

# Cek Plagiasi

*by* Eva Argarini

---

**Submission date:** 17-Dec-2021 04:34AM (UTC-0500)

**Submission ID:** 1732489695

**File name:** Template\_Penulisan\_Artikel-Jurnal\_Teknoinfo-lates-FIX.docx (211.28K)

**Word count:** 3115

**Character count:** 19893

## PERBANDINGAN 3 ALGORITMA KLASIFIKASI DATA MINING DALAM PRO-KONTRA BAHAYA ROKOK ELEKTRIK

Eva Argarini Pratama<sup>1)</sup>, Corie Mei Hellyana<sup>2)</sup>, Nuzul Imam Fadlilah<sup>3)</sup>

<sup>1</sup>Sistem Informasi, Universitas Bina Sarana Informatika  
<sup>2</sup>Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika  
<sup>3</sup>Teknologi Komputer, Universitas Bina Sarana Informatika  
<sup>1,2,3</sup>Jl. Kramat Raya No. 98 Kwitang, Senen, Jakarta Pusat  
 Email: <sup>1</sup>[eva.eap@bsi.ac.id](mailto:eva.eap@bsi.ac.id), <sup>2</sup>[corie.cma@bsi.ac.id](mailto:corie.cma@bsi.ac.id), <sup>3</sup>[nuzul.nfh@bsi.ac.id](mailto:nuzul.nfh@bsi.ac.id)

### Abstract

Many studies and facts exist in society that smoking is bad for health. This causes smokers to look for alternatives from traditional cigarettes to e-cigarettes (vaping) to help quit smoking slowly. However, recently the use of e-cigarettes has become its own pros and cons in society because of the acceptance of information from the community itself. This study tries to compare the use of 3 data mining classification methods, namely Decision Tree, Naïve Bayes and Logistic Regression to get the most accurate algorithm and find out the results of the classification of dangers or not using e-cigarettes based on the views of the community itself using 4 comparison factors, namely how often. see how people use/offer/advertise/promote e-cigarettes, how often they see shows about the dangers of e-cigarettes, have close friends who smoke vaping, and have family (father/mother/siblings) who smoke. From the results of data processing and testing by measuring the performance of the three algorithms using the confusion matrix procedure, operator cross validation and the ROC curve, the decision tree algorithm produces the highest level of accuracy value of 81.00% and based on the decision tree algorithm graph it can also be seen that you have Close friends who smoke e-cigarettes are the main factors that people think vaping is dangerous.

**Keyword:** e-cigarette, datamining, classification

### Abstrak

Banyak penelitian dan fakta yang ada di masyarakat bahwa merokok itu buruk bagi kesehatan. Hal ini menyebabkan perokok mencari alternatif dari rokok tradisional hingga rokok elektrik (vaping) untuk membantu berhenti merokok secara perlahan. Namun akhir-akhir ini penggunaan rokok elektrik menjadi pro dan kontra tersendiri di masyarakat karena faktor penerimaan informasi dari masyarakat itu sendiri. Penelitian ini mencoba membandingkan penggunaan 3 metode klasifikasi data mining yaitu Decision Tree, Naïve Bayes dan Logistic Regression untuk mendapatkan algoritma yang paling akurat dan mengetahui hasil klasifikasi bahaya atau tidaknya penggunaan rokok elektrik berdasarkan pandangan masyarakat itu sendiri menggunakan 4 faktor pembandingan yaitu seberapa sering. melihat bagaimana orang menggunakan/ menawarkan/ mengiklankan/ mempromosikan rokok elektrik, seberapa sering mereka melihat tayangan tentang bahaya rokok elektrik, memiliki teman dekat yang merokok vape, dan memiliki keluarga (ayah/ibu / saudara kandung) yang merokok. Dari hasil pengolahan data dan pengujian dengan melakukan pengukuran performansi ketiga algoritma tersebut menggunakan prosedur confusion matrix, operator cross validation dan kurva ROC, algoritma decision tree menghasilkan tingkat nilai akurasi yang tertinggi sebesar 81,00% dan berdasarkan grafik algoritma decision tree juga terlihat bahwa Anda memiliki teman dekat yang merokok dengan rokok elektrik menjadi faktor utama seseorang menganggap vaping berbahaya.

**Kata Kunci:** rokok elektrik, datamining, klasifikasi

### 1. Pendahuluan

Sekitar 1,1 miliar orang di dunia berusia 15 tahun ke atas memiliki kebiasaan merokok, dengan 80% perokok adalah orang yang tinggal di negara berkembang atau negara dengan pendapatan rendah dan menengah. Pertumbuhan konsumsi tembakau di negara berkembang menjadi beban negara di bidang kesehatan, ekonomi dan sosial dan ini merupakan masalah yang serius dan ada kecenderungan

meningkat [1]. Pada dasarnya, efek berbahaya dari merokok tembakau sudah diketahui sebelumnya. Penyebab kanker paru-paru adalah 90% karena merokok, karena perokok memiliki kemungkinan empat kali lebih tinggi untuk mengembangkan penyakit ini daripada non-perokok [2].

Dalam beberapa tahun terakhir, rokok elektronik (EC) telah banyak diiklankan sebagai perangkat merokok

alternatif yang diklaim dapat membantu perokok berhenti merokok [3]. E-rokok pertama kali diperkenalkan ke pasar pada tahun 2003, mereka dipromosikan sebagai cara untuk mengurangi konsumsi rokok. Jika seseorang mengidap penyakit paru-paru, tentunya kebiasaan berhenti merokok merupakan salah satu upaya terpenting untuk menjaga kesehatan. Namun, bukti yang mendukung penggunaan rokok elektrik sebagai strategi efektif untuk membantu perokok berhenti merokok masih kurang. Selain itu, dampak penggantian rokok tradisional atau tembakau dengan rokok elektrik masih belum jelas [4].

Faktanya, rokok elektrik tidak membakar tembakau, sehingga dianggap memiliki risiko/ dampak buruk yang jauh lebih rendah daripada rokok tradisional. Selain hanya menghasilkan uap yang secara visual menyerupai asap, untuk alasan yang sama, efek uap inilah yang diduga dapat digunakan sebagai pengganti rokok tembakau. Selain itu, beberapa survei di internet dan uji klinis menunjukkan bahwa rokok elektrik dapat membantu perokok berhenti merokok atau mengurangi bahaya dengan merokok lebih sedikit, tanpa risiko bahaya yang luar biasa bagi pengguna atau pengguna pasif. Efek kesehatan dari pemakaian rokok elektrik dirasa lebih sedikit/ terbatas. Sementara rokok elektrik aerosol mungkin mengandung lebih sedikit racun daripada rokok tradisional biasa, penelitian yang meneliti apakah rokok elektrik lebih tidak berbahaya daripada rokok tradisional tidak meyakinkan. Beberapa bukti menunjukkan bahwa penggunaan rokok elektrik dapat memfasilitasi penghentian merokok, tetapi data definitif masih kurang [5]. Inilah sebabnya mengapa pemakaian rokok elektrik telah memicu perdebatan diantara para profesional kesehatan yang memiliki tujuan yang sama yaitu mengurangi penyakit dan kematian terkait tembakau. Tetapi mereka tidak setuju tentang rokok elektrik yang dianggap membuat masalah penyakit dan kematian menjadi lebih menurun atau lebih meningkat [2].

Dengan permasalahan tersebut, menggunakan pengetahuan datamining yang ada. Penelitian ini mencoba membandingkan penggunaan algoritma yang ada untuk mengklasifikasikan opini tentang berbahaya atau tidaknya penggunaan rokok elektrik bagi masyarakat umum saat ini. Algoritma yang dibandingkan adalah pohon keputusan, nave Bayes dan Logistic Regression. Dimana pada penelitian sebelumnya algoritma pohon keputusan merupakan salah satu algoritma yang paling akurat dalam mengklasifikasikan beberapa masalah seperti memprediksi status pembayaran non finansial terburuk pada kredit komersial dimana pohon keputusan memiliki akurasi yang lebih baik daripada regresi logistik [6]. Pada kasus perbandingan naive bayes dan pohon keputusan pada seleksi fitur menggunakan algoritma genetika untuk masalah klasifikasi, juga menunjukkan bahwa pohon keputusan memiliki akurasi yang sedikit lebih baik daripada Naïve Bayes. Namun pada penelitian yang membandingkan penggunaan nave bayes, decision tree dan k-Nearest Neighbor dalam menemukan desain alternatif pada alat simulasi energi, menunjukkan bahwa

pada penelitian ini, Naïve Bayes mengungguli Decision Tree dan kNearest Neighbor [7].

Perbandingan penggunaan algoritma naïve Bayes dan regresi logistik pada Teori, Implementasi, dan Validasi Eksperimental dengan hasil yang menunjukkan bahwa Regresi Logistik dengan teknik gradient climbing dapat mengungguli classifier dari Naïve Bayes umum. Namun, dengan asumsi pengklasifikasi Gaussian Naive Bayes, baik pengklasifikasi Naïve Bayes maupun Regresi Logistik memiliki kinerja yang sama [8]. Dengan adanya perbedaan hasil perbandingan tersebut, penelitian ini juga perlu membandingkan algoritma klasifikasi mana yang paling baik diterapkan dalam meningkatkan akurasi pengklasifikasian opini tentang berbahaya atau tidaknya penggunaan rokok elektrik.

## 22 2. Metode Penelitian

Metode penelitian yang dilakukan pada penelitian ini yaitu penelitian eksperimen, dan memiliki beberapa tahapan yaitu:

### 2.1. Metode Pengumpulan Data

Langkah awal adalah pengumpulan data kuisisioner terkait rokok elektrik. Dataset berfokus pada opini yang membahas hal-hal terkait aspek pendukung utama dalam pengenalan rokok elektrik hingga penggunaan rokok elektrik. Tahap selanjutnya adalah proses pemberian label pada data melalui penyematan status tweet yang dilihat dari sentimen negatif dan sentimen positif. Melalui cara manual (memberikan informasi nilai untuk setiap kuisisioner yang diisi) proses pelabelan dilakukan.

Pada tahap awal, data mentah dikumpulkan dengan menggunakan pertanyaan dan pernyataan dalam kuisisioner sebagai variabel yang diolah, yaitu Apakah Anda pernah melihat kesan orang menggunakan/ menawarkan/ mengiklankan/ mempromosikan rokok elektrik/ vape? Pernahkah Anda melihat tayangan tentang bahaya rokok elektrik? Saya punya teman dekat perokok elektrik, saya punya keluarga (ayah/ ibu/ kakak) perokok elektrik. Pertanyaan dan pernyataan dalam kuisisioner diambil sebagai data dari beberapa responden yang tersebar di beberapa kota besar di Indonesia seperti Jakarta, Medan, Semarang, Surabaya, dan lain-lain. Data yang terkumpul disimpan dalam format file excel. Dataset yang telah dikumpulkan dari kuisisioner ini adalah data unsupervised. Untuk diolah menggunakan teknik supervised learning, data angket yang telah dikumpulkan sebelumnya perlu dibuatkan atau diberi label, proses pelabelan dilakukan secara manual dengan memberikan status pada setiap pertanyaan. Sukai label A untuk pertanyaannya Pernahkah Anda melihat kesan orang menggunakan/ menawarkan/ mengiklankan/ mempromosikan rokok elektrik/ vape? Dengan jawaban Ya dan Tidak, beri label B untuk pertanyaan ?, Pernahkah Anda melihat kesan tentang bahaya rokok elektrik? Dengan tidak pernah, sangat jarang, jarang dan sering, label C untuk pernyataan saya

memiliki teman dekat yang merokok vape/ e-rokok dengan Ya dan Tidak, dan label D dengan pernyataan saya memiliki keluarga (ayah/ ibu/ saudara) yang merokok elektrik dengan Ya dan Tidak.

**2.2. Cross Validation**

Dalam memvalidasi atau menilai keakuratan dalam suatu model yang ada pada suatu dataset dapat menggunakan suatu teknik yang disebut dengan dengan cross validation. Validasi juga dilakukan sebagai tes standar untuk memprediksi tingkat kesalahan. Ada 3 langkah utama dalam validasi croos yaitu:

1. Cadangan sebagian dari kumpulan data sampel.
2. Menggunakan sisa data-set melatih model.
3. Uji model menggunakan bagian cadangan dari kumpulan data.

Pengolahan data awal dilakukan pada penggunaan RapidMiner, terdapat operator *cross validation* yang merupakan suatu operator yang ada pada rapidminer dan memiliki 2 subproses: yang pertama adalah subproses *training* dimana subproses ini dapat dipakai dalam mencoba atau melatih model yang sudah dibuat dan yang kedua subproses *testing* yaitu subproses yang digunakan sebagai pengujian dan pengukuran dari kinerja model tersebut, untuk memberikan nilai k (jumlah iterasi) maka pada operator *cross validation* terdapat parameter yang dapat digunakan yaitu *Number of fold*, dengan penggunaan operator ini terdapat *sampling type* digunakan untuk memilih teknik *sampling* yang dapat membagi dataset [9]. Standar deviasi yang merupakan hasil perolehan dari pengukuran dalam penyebaran data yang menggambarkan atau menunjukkan jarak rata-rata dari nilai tengah menuju suatu titik nilai tertentu akan dihasilkan berdasarkan kinerja model yang menggunakan cross validation dalam penentuan akurasinya. Hal ini dapat diartikan semakin besar standar deviasi yang dihasilkan, maka penyebaran dari nilai tengahnya juga akan semakin besar, dan begitu juga sebaliknya.

**2.3. Confusion Matrix**

Terdapat suatu konsep dalam datamining dimana penghitungan akurasi dapat menggunakan salah satu metode yaitu confusion matrix, dimana pada confusion matrix ini perhitungan dapat menghasilkan 4 keluaran yaitu akurasi, recall, precision, dan error rate. Hasil evaluasi model klasifikasi didasarkan pada pengujian untuk memperkirakan objek benar dan salah [10]. Pada tahap confusion matrix kinerja dapat diukur dengan TP, TN, FP dan FN, seperti gambaran dan rumusan berikut [11]:

Tabel 1. Komposisi Tabel Hasil Confusion Matrix

Kelas	Terklarifikasi Positif	Terklarifikasi Negatif
<b>Positif</b>	True Positif (TP)	False Negative (FN)

Negatif	False Positif (FP)	True Negative (TN)
---------	--------------------	--------------------

True Positive (TP) merupakan jumlah dari data bernilai benar dan memiliki nilai kebenaran datanya adalah benar. False Negative (FN) adalah jumlah dari data salah dan memiliki nilai kebenaran datanya adalah salah. False Negative (FN) adalah jumlah dari data benar dan dianggap oleh sistem memiliki nilai kebenaran datanya adalah salah. True Negative (TN) adalah jumlah dari data salah dan dianggap oleh sistem memiliki nilai kebenarannya benar.

Adapun rumusan dalam menghitung keakuratan dalam mengklasifikasikan data adalah:

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

**2.4. ROC Curve**

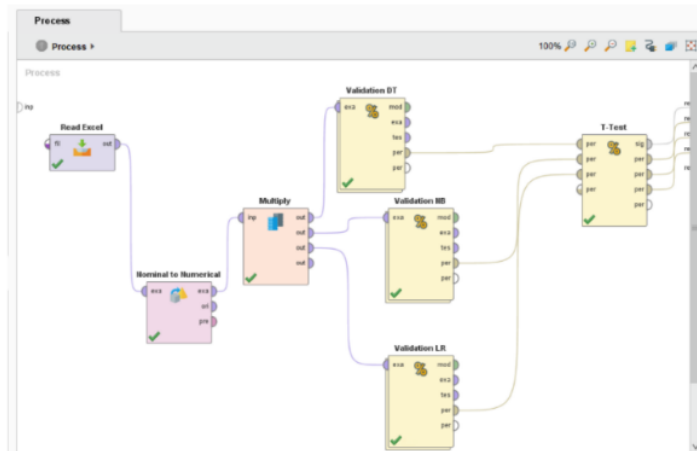
ROC Curve merupakan salah satu prosedur analisis model klasifikasi yang sudah ditentukan sebelumnya. Salah satu manfaat dari ROC curve adalah parameter model yang diinginkan berdasarkan sifat kategoris modelnya. Tingkat akurasi, skalabilitas, kecepatan dan interpretabilitas merupakan beberapa kriteria dari penggunaan metode [12].

**2.5. Validasi**

Kinerja dari suatu metode perlu diuji atau dievaluasi secara sistematis [13] dan evaluasi klasifikasi dilakukan berdasarkan pada pengujian diobjek benar dan salah sehingga menghasilkan suatu validasi data yang dapat digunakan untuk menentukan jenis skema pembelajaran atau hasil yang terbaik, berdasarkan data pelatihan untuk melatih rencana pembelajaran untuk memaksimalkan penggunaan data.

**3. Hasil dan Pembahasan**

Dalam melakukan perbandingan 3 algoritma klasifikasi ini, terdapat beberapa tahapan, yang diawali dengan tahap modeling. Pada tahap ini, pemilihan teknik datamining dilakukan dengan menentukan algoritma yang akan digunakan dalam penelitian dengan menggunakan sebuah tools yaitu Rapid Miner versi 9.5. Adapun hasil pengujian model yang telah dilakukan dengan menggunakan tools tersebut adalah mengklasifikasikan rokok elektrik aman/tidak berbahaya bagi kesehatan dan rokok elektrik memang berbahaya bagi kesehatan menggunakan algoritma Decision Tree, Naïve Bayes dan Logistic Regression untuk mendapatkan hasil nilai akurasi tertinggi atau terbaik. Berikut ini adalah desain model Rapidminer yang digunakan:

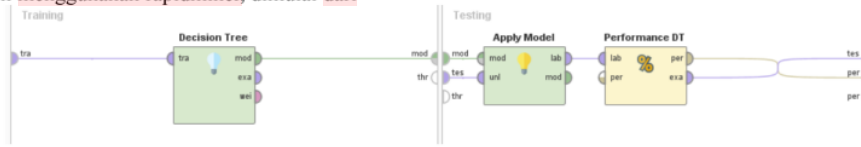


Gambar 1. Perbandingan Desain Model Algoritma Pohon Keputusan, Naïve Bayes dan Regresi Logistik

Hasil akurasi yang diperoleh akan dipengaruhi oleh setting dan penggunaan operator maupun parameter yang dipakai pada framework rapid miner, selain itu model yang terbentuk pun juga akan terpengaruh oleh hal tersebut, sebagai gambaran adalah penggunaan bantuan multi operator dalam melakukan perbandingan.

memasukkan data kemudian setting role set yang kemudian menentukan label yang ada kemudian memproses dokumen yang berisi seperti pada Gambar 1. masing-masing algoritma adalah seperti contoh gambar dibawah untuk salah satu algoritma yaitu Decision Tree :

Gambar diatas juga menggambarkan model pengujian algoritma Decision Tree, Naïve Bayes dan Logistic Regression menggunakan rapidminer, dimulai dari



Gambar 2. Contoh Desain Proses 10-Fold Cross Validation untuk salah satu algoritma yaitu Decision Tree

Gambar 2 menjelaskan desain proses pada operator Decision Tree validasi silang pada Gambar 1. Pengujian ini dilakukan dengan mengambil data dari operator read excel dan data ini merupakan data bersih yang sudah melalui preprocessing. Mengolah dokumen dari file untuk mengubah file menjadi dokumen. Proses validasi terdiri dari data latih dan data uji. Kemudian masuk ke model algoritma pohon keputusan yang didalamnya terdapat perhitungan algoritma, kemudian model diterapkan, setelah itu masuk evaluasi kinerja, kemudian muncul nilai akurasi dan AUC.

sebelumnya telah dilakukan tergambar hasil nilai Accuracy (Confusion Matrix) nya adalah sebagai berikut:

Tabel 2. Nilai Akurasi Algoritma Decision Tree  
accuracy: 81.00% +/- 6.30% (micro average: 81.00%)

	True Yes	True No	Class Precision
Pred. Yes	229	48	82.67%
Pred. No	9	14	60.87%
Class recall	96.22%	22.58%	

Tahap selanjutnya yaitu evaluasi model. Pada tahap evaluasi ini akan diketahui model yang sudah berhasil dibuat pada langkah sebelumnya akan memiliki nilai utilitas sebanyak apa. Validasi silang 10 kali lipat digunakan untuk evaluasi. Dari hasil pengujian yang telah dilakukan pada model algoritma decision tree yang digunakan menghasilkan nilai Accuracy (Confusion Matrix). Setelahnnya adalah menghitung akurasi dari masing-masing algoritma. Dari hasil evaluasi model dengan menggunakan algoritma Decision Tree yang

Dari tabel 2 akurasi model algoritma dapat dihitung dengan menggunakan persamaan berikut:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\
 &= \frac{229+14}{229+48+9+14} \\
 &= \frac{243}{300} \\
 &= 0,81
 \end{aligned}$$

Jumlah True Yes (TP) sebanyak 229 record yang tergolong aman dan False Not yes (FN) sebanyak 9 record yang tergolong False unsafe. 48 catatan Benar Tidak berikutnya diklasifikasikan sebagai tidak aman dan 14 catatan Tidak Benar diklasifikasikan sebagai Tidak Aman. Berdasarkan tabel 2 diatas menunjukkan bahwa tingkat akurasi menggunakan algoritma Decision Tree adalah 81,00%.

Perhitungan untuk akurasi algoritma Naïve Bayes, didapatkan bahwa dengan algoritma Naïve Bayes nilai Accuracy (Confusion Matrix) yang diperoleh adalah:

Tabel 3. Nilai Akurasi Algoritma Naïve Bayes accuracy: 73,33% +/- 9.81% (micro average: 73.33%)

	True Yes	True No	Class Precision
Pred. Yes	195	37	84.05%
Pred. No	43	25	36.76%
Class recall	96.22%	22.58%	

Dari tabel 3 akurasi model algoritma dapat dihitung dengan menggunakan persamaan berikut:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\
 &= \frac{195+25}{195+37+43+25} \\
 &= \frac{243}{300} \\
 &= 0,7333
 \end{aligned}$$

Jumlah True Yes (TP) sebanyak 195 record yang tergolong aman dan False Not yes (FN) sebanyak 43 record yang tergolong False unsafe. 37 catatan Benar Tidak berikutnya diklasifikasikan sebagai tidak aman dan 25 catatan Benar Tidak diklasifikasikan sebagai Tidak Aman. Pada tabel 3 di atas memperlihatkan jumlah akurasi dari penggunaan algoritma Naïve Bayes adalah 73,33%.

Yang terakhir adalah penilaian akurasi dari algoritma Logistic Regression. Berdasarkan hasil pengujian model sebelumnya yang menggunakan algoritma Logistic Regression dapat terlihat hasil nilai Accuracy (Confusion Matrix) sebagai berikut:

Tabel 4. Nilai Akurasi Algoritma Regresi Logistik accuracy: 79,67% +/- 8.53% (micro average: 79.67%)

	True Yes	True No	Class Precision
Pred. Yes	223	46	82.90%
Pred. No	15	16	51.61%
Class recall	93.70%	25.81%	

Dari table 4 akurasi model algoritma Naïve Bayes dapat dihitung dengan menggunakan persamaan berikut:

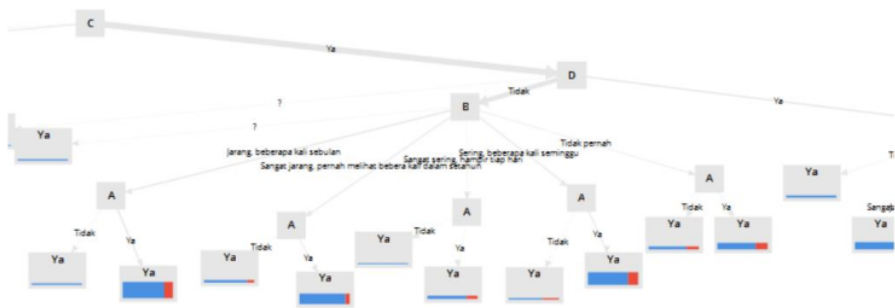
$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\
 &= \frac{223+16}{223+46+15+16} \\
 &= \frac{239}{300} \\
 &= 0,7966
 \end{aligned}$$

Jumlah True Yes (TP) sebanyak 223 record yang tergolong aman dan False Not yes (FN) sebanyak 15 record yang tergolong False unsafe. Selanjutnya 46 Benar Tidak diklasifikasikan sebagai tidak aman dan 16 catatan Benar Tidak diklasifikasikan sebagai Tidak Aman. Berdasarkan tabel 4 di atas menunjukkan bahwa tingkat akurasi menggunakan algoritma Regresi Logistik adalah 79,67%. Tahapan terakhir yaitu membandingkan hasil akurasi dari masing-masing algoritma. Berdasarkan hasil olah analisis dari 3 algoritma yang sudah dilakukan sebelumnya, maka dapat dirangkum hasilnya seperti pada tabel berikut ini:

Tabel 5. Hasil dari Perbandingan Performa ke-3 Algoritma

	Accuracy	AUC
<b>Decision Tree</b>	81,00%	0.628
<b>Naïve Bayes</b>	73.33%	0.679
<b>Logistic Regression</b>	79.67%	0.663

Dapat dilihat pada tabel bahwa akurasi tertinggi terdapat pada algoritma Decision Tree yang menghasilkan grafik seperti berikut:



Gambar 3. Grafik Hasil Decision Tree

#### 4. Kesimpulan

algoritma decision tree menghasilkan tingkat nilai akurasi yang tertinggi dibandingkan algoritma Naïve Bayes dan Logistic Regression dalam mengklasifikasi opini bahaya penggunaan rokok elektrik dengan akurasi 81,00% dan nilai AUC 0,628. Grafik hasil pengolahan data menggunakan algoritma Decision Tree menunjukkan bahwa pendapat bahwa rokok elektrik memang berbahaya. Dan faktor terbesar yang mempengaruhi pendapat tersebut adalah dari variabel C yaitu seseorang yang memiliki teman dekat yang merokok vape/rokok elektrik. Dengan hasil akurasi yang tinggi tersebut, penerapan algoritma Decision Tree pada penelitian ini memiliki akurasi yang lebih tinggi sehingga dapat digunakan untuk memberikan solusi atas masalah pro dan kontra apakah penggunaan rokok elektrik berbahaya atau tidak.

#### Daftar Pustaka

- [1] World Heald Organization, *Who Global Report on Trends in Prevalence of Tobacco Smoking 2000-2025, Second Edition*. 2018.
- [2] A. Sameera, R. K. Patel, K. Bharadwaj, M. M. A. Mir, and S. T. Ahmed, "Pros and Cons of e-cigarettes- A brief note," *J. Chem. , Biol. Phys. Sci.*, vol. 5, no. 4, pp. 4083–4088, 2015.
- [3] G. Heydari, A. E. Ahmady, F. Chamyani, M. Masjedi, and L. Fadaizadeh, "Electronic Cigarette, Effective or Harmful for Quitting Smoking and Respiratory Health: A Quantitative Review Papers," *Lung India*, vol. 34, no. 1, pp. 25–28, 2017, doi: 10.4103/0970-2113.197119.
- [4] D. Leader, "The Pros and Cons of Vaping Are They a Safer Alternative For People Living with COPD?," 2020. <https://www.verywellhealth.com/the-pros-and-cons-of-e-cigarettes-915015>.
- [5] P. Callahan-Lyon, "Electronic Cigarettes: Human Health Effects," *Tob. Control*, vol. 23, no. SUPPL. 2, 2014, doi: 10.1136/tobaccocontrol-2013-051470.
- [6] J. M. Rudd and J. Lewis Priestley, "A Comparison of Decision Tree with Logistic Regression Model for Prediction of Worst Non-Financial Payment Status in Commercial Credit," *Grey Lit. from PhD Candidates*, vol. 5, no. January, 2017, [Online]. Available: <http://digitalcommons.kennesaw.edu/dataphdgreylit/5>.
- [7] A. Ashari, I. Paryudi, and A. M. Tjoa, "Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool," *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 11, pp. 33–39, 2013.
- [8] T. K. Bhowmik, "Naive bayes vs logistic regression: Theory, implementation and experimental validation," *Intel. Artif.*, vol. 18, no. 56, pp. 14–30, 2015, doi: 10.4114/ia.v18i56.1113.
- [9] K. S. Nugroho, "Validasi Model Klasifikasi Machine Learning pada RapidMiner," <https://ksnugroho.medium.com/>, 2020. <https://ksnugroho.medium.com/validasi-model-machine-learning-pada-rapidminer-50be0080df14>.
- [10] X. Wu and V. Kumar, *The Top Ten Algorithms in Data Mining, Chapman & Hall/CRC Data Mining and Knowledge Discovery*, vol. 53, no. 9. Boca Raton: CRC Press, 2009.
- [11] Ainurrohmah, "Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka," in *Prisma*, 2021, vol. 4, pp. 493–499, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>.
- [12] C. Vercellis, *Business Intelligence: Data Mining and Optimization for Decision Making*. West Sussex: John Wiley & Sons Ltd Registered, 2009.
- [13] J. Han, M. Kamber, and J. Pei, *Data mining Concepts and Techniques (Third Edition)*. Waltham: Morga Kaufmann Publishers, 2015.

# Cek Plagiasi

## ORIGINALITY REPORT

24%

SIMILARITY INDEX

22%

INTERNET SOURCES

11%

PUBLICATIONS

11%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://ejournal.nusamandiri.ac.id">ejournal.nusamandiri.ac.id</a> Internet Source	4%
2	<a href="https://medium.com">medium.com</a> Internet Source	2%
3	Submitted to Universitas Bunda Mulia Student Paper	2%
4	<a href="http://123dok.com">123dok.com</a> Internet Source	1%
5	<a href="http://amikjtc.com">amikjtc.com</a> Internet Source	1%
6	<a href="http://ejournal.bsi.ac.id">ejournal.bsi.ac.id</a> Internet Source	1%
7	Submitted to Grand Canyon University Student Paper	1%
8	<a href="http://jurnal.kaputama.ac.id">jurnal.kaputama.ac.id</a> Internet Source	1%
9	Riska Aryanti, Atang Saepudin, Eka Fitriani, Rifky Permana, Dede Firmansyah Saefudin. "Komparasi Algoritma Naive Bayes Dengan	1%

# Algoritma Genetika Pada Analisis Sentimen Pengguna Busway", Jurnal Teknik Komputer, 2019

Publication

---

10	<a href="https://digitalcommons.kennesaw.edu">digitalcommons.kennesaw.edu</a> Internet Source	1 %
11	Submitted to Pennsylvania State System of Higher Education Student Paper	1 %
12	<a href="https://sistemasi.ftik.unisi.ac.id">sistemasi.ftik.unisi.ac.id</a> Internet Source	1 %
13	Submitted to Delaware Center for Distance Adult Learning, Inc. Student Paper	1 %
14	<a href="https://www.researchgate.net">www.researchgate.net</a> Internet Source	1 %
15	<a href="https://www.thieme-connect.com">www.thieme-connect.com</a> Internet Source	1 %
16	<a href="https://eprints.umm.ac.id">eprints.umm.ac.id</a> Internet Source	1 %
17	Submitted to University of Southampton Student Paper	1 %
18	<a href="https://serwiss.bib.hs-hannover.de">serwiss.bib.hs-hannover.de</a> Internet Source	1 %
19	<a href="https://journal.ubm.ac.id">journal.ubm.ac.id</a> Internet Source	1 %

---

<1 %

20

[repository.its.ac.id](https://repository.its.ac.id)

Internet Source

<1 %

21

Ryo Kusnadi, Yusuf Yusuf, Andriantony Andriantony, Richard Ardian Yaputra, Melna Caintan. "ANALISIS SENTIMEN TERHADAP GAME GENSHIN IMPACT MENGGUNAKAN BERT", *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, 2021

Publication

<1 %

22

[ejournal-binainsani.ac.id](https://ejournal-binainsani.ac.id)

Internet Source

<1 %

23

Islamiyah, Anisa Nur Afiyah, Nataniel Dengen, Medi Taruk. "Comparison Performance of C4.5, Naïve Bayes and K-Nearest Neighbor in Determination Drug Rehabilitation", 2019 5th International Conference on Science in Information Technology (ICSITech), 2019

Publication

<1 %

24

Warjiyono Warjiyono, Fandhilah Fandhilah, Amin Nur Rais, Ahmad Ishaq. "Metode FAST & Framework PIECES : Analisis & Desain Sistem Informasi Penjualan Berbasis Website", *Indonesian Journal on Software Engineering (IJSE)*, 2020

Publication

<1 %

25	<a href="http://ejournal.unitomo.ac.id">ejournal.unitomo.ac.id</a> Internet Source	<1 %
26	<a href="http://ikbalburhan.wordpress.com">ikbalburhan.wordpress.com</a> Internet Source	<1 %
27	<a href="http://kamu-harus-sehat.blogspot.com">kamu-harus-sehat.blogspot.com</a> Internet Source	<1 %
28	<a href="http://www.nap.edu">www.nap.edu</a> Internet Source	<1 %
29	Muhammad Rizki Fahdia, Dwiza Riana, Fachri Amsury, Irwansyah Saputra, Nanang Ruhyana. "Komparasi Algoritma Klasifikasi untuk Orientasi Minat Mahasiswa dalam Penuntasan Studi", JIRA: Jurnal Inovasi dan Riset Akademik, 2021 Publication	<1 %
30	R A Saputra, C Agustina, D Puspitasari, R Ramanda, Warjiyono, D Pribadi, Lisnawanty, K Indriani. "Detecting Alzheimer's Disease by The Decision Tree Methods Based On Particle Swarm Optimization", Journal of Physics: Conference Series, 2020 Publication	<1 %
31	<a href="http://journal.uinjkt.ac.id">journal.uinjkt.ac.id</a> Internet Source	<1 %
32	<a href="http://repository.bsi.ac.id">repository.bsi.ac.id</a> Internet Source	<1 %

---

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off