

Penerapan Algoritma C4.5 Untuk Memprediksi Blog Yang Memiliki Peluang Juara

Rita Wati^{1*}, Kusrini², Kusnawi³

^{1,3}Program Studi Teknik Informatika, Universitas Amikom Yogyakarta

²Program Studi Sistem Informasi, Universitas Amikom Yogyakarta

*ritawati@students.amikom.ac.id

Abstrak

Blog saat ini tidak hanya digunakan sebagai buku harian online atau jurnal online saja. Konten blog yang mendukung teks, gambar, video, gif, animasi, pdf dan youtube membuat pemanfaatan blog lebih dari sekedar dairy online belaka. Saat ini blog termasuk salah satu media yang dimanfaatkan oleh perusahaan atau organisasi dalam mempromosi barang dan jasa. Keberadaan blogger sangat dibutuhkan oleh perusahaan, organisasi untuk mereview produk barang maupun jasa yang dimilikinya dengan tujuan menarik pelanggan online. Salah satu upaya untuk mengajak blogger agar dapat mereview produk dengan menarik maka pelaku usaha mulai tertarik mengadakan lomba blog. Dalam penelitian ini sebuah model dibangun untuk memprediksi blog yang memiliki peluang juara dengan menerapkan algoritma pohon keputusan C4.5. Model prediksi dibuat dengan menggunakan 100 data peserta lomba blog yang diperoleh dari tiga kompetisi blog yang diselenggarakan oleh ASUS yang diperoleh melalui link tautan peserta lomba yang terdapat pada blog panitia. Dari dataset yang diolah digunakan tujuh variabel pendukung diantaranya, Jumlah Kata, DA, PA, Gambar, Template, Domain dan Juara. Model prediksi yang dihasilkan berupa pohon keputusan dengan 7 atribut yang menghasilkan 11 daun dan 18 pohon dengan akurasi sebesar 74% precision sebesar 0,735 dan recall sebesar 0,740.

Kata Kunci: Data Mining, Prediksi Juara, Blog, Algoritma C.45, Weka

Abstract

Blogs are currently not only used as online diaries or online journals. Blog content that supports text, images, videos, gifs, animations, pdfs and YouTube makes the use of blogs more than just online dairy. Currently blog is one of the media used by companies and organizations in promoting goods and services. The existence of bloggers is needed by companies and organizations to review their products and services with the aim of attracting online customers. One of the efforts to invite bloggers to be able to review products in an interesting way has led business actors to become interested in holding blog contests. In this research, a model was built to predict blogs that have a chance to win by applying the C4.5 decision tree algorithm. The prediction model was created using 100 blog contest participant data obtained from three blog competitions organized by ASUS which were obtained through the contest participant links found on the committee blog. From the processed dataset, supporting seven variables including Word Count, DA, PA, Image Template, Domain and Champion. The resulting prediction model is a decision tree with 7 attributes which produces 11 leaves and 18 trees with an accuracy of 74% with a precision of 0.735 and a recall of 0.740.

Keywords: Data Mining, Champion Prediction, Blog, C4.5 Algorithm, Weka

1. Pendahuluan

Menurut survei yang telah dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) dikutip dan di-publish oleh data

indonesia.id menyatakan bahwa jumlah pengguna internet di Indonesia lebih dari 210 juta orang. Hal ini membuat masyarakat tertarik memanfaatkan inovasi produk distruksi digital

untuk memancing dunia usaha, organisasi maupun perorangan memilih cara dengan mempromosikan produk barang maupun jasanya melalui internet dan media sosial [1].

Salah satu upaya agar produk, barang dan jasa mudah ditemukan dalam mesin telusur (*search engine*) yaitu dengan cara memposting artikel atau tulisan kedalam situs web atau blog [2]. Hampir sebagian besar perusahaan menggunakan web sebagai salah satu media promosi akan tetapi jika hanya memiliki satu web atau blog, maka yang akan muncul dalam mesin telusur juga hanya satu web atau blog saja, hal demikian menjadi kurang maksimal jika digunakan sebagai media promosi. Untuk menyiasati hal tersebut maka pelaku usaha saat ini mulai melirik keberadaan *blogger* untuk mereview produk barang atau jasa mereka yang dikemas dalam kegiatan lomba blog (*blog competition*) [3].

Kompetisi blog adalah salah satu cara yang paling berpengaruh dan efektif untuk menarik para *blogger* agar ikut berpartisipasi mereview produk, jasa yang dimiliki oleh perusahaan atau organisasi dengan tulisan dan gambar pendukung yang menarik dengan imbalan hadiah. Jika terdapat 100 *blogger* saja yang mengikuti kompetisi maka dampak yang didapatkan oleh perusahaan dan organisasi adalah penelusuran produk barang dan jasanya di internet akan

memiliki *traffic* yang tinggi dengan biaya promosi yang relatif murah.

Dalam kompetisi tentu saja persaingan ketat sangat dirasakan dari para *blogger*. Terlebih lomba blog yang diselenggarakan dari perusahaan ternama dengan hadiah yang besar dan terbuka untuk umum, karena pesertanya homogen dari berbagai kalangan profesi, yang membuat setiap artikel dan blog memiliki keunikan sendiri. Apalagi jika pesertanya memang bekerja sebagai *content writer*. Selain ide dan gaya tulisan penilaian tambahan lainnya adalah tampilan blog, pemanfaatan SEO, Domain Name System (DNS) penggunaan judul, heading akan menjadi nilai plus [4]-[5].

Untuk mengatasi hal tersebut maka para peserta *blog competition* harus memperhatikan tulisan dan hal-hal yang membuat blog mereka terlihat menarik dan unik sehingga memiliki peluang menjadi juara.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Hingga penelitian ini dilakukan belum ditemukan secara spesifik penelitian data mining yang berkaitan dengan *blog competition*. Adapun penelitian sejenis tentang blog sebelumnya yang pernah diteliti oleh Robby, Lavinia, Darsono Nababan tentang "Analisa Algoritma C4.5 dalam Menentukan Faktor yang Mempengaruhi Munculnya Profesional Blogger". Penelitian ini

menghasilkan kesimpulan bahwa topik atau tema sangat berpengaruh terhadap munculnya professional blogger [6].

Penelitian lainnya dilakukan oleh Fabriyan Fandi Dwi Imaniawan, Ragil Wijianto, Joko Dwi Mulyanto dengan judul “Penerapan Search Engine Optimization Untuk Meningkatkan Peringkat Website Pada Search Engine Result Page Google” yang menitikberatkan pada Penggunaan *Search Engine Optimization* (SEO) untuk menaikkan Peringkat Website di halaman awal pencarian pada google [7].

Penelitian berikutnya oleh Ardiyansyah, Panny Agustia Rahayuningsih, Reza Maulana, dengan judul “Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner”. Penelitian ini menganalisis perbandingan lima algoritma klasifikasi data mining yaitu decision tree, Naïve bayes, k-nearest neighbour, ID3, dan CHAID dan mendapatkan nilai akurasi tertinggi sebesar 85% untuk algoritma KKN [8].

Penelitian berikutnya yang dilakukan oleh Fadila Darojatu Sidqi, Hendra Marcos dengan judul “Optimasi Website Menggunakan Metode Search Engine Optimization (SEO) On Page Off Page Dalam Meningkatkan Trafik Kunjungan”. Penelitian ini menitik beratkan pada metode SEO menggunakan Google Analytics [9].

Penelitian lainnya oleh Taftazani Ghazi Pratama, Agung Prihandono, Achmad Ridwan Leonardo

Agustinus Halim, dengan judul “Penerapan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Pada Algoritma C4.5 Dalam Menentukan Blogger Profesional”. Penelitian ini menitikberatkan pada algoritma *ensemble bagging* untuk meningkatkan akurasi [10].

Pentingnya keberadaan *blogger* dalam memperkenalkan merek produk kepada komunitas *online* dan mulai maraknya kompetisi blog yang diselenggarakan oleh pelaku usaha maka penelitian ini bertujuan untuk menganalisa dan memprediksi blog yang memiliki peluang juara dengan penerapan klasifikasi data mining algoritma C4.5. Penggunaan data mining karena dapat merubah big data menjadi informasi dan pengetahuan, dapat membuat pohon keputusan yang mudah dianalisis serta dapat menghasilkan akurasi yang wajar [11]

2.2. Landasan Teori

1. Data Mining

Data mining adalah metode pengolahan data untuk menemukan pola yang tersembunyi yang digunakan sebagai pertimbangan pengambilan keputusan. Dalam prosesnya Data mining terdiri dari 5 tahapan yaitu pembersihan data, Integrasi data, seleksi data, transformasi data dan proses mining [12].

2. Algoritma C4.5

Algoritma C4.5 merupakan sebuah struktur yang bisa digunakan untuk membagi kumpulan data

besar menjadi himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain. Output dari algoritma C4.5 berupa pohon keputusan yang dikembangkan J.Ross Quinlan. Algoritma C4.5 bagian dari kelompok *decision tree* yang memiliki input *training samples* dan *samples* [13].

3. Klasifikasi

Klasifikasi merupakan suatu metode dari data mining yang biasa digunakan dalam bidang pendidikan dengan melakukan proses Analisa data sehingga menciptakan model yang mendeskripsikan kelas-kelas dari data yang sedang dianalisa, biasa disebut *supervised learning* yang terdiri dari dua Langkah model konstruksi dan model *usage*. Tujuan klasifikasi untuk membedakan atribut ke dalam kategori kelas [14].

4. K-Fold Cross Validation

K-Fold Cross Validation adalah salah satu metode yang melakukan pengujian pada dataset, kendatipun pembagian dilakukan secara efektif dengan ukuran K sama untuk mendapatkan nilai akurasi. Dalam prosesnya nilai k biasanya ditentukan sebagai sepuluh partisi yang diulangi sepuluh kali hingga keakuratan dan kinerja model dapat ditemukan dengan nilai rata-rata sepuluh kali lipat [15].

5. Confusion Matrix

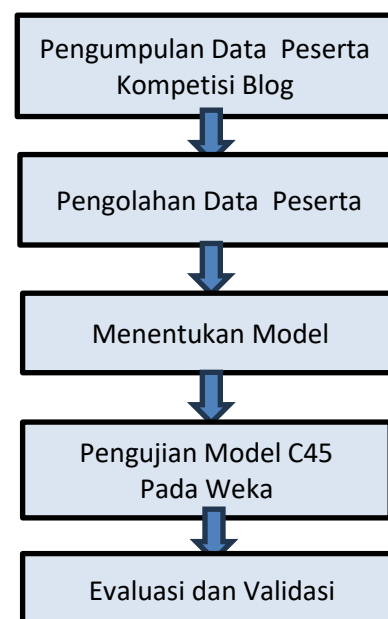
Confusion Matrix adalah sebuah cara yang digunakan untuk melihat hasil kinerja dari metode klasifikasi, dengan menggunakan empat istilah hasil proses klasifikasi yaitu True Positif (TP), True Negative (TN), False Positif (FP) dan False Negative (FN) [16].

6. Weka

Weka merupakan aplikasi yang digunakan untuk memvisualisasi algoritma analisis data dan prediksi pemodelan dengan antarmuka grafis untuk memudahkan user dalam melakukan pengolahan data mining yang diimplementasikan dengan bahasa pemrograman java [17].

2.3. Tahapan Penelitian

Proses alur dari penelitian ini dapat dilihat pada gambar 1 dibawah ini



Gambar 1. Alur Proses Penelitian

Alur proses dari penelitian diatas terdiri dari beberapa tahapan yaitu:

1. Dimulai dengan pengumpulan dataset yang diperoleh dari Person in Charge (PIC) penyelenggara Lomba Blog Asus.
2. Data diolah dengan mengambil secara random data peserta dan pemenang ketiga kompetisi sebanyak 100 data.
3. Langkah selanjutnya menentukan model
4. Kemudian dilakukan pengujian model C4.5 yang dengan pilihan J48 dengan aplikasi weka.
5. Hasil dari pemodelan berupa nilai akurasi, precision dan recall.
6. Proses evaluasi dan validasi

3. Metode Penelitian

3.1. Teknik Pengumpulan Data

Dataset yang digunakan dalam penelitian ini yaitu data yang didapat langsung dari Person in Charge (PIC) sedangkan teknik pengumpulan data menggunakan observasi dan wawancara.

Teknik observasi dilakukan dengan melakukan pengamatan langsung dari link blog PIC Lomba Blog Asus sedangkan teknik wawancara digunakan untuk mendapatkan izin penggunaan data peserta lomba dan data pemenang dari PIC yang ditunjuk oleh pihak penyelenggara serta informasi tambahan mengenai kriteria penilaian dalam lomba.

3.2. Teknik Pengolahan Data

Perangkat lunak yang digunakan dalam pengolahan data penelitian ini adalah aplikasi Weka dengan metode klasifikasi dengan menggunakan algoritma C4.5 yang diwakilkan pada pilihan J48 dalam aplikasi Weka [18].

3.3. Model Yang Diusulkan

Penelitian ini mengusulkan menggunakan metode klasifikasi algoritma C4.5 dengan pilihan J48 dengan k-fold cross validation dan untuk evaluasi menggunakan confusion matrix dengan hasil akhir nilai akurasi, recall dan precision

4. Hasil dan Pembahasan

4.1. Data Pra Proses

Dataset yang diolah dalam penelitian ini yaitu data primer yang didapatkan dari *Person in Charge* (PIC) penyelenggara lomba melalui tautan *link* blog PIC Lomba Blog Asus Zenbook 14X Oled (UX5400) www.bairuindra.com yang diambil sebanyak 25 data, Lomba Asus ExpertBook B3 Flip (B3402) sebanyak 50 data dan Lomba Blog Asus ExpertBook B7 Flip (B7402) sebanyak 25 data melalui blog PIC www.handikoo.com yang diambil data pemenang dan data peserta secara *random* [19].

Data yang diperoleh dari link PIC lomba blog ASUS hanya terdiri dari 3 atribut yaitu no, nama lengkap dan link artikel dapat dilihat pada tabel 1-

3. Sedangkan pengumuman juara diumumkan secara terpisah.

Tabel 1. Data Peserta Lomba Blog Asus Zenbook 14X Oled (UX5400)

No.	Nama	Link Blog
1	Adetya	https://www.misterpangalayo.com/2022/03/asus-zenbook-14x-oled-ux5400-11th-gen-intel.html
2	Adika Frisky Araaf	https://pencarinafkah.com/asus-zenbook-14x-oled-bikin-freelancer-jadi-lebih-produktif/
3	
144	Yudhistira Eko Wibowo	https://www.masekomer.com/2022/03/asus-zenbook-14x-oled-ux5400-penakluk-tantangan-hidup.html
145	Zahira Nur Arifah	https://dapurfrasa.blogspot.com/2022/03/asus-melahirkan-cendekiawan-bangsa.html

Tabel 2. Data Peserta Lomba Blog Asus ExpertBook B3 Flip (B402)

No.	Nama	Link Blog
1	Adam Nazar Yasin	https://tanivisual.blogspot.com/2022/06/asus-expertbook-b3-flip-b3402-laptop.html
2	Afra Dini Ashshiddiq	https://afraashshiddiq.wixsite.com/website/post/template-product-review-3
3	
154	Yvonie Oktharia Hutabarat	https://yvonieoh.blogspot.com/2022/06/5-alasan-asus-expertbook-b3-flip-bisa.html
155	Zahra Rabbiradlia	https://zahra-rabbiradlia.com/2022/06/asus-

Tabel 3. Data Peserta Lomba Blog Asus ExpertBook B7 Flip (B7402)

No.	Nama	Link Blog
1	Adam Nazar Yasin	https://tanivisual.blogspot.com/2022/06/expertbook-b7-flip-b7402-perangkat.html
2	Afra Dini Ashshiddiq	https://afraashshiddiq.wixsite.com/website/post/template-product-review
3	
145	Zahra Rabbira dlia	https://zuhafarhananii.blogspot.com/2022/06/anti-mood-swing-club-dengan-asus.html

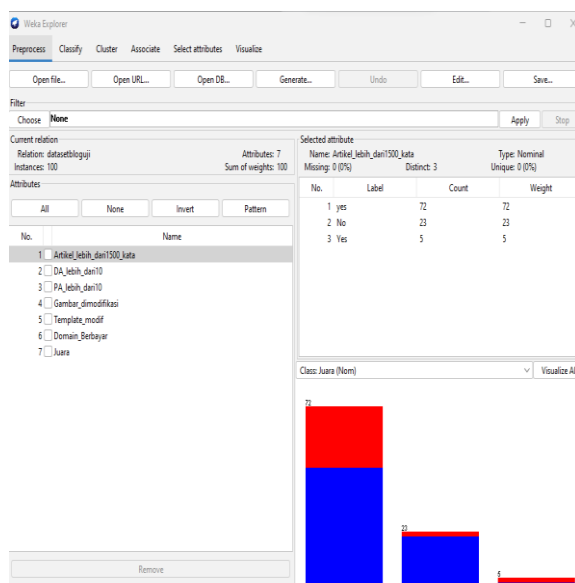
Kemudian data diolah menggunakan excel dengan menambahkan atribut jumlah kata dalam artikel dengan menggunakan *software* id.wordcounter360.com, Domain Authority (DA) dan Page Authority (PA) ditentukan dengan menggunakan *software* websiteseochecker.com sedangkan untuk gambar modifikasi, template dan domain berbayar pengecekan dilakukan secara manual dengan melihat artikel peserta lomba satu per satu. Dataset yang ada pada web panitia kemudian diolah dan menjadi dataset yang terdiri dari 7 atribut yaitu: jumlah kata dalam artikel, DA, PA, gambar dimodif, template modif, domain dan keterangan juara yang dapat dilihat pada tabel 4. Data yang telah disusun dalam excel kemudian diubah dengan format csv agar dapat diproses pada aplikasi Weka [20]-[21].

Tabel 4. Data Set yang Telah di Proses

No	Artikel >1500 Kata	DA > 10	PA > 10	Gambar modif	Template Modif	Domain TDL	Juara
1	Yes	No	No	No	No	No	No
2	No	No	No	No	No	Yes	No
3	No	Yes	Yes	Yes	Yes	Yes	Yes
4	No	Yes	Yes	No	No	No	No
5	No	No	No	No	No	Yes	No
6	No	No	No	No	No	Yes	No
7	No	No	No	No	No	No	No
8	No	No	No	No	No	No	No
9	No	No	No	Yes	Yes	Yes	No
10	Yes	Yes	Yes	Yes	Yes	Yes	No
...
97	Yes	Yes	Yes	Yes	Yes	Yes	No
98	Yes	Yes	Yes	Yes	Yes	Yes	No
99	Yes	Yes	Yes	Yes	Yes	Yes	No
100	No	Yes	Yes	Yes	Yes	Yes	No

4.2. Implementasi Pengujian C4.5 pada Weka

Dalam melakukan pengujian pada penelitian ini menggunakan aplikasi Weka dengan pemodelan klasifikasi Prediksi algoritma C4.5 *decision tree* yang dijalankan melalui *library* J48 untuk memprediksi blog yang memiliki peluang juara dengan menggunakan 100 data peserta lomba blog. Untuk tampilan pra proses data pada Weka dapat dilihat pada gambar 2



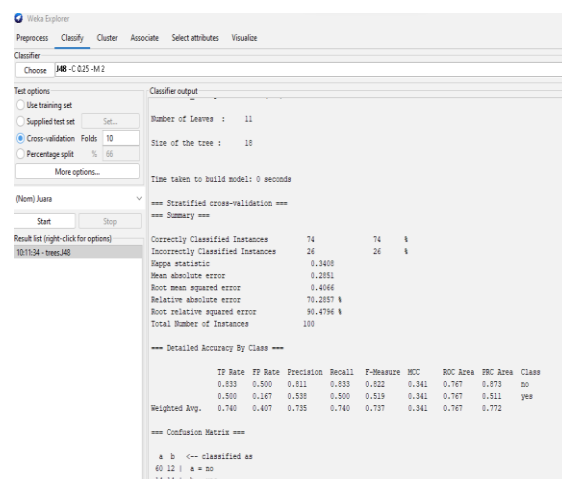
Gambar 2. Tampilan pra proses data

4.3. Model Prediksi K-Fold Cross Validation

Dalam mengevaluasi hasil, menggunakan pembagian data *training*, *testing* menggunakan *K-fold cross validation*, dengan nilai $K=2$ - $K=10$ yang menghasilkan *precision* dan *recall* berbeda-beda sedangkan untuk persentase akurasi menghasilkan hasil yang beragam yaitu untuk K3, K7 dan K9 menghasilkan akurasi yang sama yaitu sebesar 69%, berikutnya K5 dan K10 menghasilkan 74%, sedangkan hasil akurasi bervariasi ditunjukkan pada K2,K4,K6 dan K8 yang dapat dilihat pada tabel 5.

Tabel 5. Cross Validation C45

Ka	Akurasi	Precision	Recall
K=2	70%	0,681	0,700
K=3	69%	0,660	0,690
K=4	68%	0,680	0,680
K=5	74%	0,740	0,740
K=6	72%	0,714	0,720
K=7	69%	0,660	0,690
K=8	71%	0,713	0,710
K=9	69%	0,660	0,690
K=10	74%	0,735	0,740



Gambar 3. Classifier output k = 10

4.4. Pengukuran Confusion Matrix

Proses ini dilakukan dengan menggunakan aplikasi Weka berdasarkan 100 data yang ada. Hasil pengukuran dapat dilihat pada *Confusion Matrix* tabel 6 di bawah ini.

Tabel 6. *Confusion Matrix*

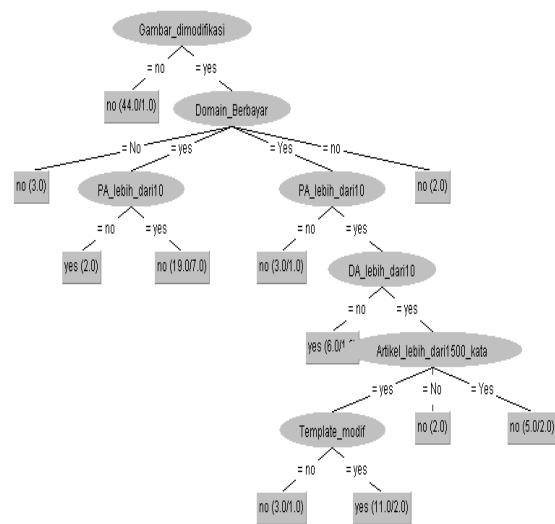
Jumlah Data	Data Error	TP	FP	FN	TN
100	24	60	12	14	14

Pada tabel 6 terdapat 24 data *error* dengan *True Positive* (TP) yaitu data yang tidak juara dan hasilnya tidak juara sebesar 60 data dengan *False Positive* (FP) data dengan hasil juara dan prediksinya tidak juara 12 data, *False Negative* (FN) data dengan hasil tidak juara tapi prediksinya juara terdapat 14 data dan *True Negative* (TN) data tidak juara dan prediksinya tidak juara sebanyak 14 data. Dari hasil ini terdapat 72 data yang tidak juara dan 28 yang juara.

4.5. Decision Tree

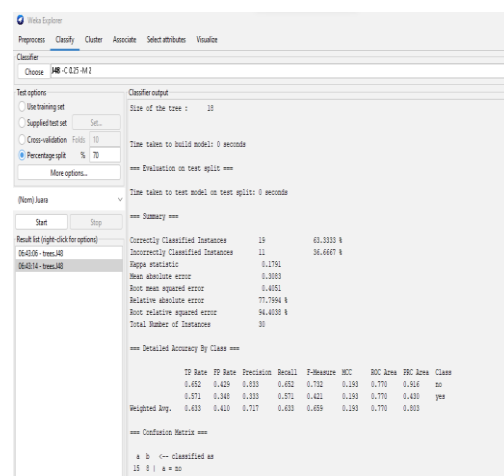
Setelah dilakukan proses prediksi, menghasilkan *decision tree* dengan *number of leaves* sebanyak 11 dan *size of tree* sebanyak 18. Jumlah daun menentukan rule hasil dari *decision tree* terlihat pada gambar 4. Dapat dilihat blog yang memiliki peluang juara adalah memiliki gambar pendukung yang dimodifikasi, kemudian cabang berikutnya menggunakan domain berbayar, PA

dan DA lebih dari 10, artikel lebih dari 1500 kata dan menggunakan *template* yang dimodifikasi.



Gambar 4. Tampilan Pohon Keputusan Pengujian Model Prediksi

4.6. Pengujian Model Prediksi Percentage Split
Pengujian kedua dilakukan dengan pembagian data *training* dan *testing* menggunakan *percentage split 70%* dan menghasilkan akurasi sebesar 63,33 % dengan *precision* 0,717 dan *recall* 0,633. Hasil pengujian dapat dilihat pada gambar 5.



Gambar 5. Classifier Output Percentage Split 70%

5. Kesimpulan

Berdasarkan hasil pengujian model prediksi yang telah diperoleh pada proses training dengan k-fold cross validation = 10 terdapat 74 data yang dapat diklasifikasikan dengan benar dan 26 data diklasifikasikan dengan salah dengan hasil akurasi sebesar 74%, precision sebesar 0,735 dan recall sebesar 0,740. Sedangkan dari hasil pengujian model prediksi yang telah diperoleh pada proses training dengan percentage split 70% dapat diklasifikasikan dengan benar 19 dan 11 data diklasifikasikan dengan salah dan menghasilkan akurasi sebesar 63,33 % dengan precision 0,717 dan recall 0,633.

Dari hasil pengujian keseluruhan proses training dengan k-fold cross validation = 10 yang dapat digunakan sebagai penentuan akurasi sebesar 74%. Berdasarkan hasil yang diperoleh dapat disimpulkan algoritma C4.5 dalam kategori cukup akurat untuk menganalisis dan memprediksi blog yang memiliki peluang juara

6. Daftar Pustaka

- [1] H. Kristanto and A. Nelson, "Strategi Promosi dan Membangun Image pada Media Digital/Online," *J. Pengabd. Kpd. ...*, vol. 3, no. 2, pp. 1545–1551, 2023.
- [2] M. Poongodi *et al.*, "5G based Blockchain network for authentic and ethical keyword search engine," *IET Commun.*, vol. 16, no. 5, pp. 442–448, 2022, doi: 10.1049/cmu2.12251.
- [3] R. A. Wilis and R. A. Augustina, "Influencer & Blogger Sebagai Strategi Pemasaran Era Digital," *JIS Wilis, R. A., Augustina, R. A. (2022). Influ. Blogger Sebagai Strateg. Pemasar. Era Digit. JIS (Jurnal Ilmu Siber), 1(4), 127–132. (Jurnal Ilmu Siber)*, vol. 1, no. 4, pp. 127–132, 2022.
- [4] T. Utari, "Pengembangan Search Engine Optimization (SEO) pada Digital Marketing Sebagai Strategi Pemasaran UMKM," *Ijd-Demos*, vol. 4, no. 2, pp. 676–685, 2022, doi: 10.37950/ijd.v4i2.247.
- [5] D. Kartika and Y. Mardiana, "Dns Server And Web Server Simulation With Debian Operating System On Local Area Network Simulasi Dns Server Dan Web Server Dengan Sistem Operasi Debian Pada Jaringan Local Area Network," vol. 2, no. 1, pp. 83–92, 2023.
- [6] R. Robby, "Analisa Algoritma C4.5 Dalam Menentukan Faktor Yang Mempengaruhi Munculnya Professional Blogger," *J. Ilm. Inform.*, vol. 7, no. 01, pp. 11–17, 2019.
- [7] F. F. Dwi Imaniawan, R. Wijianto, and J. D. Mulyanto, "Penerapan Search Engine Optimization Untuk Meningkatkan Peringkat Website Pada Search Engine Results Page Google (Studi Kasus: www.superbengkel.co.id)," *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 2, pp. 28–39, 2020, doi: 10.31294/evolusi.v8i2.9172.
- [8] Ardiyansyah, P. A. Rahayuningsih, and R. Maulana, "Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner," *J. Khatulistiwa Inform.*, vol. VI, no. 1, pp. 20–28, 2018.
- [9] F. D. Sidqi and H. Marcos, "Optimasi Website Menggunakan Metode Search Engine Optimization (Seo) on Page Dan Off Page Dalam Meningkatkan Trafik Kunjungan," *J. Inf. Syst. Manag.*, vol. 3, no. 2, pp. 55–58, 2021, doi: 10.24076/joism.2021v3i2.456.
- [10] P. G. Taftani, P. Agung, and R. Achmad, "Penerapan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Pada Algoritma C4.5 Dalam Menentukan Blogger Profesional," vol. 1, no. 1, pp. 49–55, 2020.

- [11] A. Pujiyanto, A. Mulyati, and R. Novaria, "Pemanfaatan Big Data Dan Perlindungan Privasi Konsumen Di Era Ekonomi Digital," *Maj. Ilm. Bijak*, vol. 15, no. 2, pp. 127–137, 2018, doi: 10.31334/bijak.v15i2.201.
- [12] R. Kurniah, D. Y. Surya Putra, and E. Diana, "Penerapan Data Mining Decision Tree Algoritma C4.5 Untuk Mengetahui Tingkat Kepuasan Mahasiswa Terhadap Layanan Akademik Dan Kemahasiswaan (Studi Kasus Universitas.Prof.Dr. Hazairin,SH)," *Infotek J. Inform. dan Teknol.*, vol. 5, no. 2, pp. 316–326, 2022, doi: 10.29408/jit.v5i2.5910.
- [13] M. Mahpuz, A. Muliawan Nur, and L. M. Samsu, "Penerapan Algoritma C4.5 Dalam Mengklasifikasi Status Gizi Balita Pada Posyandu Desa Dames Damai Kabupaten Lombok Timur," *Infotek J. Inform. dan Teknol.*, vol. 5, no. 1, pp. 72–81, 2022, doi: 10.29408/jit.v5i1.4414.
- [14] A. D. W. M. Sidik, I. Himawan Kusumah, A. Suryana, Edwinanto, M. Artiyasa, and A. Pradiftha Junfithrana, "Gambaran Umum Metode Klasifikasi Data Mining," *Fidel. J. Tek. Elektro*, vol. 2, no. 2, pp. 34–38, 2020, doi: 10.52005/fidelity.v2i2.111.
- [15] L. Mardiana, D. Kusnandar, and N. Satyahadewi, "Analisis Diskriminan Dengan K Fold Cross Validation Untuk Klasifikasi Kualitas Air Di Kota Pontianak," *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 11, no. 1, pp. 97–102, 2022, [Online]. Available: <https://jurnal.untan.ac.id/index.php/jbmstr/article/view/51608>
- [16] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix," *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.
- [17] K. Alpan and G. S. Ilgi, "Classification of Diabetes Dataset with Data Mining Techniques by Using WEKA Approach," *4th Int. Symp. Multidiscip. Stud. Innov. Technol. ISMSIT 2020 - Proc.*, no. October, 2020, doi: 10.1109/ISMSIT50672.2020.9254720.
- [18] M. M. Dr. Bhargava N., Sharma G., Dr. Bhargava R., "International Journal of Advanced Research in Decision Tree Analysis on J48 Algorithm for Data Mining," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 6, pp. 1114–1119, 2013.
- [19] T. Pramiyati, J. Jayanta, and Y. Yulnelly, "Peran Data Primer Pada Pembentukan Skema Konseptual Yang Faktual (Studi Kasus: Skema Konseptual Basisdata Simbumil)," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 8, no. 2, p. 679, 2017, doi: 10.24176/simet.v8i2.1574.
- [20] D. Knowles, "The domain of authority," *Philosophy*, vol. 82, no. 319, pp. 23–43, 2007, doi: 10.1017/S0031819107319025.
- [21] M. P. Gill, J. M. Prystowsky, L. Gill, and N. Page, "Calculating Web Page Authority Using the PageRank Algorithm," pp. 1–12, 2005.