k dan persentase fitur yang akan digunakan untuk menguji dan membandingkan metode knn maupun dwknn berdasarkan hasil klasifikasinya. Kontribusi penelitian ini menunjukkan bahwa akurasi metode dwknn lebih baik daripada knn, perbedaan nilai k ini tidak memiliki dampak yang terlalu berarti dalam klasifikasi komentar spam, dan seleksi fitur (features selection) memiliki hasil success rate yang baik pada penggunaan fs antara 80% - 100%. Akurasi optimal dari knn adalah 82.36% sedangkan menggunakan dwknn mencapai 91.08% pada fs 80%. [6][8]

Penelitian yang dilakukan oleh eni pudjiarti pada jurnal jurnal pilar nusa mandiri vol.xii, no.2 september 2016 dengan judul prediksi spam email menggunakan metode support vector machine dan particle swarm optimization menjelaskan bahwa spam email adalah istilah yang digunakan untuk menggambarkan pesan yang dikirim dalam email massal atau email yang masuk diterima tanpa persetujuan. Spam email filtering adalah program yang digunakan untuk mendeteksi email yang tidak diinginkan dan mencegah email yang tidak diminta dan untuk masuk ke inbox pengguna email. Setelah menguji dua model, yaitu support vector machine mesin dan support vector machine berbasis particle swarm optimization, hasil yang diperoleh dengan demikian diperoleh menguji algoritma menggunakan support vector machine yang merupakan nilai yang diperoleh akurasi 85,75% dan nilai auc adalah 0,901, sedangkan tes menggunakan support vector machine berbasis particle swarm optimization nilai yang diperoleh akurasi 89,24% dan nilai auc adalah 0,935 dengan tingkat yang baik klasifikasi diagnostik. Sehingga kedua metode memiliki berbagai tingkat akurasi yaitu sebesar 03:49% dan nilainilai auc 0,034 perbedaan (Pudjiarti, 2016) Penelitian yang dilakukan oleh nur qodariyah fitriyah, hardian oktavianito, hasbullah pada jurnal jurnal sistem & teknologi informasi indonesia dengan judul deteksi spam pada email berbasis fitur konten menggunakan naïve bayes menjelaskan bahwa penelitian menunjukkan bahwa terdapat lebih dari 3 milyar akun email di dunia dengan frekuensi pengiriman email sekitar 205 – 294 milyar setiap hari. Salah satu masalah

yang muncul dari pengiriman email yang luar biasa ini adalah adanya spam email, salah satu solusi untuk mengatasi permasalahan spam email tersebut adalah dengan teknik penyaringan spam email. Penyaringan spam email dapat dilakukan dengan menggunakan pendekatan teori berbasis pembelajaran, yaitu dengan klasifikasi. Penelitian ini menerapkan algoritma naive bayes untuk melakukan klasifikasi spam email sehingga dari dataset email, akan dikelompokkan menjadi 2 yaitu spam email dan non- spam email. Hasil uji dengan menggunakan k-fold cross validation sebagai pembagian data latih dan data uji, menghasilkan kesimpulan bahwa nilai rata – rata data terklasifikasi benar adalah sebesar 3903, sedangkan nilai rata – rata data terklasifikasi salah adalah sebesar 698, rata – rata akurasi sebesar 84.8%, sedangkan rata – rata precision dan recall berturut – turut adalah 0.86 dan 0.85. Akurasi, precision, dan recall tertinggi diperoleh ketika menggunakan nilai k=9.[8]

A. Teknik Pengumpulan Data

Adapun teknik pengumpulan data merupakan suatu cara yang dilakukan oleh peneliti maka penelitian ini teknik pengumpulan data yaitu wawancara dan obsevasi[7].

Adapun teknik pengumpulan data merupakan suatu cara yang dilakukan oleh peneliti untuk memperoleh data-data yang diperlukan. Sumber data pada penelitian ini menggunakan Data Skunder dimana data Penelitian ini menggunakan dataset spam dan not spam berasal dari <https://www.kaggle.com/> yang diakses pada tanggal 20 September 2021.

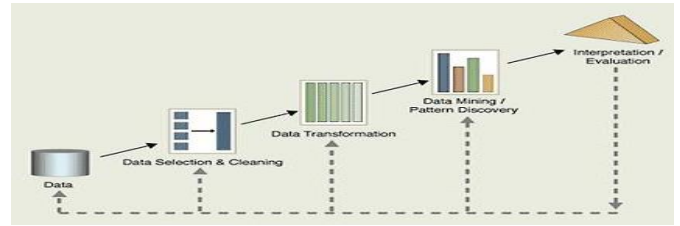
Adapun data yang didapat yaitu menggunakan data 5390 data spam dan ham, data dapat dilihat sebagai berikut :

Jumlah	label	text	label_num
605	ham	Subject: enron methanol ; n	0
2349	ham	Subject: hpl nom for january	0
3624	ham	Subject: neon retreatsho h	0
4685	spam	Subject: photoshop , windo	1
2030	ham	Subject: re : indian springst	0
2949	ham	Subject: ehronline web add	0
2793	ham	Subject: spring savings certifi	0
4185	spam	Subject: looking for medicat	1
2641	ham	Subject: noms / actual flow i	0
1870	ham	Subject: nominations for oct	0
4922	spam	Subject: vocable % rnd - wor	1
3799	spam	Subject: report 01405 lwffur	1
1488	ham	Subject: enron / hpl actuals	0
3948	spam	Subject: vic. odin n ^ owber	1
3418	ham	Subject: tenaska iv julydarre	0
4791	spam	Subject: underpriced issue v	1
2643	ham	Subject: re : first delivery - v	0
3137	ham	Subject: swift - may 2001 vol	0
1629	ham	Subject: meter variances - u	0
1858	ham	Subject: additional recruitin	0
3261	ham	Subject: fw : ercot load com	0
3447	ham	Subject: meter 6461 , concor	0
2459	ham	Subject: hpl nom for january	0

Gambar 1. Data Kagle Spam dan Ham

B. Tahapan Penelitian

Metode pengembangan data mining yang digunakan untuk menganalisis data dalam penerapan data mining ini menggunakan proses tahapan *knowledge discovery in databases* (KDD) yang terdiri dari Data, Data Cleaning, Data transformation, Data mining, Pattern evolution, knowledge



Gambar 2. Alur Penelitian [8]

Berdasarkan gambar 2 tentang alur penelitian maka dapat dijelaskan sebagai berikut :

1. *Data* merupakan sekumpulan data operasional yang diperlu sebelum dilakukan sebelum tahap penggalian informasi.[9][10]
2. *Data Selection* merupakan proses Pembersihan data yang bertujuan untuk menghilangkan data yang tidak memiliki nilai (null), data yang salah input
3. *Data Transformation* dilakukan dengan memberikan inialisasi terhadap data
4. *Data Mining* Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan
5. *Evaluation* Pada tahap evaluasi, akan diketahui apakah hasil daripada tahap data mining dapat menjawab tujuan yang telah ditetapkan

C. Data Mining

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan carayangberbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagipemilik data. “*Data mining* merupakan bidang dari beberapakeilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola[6]statistik, *database*, dan visualisasi untuk penanganan permasalahanpengambilan informasi dari *database* yang besar.”

*mining*didorong oleh beberapa faktor, antara lain :

1. Pertumbuhan yang cepat dalam kumpulan data
2. Penyimpanan data dalam data *warehouse*, sehingga seluruh perusahaanmemiliki akses ke dalam *database* yang baik.
3. Adanya peningkatan akses data melalui *navigasi web* dan *intranet*

D. Rapid Miner

RapidMiner adalah salah satu software untuk pengolahan *data mining*.Pekerjaan yang dilakukan oleh RapidMiner text mining adalah berkisar dengan analisis teks, mengekstrak pola-pola dari data set yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. Tujuan dari analisis teks ini adalah untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah.RapidMiner menyediakan prosedur data mining dan machine learning, di dalamnya termasuk: ETL (extraction, transformation, loading), data preprocessing, visualisasi, modelling dan evaluasi. Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI. Penyajiannya dituliskan dalam bahasa pemrograman Java[16].

III. HASIL DAN PEMBAHASAN

A. Data.

Penelitian ini menggunakan dataset spam dan not spam berasal dari <https://www.kaggle.com/> yang diakses pada tanggal 20 September 2021. Adapun data yang didapat yaitu menggunakan data 5390 data spam dan ham, data dapat dilihat sebagai berikut :

Tabel 1 Data Kagle Spam dan Ham

Jumlah	label	text	label_num
605	ham	Subject: enron methanol ; n	0
2349	ham	Subject: hpl nom for january	0
3624	ham	Subject: neon retreat ho	0
4685	spam	Subject: photoshop , windo	1
2030	ham	Subject: re : indian springstf	0
2949	ham	Subject: ehronline web add	0
2793	ham	Subject: spring savings certi	0
4185	spam	Subject: looking for medical	1
2641	ham	Subject: noms / actual flow	0
1870	ham	Subject: nominations for oct	0
4922	spam	Subject: vocable % rnd - wor	1
3799	spam	Subject: report 01405 lwffur	1
1488	ham	Subject: enron / hpl actuals	0
3948	spam	Subject: vic . odin n ^ owber	1
3418	ham	Subject: tenaska iv julydarre	0
4791	spam	Subject: underpriced issue v	1
2643	ham	Subject: re : first delivery - v	0
3137	ham	Subject: swift - may 2001 vol	0
1629	ham	Subject: meter variances - u	0
1858	ham	Subject: additional recruitin	0
3261	ham	Subject: fw : ercot load com	0
3447	ham	Subject: meter 6461 , concol	0
2459	ham	Subject: hpl nom for january	0

B. Preprocessing Data

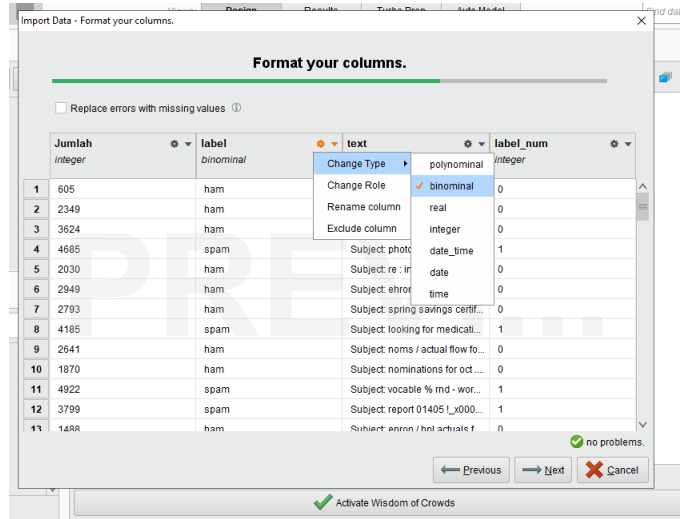
Data Cleaning atau PreProcessing bertujuan untuk melakukan klasifikasi data spam dan ham, maka hasil data cleaning atau preprocessing tersebut dapat di lihat pada tabel berikut ini :

Tabel 2 Data Preprocessing

Jumlah	label	text	label_num
605	ham	Subject: enron methanol ; n	0
2349	ham	Subject: hpl nom for january	0
3624	ham	Subject: neon retreat ho	0
4685	spam	Subject: photoshop , windo	1
2030	ham	Subject: re : indian springstf	0
2949	ham	Subject: ehronline web add	0
2793	ham	Subject: spring savings certi	0
4185	spam	Subject: looking for medical	1
2641	ham	Subject: noms / actual flow	0
1870	ham	Subject: nominations for oct	0
4922	spam	Subject: vocable % rnd - wor	1
3799	spam	Subject: report 01405 lwffur	1
1488	ham	Subject: enron / hpl actuals	0
3948	spam	Subject: vic . odin n ^ owber	1
3418	ham	Subject: tenaska iv julydarre	0
4791	spam	Subject: underpriced issue v	1
2643	ham	Subject: re : first delivery - v	0
3137	ham	Subject: swift - may 2001 vol	0
1629	ham	Subject: meter variances - u	0
1858	ham	Subject: additional recruitin	0
3261	ham	Subject: fw : ercot load com	0
3447	ham	Subject: meter 6461 , concol	0
2459	ham	Subject: hpl nom for january	0

C. Data Transformation

Data transformation dilakukan dengan memberikan inialisasi terhadap data yang memiliki nilai akan disesuaikan type data yang dibutuhkan pada algoritma naïve bayes dan algoritma C.45. Data Tersebut dapat dilihat pada gambar berikut ini :

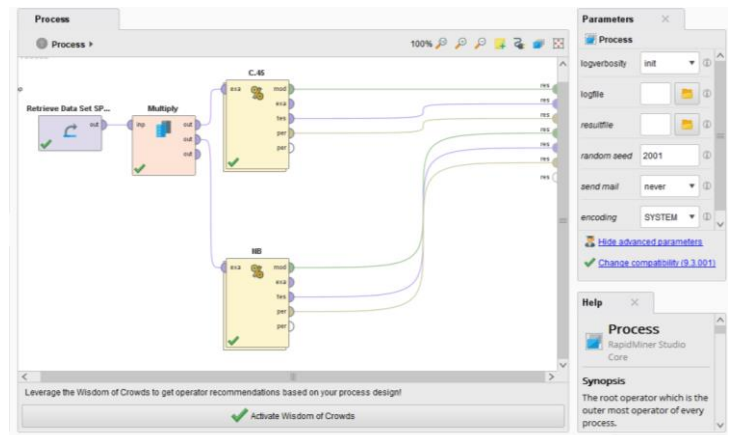


Gambar 2 Data Transformation

D. Model Algoritma

Pernarapan model dalam uji komparasi dua algoritma yaitu algoritma naïve bayes dan algoritma c.45 sebagai berikut :

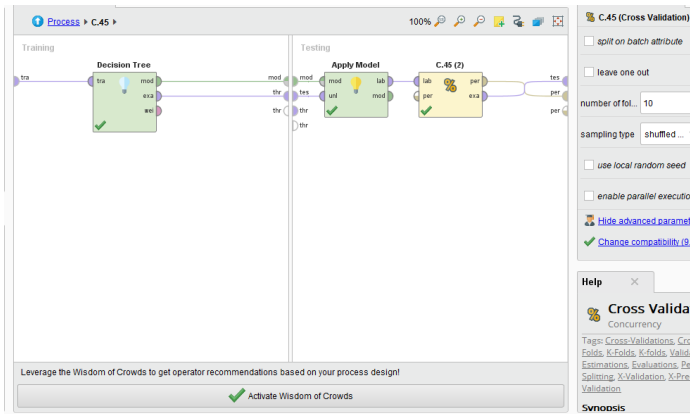
Tahap Pertama



Gambar 3. Model Panggil Data

Berdasarkan gambar diatas menjelaskan bahwa dalam melakukan penerapan uji komparasi dua algoritma antara algoritma naïve bayes dan algoritma C.45 menggunakan tools rapidminer versi 9.3 dengan bantuan operator retrieve data berfungsi untuk memanggil data dalam bentuk file excel maupun file .csv. kemudian operator multiply berfungsi untuk menduplicate data set. Sedangkan croes validation untuk model uji tiap algoritma.

Tahap Kedua



Gambar 4. Model Algoritma C.45

Berdasarkan gambar diatas menjelaskan bahwa dalam menguji komparasi diperlukan model dengan menggunakan opertor decision tree, apply model dan performance.

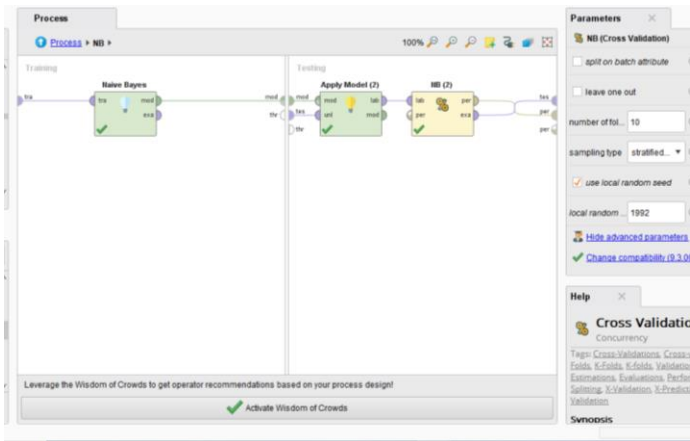
Tahap Ketiga

- a) Prediksi Ham dan true Ham sebanyak 3385 Data
- b) Prediksi Ham dan true Spam sebanyak 166 Data
- c) Prediksi Spam dan true Ham sebanyak 0 Data
- d) Prediksi Spam dan true Spam sebanyak 1447 Data

Hasil klasifikasi Algoritma C.45, Berdasarkan Hasil klasifikasi Algoritma C.45 didapatkan nilai sebagai berikut :

Tabel 3. Hasil Klasifikasi Algoritma C.45

No	Label	Prediksi
1	spam	spam
2	spam	spam
3	spam	ham
4	spam	ham
5	spam	ham
...
4998	ham	ham



Gambar 5 Model Algoritma C.45

Berdasarkan gambar diatas menjelaskan bahwa dalam menguji komparasi diperlukan model dengan menggunakan opertor decision tree, apply model dan performance.

E. Hasil Pola

Hasil pola yang didapat dari uji komparasi algoritma dibahas sebagai berikut :

Hasil Pola Algoritma C.45Berdasarkan hasil pola algoritma C.45 didapatkan nilai akurasi sebagai berikut :

	true ham	true spam	class precision
pred ham	3385	166	95.33%
pred spam	0	1447	100.00%
class recall	100.00%	89.71%	

Gambar 6 Hasil Akurasi Algoritma C.45

Berdasarkan hasil akurasi algoritma C.45 menghasilkan akurasi sebesar 96,68% dengan rincian sebagai berikut :

Hasil Pola Algoritma Naïve Bayes. Berdasarkan hasil pola algoritma Naïve Bayes didapatkan nilai akurasi sebagai berikut :

	true ham	true spam	class precision
pred ham	3385	165	95.30%
pred spam	0	1448	100.00%
class recall	100.00%	89.77%	

Gambar 7. Hasil Akurasi Algoritma Naïve Bayes

Berdasarkan hasil akurasi algoritma Naïve Bayes menghasilkan akurasi sebesar 96,70% dengan rincian sebagai berikut :

- 1) Prediksi Ham dan true Ham sebanyak 3385 Data
- 2) Prediksi Ham dan true Spam sebanyak 165 Data
- 3) Prediksi Spam dan true Ham sebanyak 0 Data
- 4) Prediksi Spam dan true Spam sebanyak 1448 Data

Hasil Error Algoritma Naïve Bayes. Berdasarkan Hasil Error Algoritma Naïve Bayes. didapatkan nilai sebagai berikut :

	true ham	true spam	class precision
pred ham	3385	165	95.30%
pred spam	0	1448	100.00%
class recall	100.00%	89.77%	

Gambar 8. Hasil Error Algoritma Naïve Bayes.

Berdasarkan hasil gambar diatas menjelaskan bahwa nilai error pada algoritma naïve bayes sebesar 3,30%

Hasil klasifikasi Algoritma Naïve Bayes. Berdasarkan Hasil klasifikasi Algoritma naïve bayes didapatkan nilai sebagai berikut :

Tabel 4 Hasil klasifikasi Algoritma Naïve Bayes

No	Label	Prediksi
1	ham	ham
2	ham	ham
3	ham	ham
4	ham	ham
5	ham	ham
...
220	spam	spam

Berdasarkan hasil pembahasan pada uji komparasi antara algoritma c.45 dan algoritma naïve bayes dilihat dari besaran akurasi yang didapat sebagai berikut :

Tabel 5 Hasil Algoritma Terbaik.

No	Algoritma	Hasil Akurasi	Keterangan
1	Algoritma C.45	96.68%	
2	Algoritma Naïve Bayes	96.70%	

Berdasarkan hasil uji komparasi diperoleh hasil algoritma terbaik dengan mengukur tingkat hasil akurasi sehingga dapat diperoleh algoritma C.45 memiliki nilai sebesar 96.68% Kemudian pada penerapan model algoritma naïve bayes menjelaskan bahwa tingkat hasil akurasi dapat diperoleh dari algoritma naïve bayes dengan nilai sebesar 96.79%. bisa di artikan bahwa algoritma naïve bayes data dikategorikan sebagai pedoman pengambilan keputusan.

IV. KESIMPULAN

Berdasarkan Hasil penelitian yang telah dilakukan penulis dapat menarik beberapa kesimpulan sebagai berikut : Berdasarkan hasil akurasi algoritma Naïve Bayes menghasilkan akurasi sebesar 96,70% dengan rincian yaitu Prediksi Ham dan true Ham sebanyak 3385 Data, Prediksi Ham dan true Spam sebanyak 165 Data, Prediksi Spam dan true Ham sebanyak 0 Data, Prediksi Spam dan true Spam sebanyak 1448 Data. Berdasarkan hasil akurasi algoritma C.45 menghasilkan akurasi sebesar 96,68% dengan rincian yaitu Prediksi Ham dan true Ham sebanyak 3385 Data, Prediksi Ham dan true Spam sebanyak 166 Data, Prediksi Spam dan true Ham sebanyak 0 Data Prediksi Spam dan true Spam sebanyak 1447 Data. Berdasarkan hasil uji komparasi diperoleh hasil algoritma terbaik dengan mengukur tingkat hasil akurasi sehingga dapat diperoleh algoritma C.45 memiliki nilai sebesar 96.68% Kemudian pada penerapan model algoritma naïve bayes menjelaskan bahwa tingkat hasil akurasi dapat diperoleh dari algoritma naïve bayes dengan nilai sebesar 96.79%. bisa di artikan bahwa algoritma naïve bayes data dikategorikan sebagai pedoman pengambilan keputusan.

V. DAFTAR PUSTAKA

- [1] m. Abdurohman, r. Husna, i. Ali, g. Dwilestari, and n. Rahaningsih, "penerapan model klasifikasi dalam tingkat kepuasan layanan publik Kelurahan Karyamulya dengan menggunakan algoritma decision tree," *inf. Manag. Educ. Prof. J. Inf. Manag.*, vol. 6, no. 1, p. 81, 2022, doi: 10.51211/imbi.v6i1.1678.
- [2] p. Studi, t. Informatika, p. Studi, s. Informasi, p. Studi, and r. Perangkat, "pengelompokkan hasil belajar siswa dengan metode clustering k-means saeful anwar 1) , tati suprapti 2) , gifthera dwilestari 3) irfan ali 4)," vol. 4, no. 2, pp. 60–72, 2022.
- [3] f. M. Basysyar, g. Dwilestari, a. Bahtiar, martanto, and d. N. Nuris, "market basketball analysis algorithm for determining products association," *iop conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012040, 2021, doi: 10.1088/1757-899x/1088/1/012040.
- [4] f. M. Basysyar, "clustering data disabilitas menggunakan algoritma k-means di kabupaten cirebon," *jursima (jurnal sist. Inf. Dan ...)*, vol. 9, no. 3, 2021.
- [5] s. Suhari, a. Faqih, and f. M. Basysyar, "sistem informasi kepegawaian menggunakan metode agile development di cv. Angkasa raya," *j. Teknol. Dan inf.*, vol. 12, no. 1, pp. 30–45, 2022, doi: 10.34010/jati.v12i1.6622.
- [6] s. M. A. K-means, . "kata kunci : data mining, kualitas, ujian nasional, algoritma k-means,," vol. 10, no. 1, 2022.
- [7] c. L. Rohmat, i. Ali, t. Suprapti, and u. Aryanti, "aplikasi pemesanan online barbershop berbasis android untuk meningkatkan layanan," vol. 4, no. 2, pp. 37–45, 2021.
- [8] y. A. Wijaya, n. Suarna, iin, r. Hamonangan, and r. Nining, "comparison of machine learning algorithm for santander dataset," *iop conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012032, 2021, doi: 10.1088/1757-899x/1088/1/012032.
- [9] n. Suarna, y. A. Wijaya, mulyawan, t. Hartati, and t. Suprapti, "comparison k-medoids algorithm and k-means algorithm for clustering fish cooking menu from fish dataset," *iop conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012034, 2021, doi: 10.1088/1757-899x/1088/1/012034.
- [10] s. Turangga and y. A. W, "analisis internet menggunakan parameter quality of service pada alfamart tuparev 70," vol. 6, no. 1, pp. 392–398, 2022.
- [11] t. Hartati and y. A. Wijaya, "analisis data lalu lintas jaringan di kantor cangehgar cyber operation center menggunakan algoritma k-means network traffic data analysis at cangehgar cyber operation center office using k-means algorithm," vol. 7, no. 1, pp. 75–84, 2022.
- [12] h. Putri, a. I. Purnamasari, a. R. Dikananda, o. Nurdiawan, and s. Anwar, "penerima manfaat bantuan non tunai kartu keluarga sejahtera menggunakan metode naïve bayes dan knn," *build. Informatics, technol. Sci.*, vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.
- [13] a. Z. Zami, o. Nurdiawan, and g. Dwilestari, "klasifikasi kondisi gizi bayi bawah lima tahun pada posyandu melati dengan menggunakan algoritma

- decision tree,” *j. Sist. Komput. Dan inform.*, vol. 3, pp. 305–310, 2022, doi: 10.30865/json.v3i3.3892.
- [14] e. W. Ramadhona, t. Prasetya, and a. I. Purnamasari, “game edukasi ‘ nihongo kurabu ’ belajar bahasa menggunakan unity 2d berbasis android,” *inf. Manag. Educ. Prof.*, vol. 6, no. 1, pp. 71–80, 2022.
- [15] a. Z. Zami, o. Nurdiawan, and g. Dwilestari, “klasifikasi kondisi gizi bayi bawah lima tahun pada posyandu melati dengan menggunakan algoritma decision tree,” *j. Sist. Komput. Dan inform.*, vol. 3, pp. 305–310, 2022, doi: 10.30865/json.v3i3.3892.
- [16] h. Putri, a. I. Purnamasari, a. R. Dikananda, o. Nurdiawan, and s. Anwar, “penerima manfaat bantuan non tunai kartu keluarga sejahtera menggunakan metode naïve bayes dan knn,” *build. Informatics, technol. Sci.*, vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.
- [17] d. A. K. Irfan nurdiyanto, odi nurdiawan, nining rahaningsih, ade irfma purnamasari, “penentuan keputusan pemberian pinjaman kredit menggunakan algoritma c.45,” *j. Data sci. Dan inform.*, vol. 1, no. 1, pp. 16–20, 2021.
- [18] a. Faqih, o. Nurdiawan, and a. Setiawan, “ethnomathematics : utilization of crock , ladle , and chopping board for learning material of geometry at the elementary school,” vol. 4, no. 1, pp. 46–55, 2021.
- [19] o. Nurdiawan, f. A. Pratama, d. A. Kurnia, kaslani, and n. Rahaningsih, “optimization of traveling salesman problem on scheduling tour packages using genetic algorithms,” *j. Phys. Conf. Ser.*, vol. 1477, no. 5, 2020, doi: 10.1088/1742-6596/1477/5/052037.
- [20] f. Arie pratama, k. Kaslani, o. Nurdiawan, n. Rahaningsih, and n. Nurhadiansyah, “learning innovation using the zahir application in improving understanding of accounting materials,” *j. Phys. Conf. Ser.*, vol. 1477, no. 3, pp. 0–6, 2020, doi: 10.1088/1742-6596/1477/3/032018.