

Akurasi Algoritma Klasifikasi pada *Software* Rapidminer dan Weka

Ainurrohmah^{a,*}

^a Universitas Negeri Semarang, Sekaran, Gunungpati, Semarang 50229, Indonesia

* Alamat Surel: aiurrohmah@students.unnes.ac.id

Abstrak

Penelitian ini bertujuan untuk mengukur nilai akurasi dari pemrosesan data menggunakan algoritma-algoritma klasifikasi yang ada pada *software data mining* Rapidminer dan Weka. Nilai akurasi yang diperoleh dari algoritma-algoritma tersebut juga bisa digunakan untuk membandingkan *software data mining* yang lebih akurat antara *software* Rapidminer dan Weka. Nilai akurasi bisa dilakukan dengan menggunakan *Confusion matrix*. Penelitian ini merupakan penelitian kajian pustaka. Metode yang digunakan adalah membaca dan memahami keakuratan algoritma klasifikasi pada Rapidminer dan Weka, serta melihat hasil penelitian sebelumnya mengenai akurasi algoritma klasifikasi pada Rapidminer dan Weka dengan menggunakan data spam. Hasil dari penelitian ini adalah (1) Algoritma klasifikasi yang ada di *software* Rapidminer dan Weka adalah *Decision Tree*, *Random Forest*, K-NN, Naives Bayes, dan MLP. (2) Akurasi yang didapatkan dari satu algoritma dengan algoritma lain atau dari satu aplikasi dengan aplikasi lainnya pasti akan berbeda; Perbedaan ini dikarenakan jumlah data yang dipakai, tipe data yang dipakai dan keterkaitan antar data. (3) Dalam tiga penelitian terdahulu yang sudah dikaji dengan data text spam didapatkan bahwa dua penelitian mengatakan Weka lebih akurat sedangkan satu penelitian mengatakan Rapidminer lebih akurat. Algoritma yang mempunyai nilai akurasi terbaik adalah SVM, K-NN dan MLP.

Kata kunci:

Akurasi, Klasifikasi, Rapidminer, Weka

© 2021 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

1. Pendahuluan

Data mining merupakan proses menggunakan teknik atau metode tertentu untuk menemukan pola atau informasi yang menarik dalam data yang dipilih, menggunakan teknik atau metode tertentu (Hidayanti et al., 2020). *Data mining* merupakan teknik penemuan pola yang tersembunyi dari sebuah data. Pola yang didapatkan akan digunakan untuk memperoleh pengetahuan yang berguna. Banyak fungsi yang dapat dilakukan menggunakan *data mining* di antaranya adalah *classification*, *clustering*, *feature selection* dan *association rule mining* (Lishania et al., 2019). Klasifikasi merupakan pengambilan pola dari informasi yang ada sehingga mendapatkan sebuah aturan yang diinginkan. Klasifikasi mempunyai berbagai macam algoritma yang dapat digunakan serta diterapkan dalam satu data. Dari berbagai macam algoritma yang ada, tentunya ada salah satu algoritma yang memiliki akurasi paling tinggi. Akurasi klasifikasi akan mempengaruhi kinerja klasifikasi (Riyanto, 2019). Pengaruh dalam penilaian akurasi ini dinilai baik dengan data dan juga dengan algoritma yang dipakai. Akurasi dari sebuah algoritma tentunya dapat dihitung baik menggunakan manual seperti *Confusion Matrix* atau dapat dilihat dengan bantuan *software* seperti Rapidminer.

Terdapat beberapa jenis aplikasi atau *Software data mining* terutama pada klasifikasi seperti Rapidminer, Weka, Orange, KNIME, SPSS Climatte dan sebagainya. Aplikasi yang dianggap mirip, mudah digunakan dan tidak perlu bingung dengan penggunaan bahasa pemrograman yang rumit adalah Rapidminer dan Weka. Langkah-langkah keduanya hampir sama yaitu hanya memasukkan data (*import data*) ke dalam aplikasi kemudian memilih algoritma lalu mendapatkan hasilnya, kedua aplikasi bersifat terbuka sehingga dapat digunakan tanpa pembayaran dan mempunyai beberapa algoritma klasifikasi yang sama seperti Naives Bayes, *Random Forest*, *Random Tree*, SVM, *Decision Tree* dan masih banyak lagi.

To cite this article:

Ainurrohmah. (2021). Akurasi Algoritma Klasifikasi pada *Software* Rapidminer dan Weka. *PRISMA, Prosiding Seminar Nasional Matematika 4*, 493-499

Sehingga kedua aplikasi ini dapat kita bandingkan satu sama lain. Perbandingan yang ingin dilihat dalam kajian pustaka ini adalah perbandingan akurasi khususnya algoritma klasifikasi yang sama pada aplikasi Rapidminer dan Weka. Akurasi yang didapatkan dari algoritma yang sama dengan data yang sama belum tentu akan mendapatkan hasil akurasi yang sama pada aplikasi yang berbeda. Perbedaan hasil akurasi ini dipengaruhi oleh faktor tingkat akurasi pada saat data diproses pada aplikasi atau *software* yang digunakan. Sehingga perbandingan akurasi tidak hanya dilakukan untuk menentukan algoritma mana yang lebih unggul tingkat akurasinya. Namun, juga dalam penelitian kajian pustaka ini digunakan untuk melihat aplikasi atau *software* mana yang memiliki tingkat akurasi yang lebih baik. Rumusan Masalah kajian ini antara lain: Bagaimana cara menghitung akurasi pada algoritma klasifikasi?, Apa kegunaan perhitungan akurasi? Algoritma apa saja yang bisa dibandingkan pada Rapidminer dan Weka? Mengapa nilai akurasi bisa berbeda? Bagaimana hasil dari penelitian terdahulu mengenai akurasi dengan menggunakan data spam text? Tujuan: Mengukur nilai akurasi dari pemrosesan data menggunakan algoritma-algoritma klasifikasi yang ada pada *software data mining* Rapidminer dan Weka.

2. Pembahasan

Banyak data yang tersebar di seluruh bidang keilmuan. Data-data yang tersebar itu perlu pengolahan data agar masyarakat awan yang tidak memiliki keilmuan yang berkaitan dengan data tersebut dapat menerima informasi dengan jelas dan tanpa kesalahpahaman. Salah satu bentuk dari pengolahan suatu data yaitu *data mining*. *Data mining* adalah istilah yang digunakan untuk menggambarkan penemuan pengetahuan dalam *database*. Penambangan data juga merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mendeskripsikan dan mengidentifikasi informasi yang berguna (Lishania *et al.*, 2019). *Data mining* adalah disiplin ilmu yang bertujuan untuk menemukan, menambang, atau mengekstrak pengetahuan berdasarkan data atau informasi yang data tersebut miliki (Susanto & Suryadi, 2010). Sedangkan pada buku (Larose, 2005) Penambangan data atau penggalian data adalah proses menjelajahi dan menganalisis data dalam jumlah besar secara otomatis atau semi-otomatis untuk menemukan pola dan aturan yang bermakna. Oleh karena itu, dapat disimpulkan bahwa *data mining* adalah pengolahan data untuk mencari informasi tertentu dan dapat digunakan dalam *big data*.

Data Mining/Knowledge Discovery in Database (KDD) adalah suatu cara yang dilakukan untuk mengolah data dalam jumlah yang besar agar mendapatkan pengetahuan baru. *Data mining* memiliki 4 fungsi dasar yang biasa kita temui yaitu, (1) fungsi klasifikasi, fungsi ini merupakan fungsi untuk menemukan model/fungsi untuk menggambarkan suatu *class* dari data, dengan klasifikasi ini mampu membelajari data sehingga mampu meramalkan kecenderungan data, (2) fungsi prediksi, fungsi ini digunakan untuk menemukan pola data menggunakan variabel dalam memprediksi variabel lain yang belum diketahui jenisnya, (3) fungsi deskripsi, fungsi ini digunakan untuk menemukan karakteristik data, dan (4) fungsi asosiasi, fungsi ini digunakan untuk menemukan hubungan yang ada pada nilai atribut dari banyaknya data (Ayudhitama *et al.*, 2020).

2.1. Klasifikasi

Klasifikasi adalah proses menemukan model atau fungsi yang mendeskripsikan atau membedakan konsep atau kelas data untuk mengevaluasi kelas suatu objek yang labelnya tidak diketahui. Untuk mencapai tujuan ini, proses klasifikasi membuat model yang mampu membedakan data ke dalam kelas yang berbeda berdasarkan aturan atau fungsi tertentu. (Faid *et al.*, 2019). Proses klasifikasi didasarkan pada empat komponen dasar yaitu kelas, prediktor, *training set* dan pengujian *data set*. Model klasifikasi yang paling populer ada berbagai macam seperti pohon keputusan *Decision Trees*, *Naïve Bayes Classifiers*, jaringan syaraf, *Statistical Analysis*, *Genetic Algorithms*, *Rough Sets*, *K-Nearest Neighbor Classifier*, *Rule-based Methods*, *Memory Based Reasoning*, *Support Vector Machines* (Novianti, 2019). Klasifikasi adalah proses menemukan sekumpulan model yang membahas dan membedakan ide dan kelas data, agar model tersebut dapat digunakan untuk menebak kelas mana yang dirahasiakan oleh label atau tidak mempunyai label sehingga kelasnya masih rahasia. Membangun model klasifikasi menggunakan data pelatihan. Model yang dihasilkan diuji dengan menugaskan label kelas ke kumpulan data objek data. Model diwakili sebagai pohon keputusan, aturan klasifikasi dan rumus matematika. Ini untuk mengklasifikasikan objek yang akan datang atau yang tidak diketahui. Tingkat ketepatan adalah

persentase dari set pelatihan independen yaitu sampel uji yang diklasifikasikan dengan benar oleh model (Pujiyanto & Ristanti, 2019). Algoritma klasifikasi yang bisa dipakai pada Weka dan Rapidminer antara lain:

- *Naives Bayes*

Naive Bayes Classifier (NBC) adalah metode teknik klasifikasi dan termasuk dalam pengklasifikasi statistik yang dapat memprediksi probabilitas dari suatu kelas. NBC didasarkan pada teori Bayesian. NBC mengasumsikan bahwa nilai atribut kelas tidak bergantung pada nilai atribut lainnya:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (1)$$

Class C_i adalah nilai terbesar dari sebuah data, sedangkan $P(X)$ adalah konstanta untuk semua class. P merupakan posterior probability (Azis *et al.*, 2020).

- K-NN

Algoritma *K-Nearest Neighbor* adalah pendekatan untuk menemukan kasus dengan menghitung kedekatan antara kasus baru dan lama. Jarak data yang ada ditentukan oleh pengguna yang diwakili oleh k (Azis *et al.*, 2020). Nilai k terbaik untuk algoritma ini bergantung pada datanya. Secara umum, nilai k yang tinggi mengurangi efek noise pada klasifikasi, tetapi membuat batas antara setiap klasifikasi menjadi lebih kabur. Ada banyak cara untuk mengukur kedekatan antara data baru dan data lama (data latih), antara lain jarak Euclidean dan jarak Manhattan (*city distance*), yang paling umum adalah jarak Euclidean (Faid *et al.*, 2019).

- *Random Forest*

Random Forest (hutan acak) adalah sekumpulan pengklasifikasi yang terdiri dari beberapa pohon keputusan dan diklasifikasikan berdasarkan hasil klasifikasi untuk setiap anggota pohon keputusan. Metode hutan acak merupakan perpanjangan dari metode CART, yang terdiri dari penggunaan agregasi *bootstrap* (*bagging*) dan metode pemilihan fitur secara acak. Beberapa pohon tumbuh di hutan acak untuk membentuk hutan, kemudian dilakukan analisis terhadap kelompok pohon tersebut. Dalam kumpulan data yang terdiri dari n variabel observasi dan penjelas, hutan acak diimplementasikan oleh, (1) lakukan n pengambilan sampel acak dengan pemulihan set data. Tahap ini adalah tahap *bootstrap*, (2) dalam contoh *bootstrap*, pohon dibuat hingga mencapai ukuran maksimumnya (tanpa pemangkasan). Pada setiap node, penyortiran dilakukan dengan memilih m variabel penjelas secara acak, di mana $m \ll p$. Penyortir terbaik dipilih dari variabel penjelas ini. Langkah ini adalah langkah pemilihan fitur acak, dan (3) ulangi langkah 1 dan 2 sampai k kali untuk membuat hutan dengan k pohon (Faid *et al.*, 2019).

- *Decision Tree*

Decision Tree (pohon keputusan) adalah algoritma yang paling umum digunakan untuk menyelesaikan masalah klasifikasi. Pohon keputusan terdiri dari beberapa simpul, yaitu akar pohon, simpul dalam, dan daun. Konsep entropi digunakan untuk menentukan atribut pohon mana yang akan dipecah. Pohon keputusan adalah pohon keputusan yang dihasilkan dari ekstraksi data menggunakan algoritma klasifikasi (Faid *et al.*, 2019).

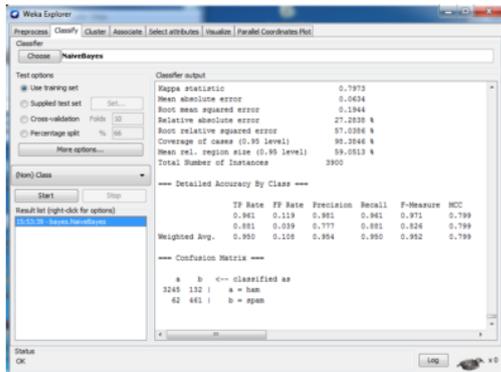
- *Support Vector Machine*

Support Vector Machine (SVM) adalah teknik prediksi untuk klasifikasi dan regresi. *Support Vector Machine* (SVM) memiliki prinsip dasar pengklasifikasi linier, yaitu kasus klasifikasi dapat dipisahkan secara linier, namun kini *Support Vector Machine* (SVM) juga mampu menyelesaikan masalah nonlinier dengan menambahkan konsep kernel pada ruang kerja multidimensi. Pencarian hyperplane akan dilakukan di ruang ini untuk memaksimalkan jarak antar kelas (Faid *et al.*, 2019).

2.2. Weka

Waikato Environment for Knowledge Analysis (WEKA) adalah perangkat lunak pembelajaran mesin populer yang ditulis dalam bahasa pemrograman java. WEKA dikembangkan oleh University of Waikato di Selandia Baru. WEKA mencakup kumpulan algoritma dan visualisasi untuk analisis data dan pemodelan prediktif. Algoritma pembelajaran mesin di WEKA dapat digunakan untuk menyelesaikan masalah di bidang data mining. WEKA menerapkan semua teknik pembelajaran untuk klasifikasi dan regresi, yaitu pohon keputusan, *Support Vector Machines* (SVM), logistik dan linier, *multi layers perceptrons* dan metode *nearest neighbour* (Lishania *et al.*, 2019). Aplikasi WEKA memungkinkan pengguna untuk mengidentifikasi informasi tersembunyi dari *database* dan sistem file dengan pilihan dan

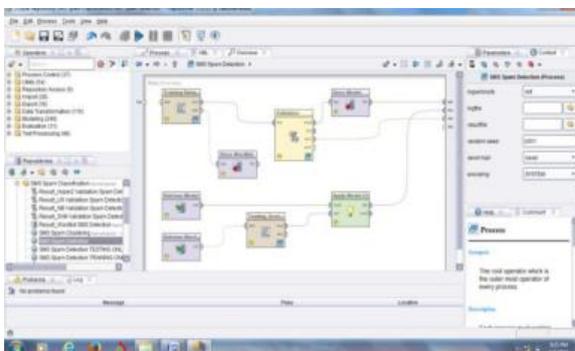
antarmuka *visual* yang mudah digunakan. *Workbench* Weka berisi koleksi alat visualisasi dan algoritma misalnya (C4.5 (C5), ID3, K- *means*, dan Apriori) untuk memecahkan masalah *data mining* dan prediksi secara nyata. (Pujiyanto & Ristanti, 2019). Tampilan aplikasi WEKA yang memperlihatkan hasil dari klasifikasi yang sudah dijalankan setelah memasukkan data atau *import* data dan memilih algoritma lalu menjalankannya dapat dilihat di Gambar 1.



Gambar 1. Interface Weka (Zainal *et al.*, 2015)

2.3. Rapidminer

Rapid Miner adalah *software* pengolah data. Menggunakan prinsip dan *algoritme data mining*, RapidMiner mengekstrak pola dari *big data*, menggabungkan metode statistik, kecerdasan buatan, dan *database*. Rapid Miner memudahkan penggunaannya untuk menghitung data dalam jumlah besar menggunakan operator. Operator ini digunakan untuk mengubah data. Data tersebut digabungkan dengan *node* di operator, dan kemudian Anda tinggal menghubungkannya ke *node* hasil untuk melihat hasilnya (Rachmat *et al.*, 2017). Rapidminer ditulis dalam bahasa pemrograman Java untuk menjalankan tugasnya. Pada dasarnya menggunakan proses XML-File yang dihasilkan oleh pengguna dan berisi urutan tugas yang diwakili oleh operator. Lebih dari 500 operator sudah termasuk dalam Rapidminer. Fungsi operator mencakup aspek utama dari analisis data seperti pemuatan dan transformasi data, *preprocessing* dan visualisasi data, pemodelan dan evaluasi model (Schlitter & Laessig, 2013). Cara menggunakan rapidminer adalah *import* data ke dalam aplikasi terlebih dahulu setelah *data set* terpasang di rapidminer maka bisa sambungkan beberapa operator, misal data ada yang kurang atau hilang bisa kita sambungkan dengan *operator missing*. Setelah data lengkap kita bisa membuat modeling dengan rapidminer atau hal lainnya. Setiap kita mengerjakan sesuatu kita akan dibawa ke halaman *design*, seperti pada Gambar 2 yang memperlihatkan halaman *design* yang sudah disambungkan dari data dan operatornya. Selanjutnya hasil dari halaman *design* yang sudah kita jalankan bisa kita lihat di halaman *page result* (GmbH, 2020).



Gambar 2. Interface Rapidminer (Zainal *et al.*, 2015)

2.4. Confusion Matrix

Confusion matrix merupakan metode evaluasi metode klasifikasi berdasarkan akurasi hasil klasifikasi. Akurasi klasifikasi akan mempengaruhi kinerja klasifikasi. Untuk melakukan analisis *Confusion matrix* dapat digunakan. *Confusion matrix* merupakan matriks prediksi yang akan dibandingkan dengan kategori

asli pada data masukan. (Riyanto, 2019). *Confusion matrix* merupakan salah satu perhitungan untuk menghitung nilai akurasi, presisi, *recall* dan *f-measure* dan *error rate*. *Confusion Matrix* ditampilkan pada Tabel 1.

Tabel 1. Tabel *Confusion Matrix*

Kelas	Terklarifikasi Positif	Terklarifikasi Negatif
Positif	TP (<i>True Positif</i>)	FN (<i>False Negative</i>)
Negatif	FP (<i>False Positif</i>)	TN (<i>True Negative</i>)

Pada kinerja *Confusion Matrix* dapat diukur dengan TP, FN, FP dan TN. TP (*True Positive*) adalah jumlah data bernilai benar yang memang memiliki nilai kebenaran datanya adalah benar. FN (*False Negative*) adalah jumlah data salah yang memang memiliki nilai kebenaran datanya adalah salah. FP (*False Negative*) adalah jumlah data benar yang dianggap sistem memiliki nilai kebenaran datanya adalah salah. TN (*True Negative*) adalah jumlah data salah yang dianggap sistem memiliki nilai kebenaran benar.

- Menghitung menghitung keakuratan sistem mengklasifikasikan data dengan tepat

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (2)$$

- Menghitung seberapa banyak nilai kebenaran (*positive*) dari *dataset* yang memang bernilai benar (*positive*) muncul

$$TPRate (Recall) = \frac{TP}{TP+FN} \quad (3)$$

- Ketepatan nilai kebenaran dalam klasifikasi sesuai dengan nilai kebenaran yang sesungguhnya.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

(Pujiyanto & Ristanti, 2019)

Gambar 3 dibawah ini, merupakan contoh *confusion Matrix* yang didapatkan secara langsung pada aplikasi Weka dan Rapidminer:

accuracy 69.2308 %		accuracy: 69.23%	
=== Confusion Matrix ===			
a b	<-- classified as	true No	true Yes
0 4	a = No	0	0
0 9	b = Yes	4	9
		class recall	0.00% 100.00%

Gambar 3. (a) *Confusion Matrix* Weka (b) *Confusion Matrix* Rapidminer

Sehingga setiap algoritma klasifikasi dapat menggunakan *confusion matrix* sebagai salah satu cara perhitungan akurasi pada algoritmanya. Kegunaan dalam mencari nilai akurasi ini adalah mengetahui nilai akurasi dalam memproses suatu data sehingga didapatkan suatu performa dalam pengolahan data tersebut, kegunaan lainnya untuk mencari performa terbaik dari algoritma yang bisa dipakai. Selain itu, perhitungan akurasi dengan *software* yang ada bertujuan agar melihat dan membandingkan *software* mana yang memiliki rata-rata performa kinerja yang baik untuk melakukan klasifikasi suatu data.

Beberapa hasil penelitian yang sudah membandingkan akurasi algoritma klasifikasi pada Weka dan Rapidminer dengan data *text spam*. Seperti penelitian oleh (Zainal *et al.*, 2015) yang berjudul “*An Analysis of Various Algorithms For Text Spam Classification and Clustering Using RapidMiner and Weka*” diperoleh hasil bahwa akurasi tertinggi pada Rapidminer dan Weka adalah menggunakan algoritma SVM, sedangkan akurasi yang tinggi didapatkan dengan aplikasi Weka. Penelitian lain oleh (Mohd Foozy *et al.*, 2017) dengan judul “*A Comparative Study with RapidMiner and WEKA Tools over some Classification Techniques for SMS Spam*” mendapatkan hasil bahwa akurasi algoritma yang paling unggul adalah KNN dengan rata rata akurasi tertinggi ada pada aplikasi Rapidminer. Dan penelitian lain dari (Hanif *et al.*, 2018) dengan judul “*Performance Evaluation of Machine Learning Algorithms for Spam Profile Detection on Twitter Using WEKA and RapidMiner*” didapatkan bahwa algoritma dengan akurasi terbaik adalah MLP (*Multilayer Perception*) dengan aplikasi yang memiliki akurasi rata rata tertinggi adalah Weka. Dari beberapa penelitian ini didapatkan bahwa akurasi yang didapatkan dari satu

algoritma dengan algoritma atau dari satu aplikasi dengan aplikasi lainnya pasti akan berbeda perbedaan ini dikarenakan jumlah data yang dipakai, tipe data yang dipakai dan keterkaitan antar data. Jumlah data akan mempengaruhi karena semakin banyak data akan semakin baik akurasi, tipe data akan berpengaruh terhadap algoritma yang dipakai dan akurasi algoritmanya, karena algoritma terkadang hanya dapat menggunakan satu tipe data. Dalam tiga penelitian terdahulu yang sudah dikaji dengan data text spam didapatkan bahwa 67% mengatakan Weka lebih akurat sedangkan 33% mengatakan Rapidminer lebih akurat dengan algoritma yang mempunyai nilai akurasi terbaik adalah SVM, KNN dan MLP.

3. Simpulan

Kesimpulan yang dapat diambil dari penelitian pustaka mengenai akurasi algoritma klasifikasi pada Rapidminer dan Weka adalah (1) akurasi Algoritma Klasifikasi pada *software* Rapidminer dan Weka dapat dihitung menggunakan dengan *Confusion Matrix*. *Confusion Matrix* selain dapat melihat akurasi juga dapat menghitung *recall* dan *precision*, (2) perhitungan akurasi baik pada algoritma klasifikasi bertujuan untuk mencari performa terbaik dari algoritma yang bisa dipakai. Selain itu, perhitungan akurasi dengan *software* yang ada bertujuan agar melihat dan membandingkan *software* mana yang memiliki rata-rata performa kinerja yang baik untuk melakukan klasifikasi suatu data, (3) algoritma klasifikasi yang ada di *software* Rapidminer dan Weka adalah *Decision Tree*, *Random Forest*, KNN, Naives Bayes, dan MLP, (4) akurasi yang didapatkan dari satu algoritma dengan algoritma atau dari satu aplikasi dengan aplikasi lainnya pasti akan berbeda perbedaan ini dikarenakan jumlah data yang dipakai, tipe data yang dipakai dan keterkaitan antar data. Jumlah data akan mempengaruhi karena semakin banyak data akan semakin baik akurasi, tipe data akan berpengaruh terhadap algoritma yang dipakai dan akurasi algoritmanya, karena algoritma terkadang hanya dapat menggunakan satu tipe data, dan (5) dalam tiga penelitian terdahulu yang sudah dikaji dengan data text spam didapatkan bahwa dua penelitian mengatakan Weka lebih akurat sedangkan satu penelitian mengatakan Rapidminer lebih akurat dengan algoritma yang mempunyai nilai akurasi terbaik adalah SVM, KNN dan MLP. Saran untuk penelitian selanjutnya dalam melihat keakuratan bisa menggunakan data dengan tipe data yang lain dan data yang lebih besar. Bisa menggunakan berbagai algoritma lagi sehingga tidak hanya algoritma klasifikasi.

Daftar Pustaka

- Ayudhitama, A. P., Pujiyanto, U., Elektro, T., Teknik, F., Malang, U. N., Tree, D., & Network, N. (2020). Analisa 4 Algoritma Dalam Klasifikasi Penyakit Liver. *Jurnal Informatika Polinema*, 6, 1–9.
- Azis, H., Tangguh Admojo, F., & Susanti, E. (2020). Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah. *Techno.Com*, 19(3), 286–294. <https://doi.org/10.33633/tc.v19i3.3646>
- Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool *Data mining* Weka dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, 8(1), 11–16. <https://doi.org/10.34148/teknika.v8i1.95>
- GmbH, R. (2020). *RapidMiner 9 Operator Reference Manual*. www.rapidminer.com
- Hanif, M. H. M., Adewole, K. S., Anuar, N. B., & Kamsin, A. (2018). Performance Evaluation of Machine Learning Algorithms for Spam Profile Detection on Twitter Using WEKA and RapidMiner. *Advanced Science Letters*, 24(2), 1043–1046. <https://doi.org/10.1166/asl.2018.10683>
- Hidayanti, I., Kurniawan, T. B., & Afriyudi, A. (2020). Perbandingan Dan Analisis Metode Klasifikasi Untuk Menentukan Konsentrasi Jurusan. *Jurnal Ilmiah Informatika Global*, 11(1), 16–21. <https://doi.org/10.36982/jig.v11i1.1067>
- Larose, D. T. (2005). *An Introduction to Data mining* (1st ed.). John Wiley & Sons, Inc.
- Lishania, I., Goejantoro, R., & Nasution, Y. N. (2019). Perbandingan Klasifikasi Metode Naive Bayes dan Metode Decision Tree Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda. *Jurnal Eksponensial*, 10(2), 135–142.

- Mohd Foozy, C. F., Ahmad, R., Faizal Abdollah, M. A., & Wen, C. C. (2017). A Comparative Study with RapidMiner and WEKA Tools over some Classification Techniques for SMS Spam. *IOP Conference Series: Materials Science and Engineering*, 226(1). <https://doi.org/10.1088/1757-899X/226/1/012100>
- Novianti, D. (2019). Pengukuran Kualitas E-Commerce Shopee Terhadap Kepuasan Pengguna. *Paradigma: Jurnal Komputer Dan Informatika Universitas Bina Sarana Informatika*, 21(2), 143–148. <https://doi.org/10.31294/p.v20i2>
- Pujianto, U., & Ristanti, P. Y. (2019). Perbandingan kinerja metode C4.5 dan Naive Bayes dalam klasifikasi artikel jurnal PGSD berdasarkan mata pelajaran. *Tekno*, 29(1), 50. <https://doi.org/10.17977/um034v29i1p50-67>
- Rachmat, B., Gafar, A. A., Fajriani, N., Ramdani, U., Uyun, F. R., Purnamasari, Y., & Ransi, N. (2017). *Implementasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan*. *April*, 58–62.
- Riyanto, U. (2019). Analisis Perbandingan Algoritma Naive Bayes Dan Support Vector Machine Dalam Mengklasifikasikan Jumlah Pembaca Artikel Online. *JIKA (Jurnal Informatika)*, 2(2), 62–72. <https://doi.org/10.31000/v2i2.1521>
- Schlitter, N., & Laessig, J. (2013). *Distributed Data Analytics using RapidMiner and BOINC Distributed Data Analytics using RapidMiner and BOINC*. August.
- Susanto, S., & Suryadi, D. (2010). *Pengantar Data mining Menggali Pengetahuan Dari Bongkahan Data* (1st ed.). CV Andi OFFSET.
- Zainal, K., Sulaiman, N. F., & Jali, M. Z. (2015). An Analysis of Various Algorithms For Text Spam Classification and Clustering Using RapidMiner and Weka. *International Journal of Computer Science and Information Security (IJCSIS)*, 13(3), 66–74.