

Prediksi Keputusan Berdonasi Pada Website Charity Box dengan Toko Waralaba Menggunakan Operator Cross Validation dan Algoritma Decision Tree

Hidayat Muhammad Nur¹, Vadlya Maarif^{2*}, Fabriyan Fandi Dwi Imaniawan³,
Sardiarinto⁴, Eko Saputro⁵

^{1,3,5} Sistem Informasi, Universitas Bina Sarana Informatika
Indonesia

² Teknologi Komputer, Universitas Bina Sarana Informatika
Indonesia

⁴ Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika
Indonesia

* Corresponding Author.E-mail:vadlya.vlr@bsi.ac.id

Abstrak

Donasi disebut sebagai sumbangan sebuah pemberian dari seseorang, tujuan bersifat sukarela untuk kemanusiaan. Penelitian ini masuk kategori survey dengan pendekatan kuantitatif melakukan pengamatan kegiatan berdonasi melalui media website berikut dengan cara konvensional pada toko waralaba, berdasarkan periode dataset dari tahun 2022 hingga tahun 2023, 3 kali putaran kegiatan donasi, tipe donasi melalui webbase atau toko waralaba, 6 kabupaten kota dan jumlah donasi. Hasil penelitian dengan operator validasi cross dimana data dibagi menjadi data testing dan training berdasarkan komposisi number of folds 10, algoritma *decision tree* untuk *training*, *apply model* untuk *testing* dan *performance accuracy main criteria*. *Decision tree* menggambarkan sistem prediksi yang dapat berguna teruntuk klasifikasi dan prediksi tugas. Kemudian memberlakukan teknik pembagi dan menaklukkan untuk membagi ruang pemeriksaan masalah menjadi himpunan masalah, pada prosesnya merombak bentuk data tabel menjadi sebuah model *tree*. Model *tree* akan menghasilkan *rule* dan disederhanakan. ini menunjukkan adanya jumlah akurasi persentase 55.00%. True menggunakan webbase 64.71%, True menggunakan toko waralaba 52.94% dari alat pengolahan data rapidminer.

Kata Kunci : Prediksi Donasi, Charity box webbase, Toko waralaba, Akurasi, *Decision tree*

Abstracts

Donations are referred to as donations from someone, whose purpose is voluntary for humanity. This research is in the survey category with a quantitative approach, observing donation activities through website media and conventional methods at franchise stores, based on the dataset period from 2022 to 2023, 3 rounds of donation activities, type of donation via website or franchise store, 6 city districts and donation amount. Research results with cross-operator validation where the data is divided into testing and training data based on the

composition of the number of folds 10, decision tree algorithm for training, application of the model for testing and the main criteria for performance accuracy. Decision trees describe prediction systems that can be useful for classification and prediction tasks. Then, dividing and placing techniques are applied to divide the problem examination space into a collection of problems, in the process of changing the data table shape into a model tree. The tree model will produce rules and balance. This shows that there is a total percentage accuracy of 55.00%. True uses a webbase 64.71%, True uses a franchise store 52.94% of the rapidminer data processing tool.

Keywords : Donation Prediction, Charity box webbase, Franchise shop, Accuracy, Decision tree

1. Introduction

Membuat program donasi pada suatu organisasi salah satunya melihat bahwa keberhasilan pengumpulan dana dilihat dari keberhasilan donasi yang terkumpul, berikut penyaluran dana donasi yang tepat pengelolaan berikut sasarannya. Namun pada masyarakat masih terdapat opini penyaluran dana donasi yang dianggap kurang tepat dikarenakan belum adanya transparansi informasi. Penyaluran dana donasi yang berpotensi kurang tepat sasaran dan tidak ada informasi transparan akan menimbulkan masalah yang akan memunculkan sentimen negatif dan penurunan kredibilitas sebuah lembaga pengumpulan dana.

Penelitian terdahulu yang terkait dengan faktor keputusan konsumen berdonasi dilakukan menggunakan algoritma K-NN, C.45, Naïve bayes dan decision tree. Berikut Prediksi atau proyeksi forecasting telah lama dipelajari dan tidak dapat disangkal telah memberikan manfaat banyak aktivitas manusia seperti sains, bisnis, pembelajaran mesin, penambangan data, pengetahuan, penemuan dan pengenalan pola (Muningsih et al., 2020).

Role akan tersimpan dengan tabel tipe donasi sebagai penentu keputusan berdonasi menempati peringkat pertama. Dalam hal

prediksi melalui track record sebuah donasi dapat diklasifikasikan, hasil donasi setiap putaran kegiatan dalam setahunnya. Hasil klasifikasi tersebut dapat diambil prediksi ataupun kesimpulan dalam berbagai macam hal, salah satunya yaitu mengklasifikasikan seberapa besar peluang sebuah metode donasi tersebut bisa diterima oleh masyarakat. Hasil ini menjadikan sebuah acuan bahwa kinerja sebuah metode donasi dengan sukarela mendapat tempat oleh seorang donator.(Resky & IAIN Parepare, 2021). Informasi file data jumlah donasi diakses, yang tersimpan pada repositori dan memuatnya ke dalam Proses. Operator Cross validation ini melakukan validasi silang untuk memperkirakan kinerja statistik model pembelajaran donasi dalam prakteknya, dimana data dibagi menjadi data testing dan training berdasarkan komposisi number of folds 10 (Debora et al., 2023). Kemudian algoritma decision tree, ini menghasilkan model pohon keputusan, yang dapat digunakan untuk klasifikasi dan regresi training, dilanjutkan menerapkan apply model contoh dataset untuk testing dan performance accuracy main digunakan untuk evaluasi kinerja statistik tugas klasifikasi. Operator ini mengirimkan daftar nilai kriteria kinerja tugas klasifikasi (Bahri & Lubis, 2020).

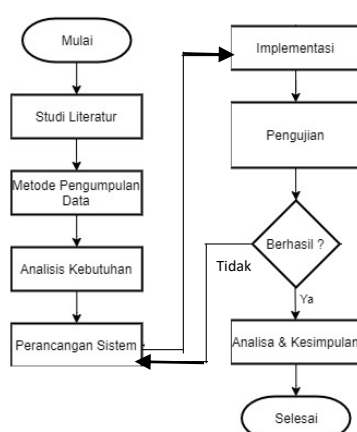
Penelitian ini menggunakan *Cross Validation* dan *Algoritma Decision Tree* dengan menggunakan data donasi pada tahun 2022 sampai tahun 2023, dengan mendapatkan data donasi berpotensi memanfaatkan *website* dan toko waralaba sebanyak 13.951.500 data dengan 5 atribut. Dan bertujuan untuk mengetahui faktor – faktor yang paling mempengaruhi masyarakat berpotensi melakukan donasi dan menerapkan algoritma *decision tree* untuk mengklasifikasikan pemberi donasi yang berpotensi memanfaatkan antara dua media tersebut sehingga hasilnya lebih akurat. Dengan adanya kriteria kinerja dan klasifikasi tersebut, dapat diambil keputusan dalam jumlah atribut mana yang paling besar menentukan peluang sebuah donasi tersebut mampu mendominasi disetiap putaran kegiatan amal kemanusiaan. Untuk mempermudah dalam klasifikasi data, penelitian ini menggunakan aplikasi RapidMiner Studio 10.3.

2. Materials and Methods

Metode penelitian yang diterapkan pada penelitian berikut berupa studi literature, selanjutnya metode pengumpulan data, melakukan analisis kebutuhan, perancangan sistem, implementasi, dan pengujian serta analisa dan kesimpulan. Flowchart dapat dilihat pada gambar 2 (Lutfi Nur Afifah, 2020)



Gambar 1. Konsep Pohon Keputusan



Gambar 2. Diagram Alir Penelitian

Preprocessing Data

Penyaringan sebelum dianalisis merupakan teknik tindakan yang berguna tatkala penambahan data untuk memperoleh pengetahuan dari basis data pada kepemilikan tabel yang saling berelasi agar mendapatkan dasar dalam pengambilan keputusan. *Preprocessing* data pada penelitian ini dilakukan dengan 3 langkah berikut :

1. Data transformasi dalam proses ini, data ditransformasikan ke dalam bentuk yang sesuai untuk proses data mining.
2. Data reduksi data dilakukan dengan menentukan atribut yang tidak diperlukan sehingga ukuran dari *database* menjadi kecil dan hanya menyertakan atribut yang diperlukan dalam proses *data mining*, karena akan lebih efisien terhadap data yang lebih kecil.
3. Data label dilakukan dengan menentukan atribut penamaan label, guna dijadikan *role* untuk dipilih atau tidaknya suatu suara (Samponu & Kusrini, 2018). Berikut ini adalah tabel data set yang akan digunakan.

Tabel 1 Data Set Hasil *Preprocessing*

No	Tahun	Putaran	Tipe Donasi	Kabupaten Kota	Jumlah Donasi
1	2022	I	Menggunakan Webbase	Kepulauan Seribu	12756
3	2022	I	Menggunakan webbase	Jakarta Selatan	947780
4	2022	I	Menggunakan alfamart	Jakarta Selatan	559201
5	2022	I	Menggunakan Webbase	Jakarta Timur	1278520
6	2022	I	Menggunakan Alfamart	Jakarta Timur	714330
7	2022	I	Menggunakan Webbase	Jakarta Pusat	488676
8	2022	I	Menggunakan Alfamart	Jakarta Pusat	298748
9	2022	I	Menggunakan Webbase	Jakarta Barat	996871
10	2022	I	Menggunakan Alfamart	Jakarta Barat	499765
11	2022	I	Menggunakan Webbase	Jakarta Utara	704930
12	2022	I	Menggunakan Alfamart	Jakarta Utara	457223
13	2023	II	Menggunakan Webbase	Kepulauan Seribu	12116
15	2023	II	Menggunakan Webbase	Jakarta Selatan	1000627
16	2023	II	Menggunakan Alfamart	Jakarta Selatan	512286
17	2023	II	Menggunakan Webbase	Jakarta Timur	1328944
18	2023	II	Menggunakan Alfamart	Jakarta Timur	670096
19	2023	II	Menggunakan Webbase	Jakarta Pusat	515729
20	2023	II	Menggunakan Alfamart	Jakarta Pusat	273755
21	2023	II	Menggunakan Webbase	Jakarta Barat	1066945
22	2023	II	Menggunakan Alfamart	Jakarta Barat	443214
23	2023	II	Menggunakan Webbase	Jakarta Utara	743580
24	2023	II	Menggunakan Alfamart	Jakarta Utara	425408

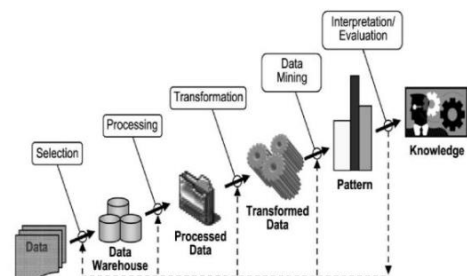
3. Hasil dan Pembahasan

Teknik dalam data mining untuk dapat temukan patokan himpunan antar item gabungan, adalah dengan atribut yang dipilih, parameter, *example set*, *performace vector* seperti berikut dibawah ini :

1. Menentukan Nilai Atribut 1

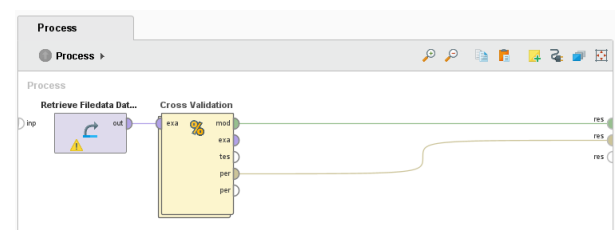
Pada tahap ini penelitian akan

dilakukan menggunakan software rapidminer dengan metode *decision tree*. Data yang diolah menggunakan Ms. Excel 2013, sehingga di aplikasi rapidminer menggunakan operator *read excel* yang berfungsi untuk membaca file yang akan diolah. Sebelumnya menerapkan algoritma operator *Cross Validation* untuk testing dan training membagi dan diacak menjadi 10 data, kemudian akan diolah menghasilkan pohon keputusan pada nilai atribut 2.



Gambar 3. Proses Knowledge Discovery In Database (I Gede Wayan Sudipa, 2023)

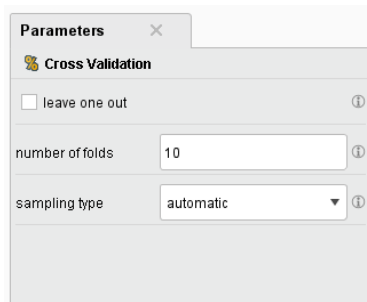
Sampailah pada tahap untuk menyuguhkan data yang tersimpan ini mempunyai corak mutu data, yaitu melepaskan data tidak lengkap atau tidak valid. Tentunya agar mendapatkan mutu data dan valid, sinkron, cocok dengan data yang di inputkan.



Gambar 4. Proses Awal Rapidminer

2. Menentukan Parameter 1

Kriteria yang digunakan adalah *number of folds* digunakan untuk memberikan nilai k (jumlah iterasi). Sampling type digunakan untuk memilih teknik sampling yang membagi dataset.

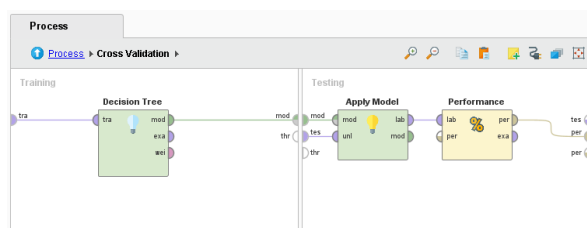


Gambar 5. Parameter pada operator *Cross Validation*

Ini adalah port dapat diperluas sesuai kebutuhan (kurang atau lebih dari 10). Kemudian menghubungkan vektor kinerja (hasil dari Operator Kinerja) ke port hasil subproses Pengujian. Port keluaran kinerja Operator Validasi Silang memberikan rata-rata kinerja selama jumlah iterasi.

3. Menentukan Nilai Atribut 2

Penentuan nilai atribut ini terdapat dua hal yang berhubungan yaitu meramalkan kategori label kelas (*training* dan *testing*) dan menggolongkan data yang didasarkan pada pelatihan menetapkan dan nilai-nilai kedalam suatu penggolongan data baru, dalam hal ini diaplikasikan secara umum untuk klasifikasi analisis keefektifan tindakan.

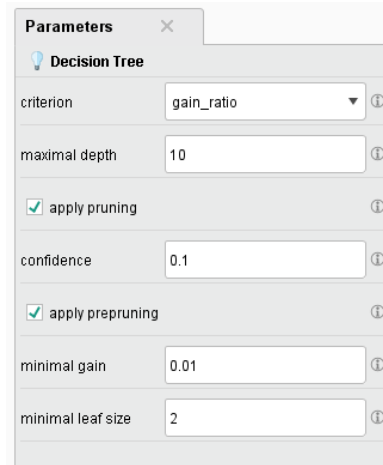


Gambar 5. Proses Penentuan Atribut Decision Tree

4. Menentukan Parameter 2

Kriteria yang digunakan adalah *gain ratio* berdasarkan informasi yang diperoleh terhadap informasi intrinsic yang berfungsi untuk mengurangi bias terhadap atribut multi nilai dengan memperhitungkan jumlah dan ukuran cabang ketika memilih

atribut. Pada pilihan maximal depth menunjukkan maksimal panjang percabangan dari puncak pohon kecabang-cabangnya.



Gambar 6. Parameter pada operator *Decision Tree*

5. Example Set

Adalah data set yang sudah terupload di aplikasi rapidminer. Berikut adalah data set yang siap untuk diproses dengan aplikasi rapidminer.

Row No.	periode_data	putaran	tipe_donasi	kabupaten_...	jumlah_don...
22	2023	I	Menggunaka...	Jakarta Barat	499765
23	2023	I	Menggunaka...	Jakarta Utara	704930
24	2023	I	Menggunaka...	Jakarta Utara	457223
25	2023	II	Menggunaka...	Kepulauan S...	12116
26	2023	II	Menggunaka...	Kepulauan S...	4251
27	2023	II	Menggunaka...	Jakarta Selat...	1000627
28	2023	II	Menggunaka...	Jakarta Selat...	512286
29	2023	II	Menggunaka...	Jakarta Timur	1328944
30	2023	II	Menggunaka...	Jakarta Timur	670096
31	2023	II	Menggunaka...	Jakarta Pusat	515729
32	2023	II	Menggunaka...	Jakarta Pusat	273755
33	2023	II	Menggunaka...	Jakarta Barat	1066945
34	2023	II	Menggunaka...	Jakarta Barat	443214
35	2023	II	Menggunaka...	Jakarta Utara	743580
36	2023	II	Menggunaka...	Jakarta Utara	425408

ExampleSet (36 examples, 1 special attribute, 4 regular attributes)

Gambar 7. *Example Set*

6. Performance Vector

Pada hasil akhir proses ini performance akan menampilkan hasil. Data hasil pembagian masing-masing data sesuai prediksi atau tidak hingga presentasi keakurasian pengkategorian *decision tree*,

dengan memberikan dua hasil pokok yaitu bagan *decision tree* dan tabel *performance*. Bila dikelompokkan dari setiap kategori maka akan menjadi seperti pada gambar 8. Prediksi True positive baris pertama menggunakan webbase mendapat point 11+1, true positive menggunakan waralaba 8. Baris kedua true positive menggunakan webbase mendapat point 6+1, true positive waralaba 9. Class recall (sensitifitas), merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Recall menjawab pertanyaan persentase donasi webbase yang diprediksi diberikan dibandingkan keseluruhan donasi yang sebenarnya yaitu true menggunakan webbase berjumlah 65.71%, true menggunakan waralaba 52.94%.

Kelompok kategori yang paling akurat dan *confusion matrix* nilai akurasi yang dihasilkan dari 10 *fold cross validation* sebesar 55.00% dengan standar deviasi (+/-) sebesar 15.32%.

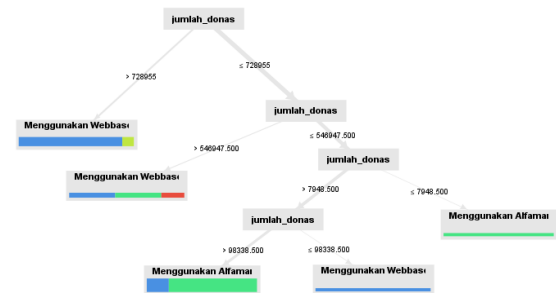
accuracy: 55.00% +/- 15.32% (micro average: 55.56%)

	true Menggunakan ...	true Menggunakan ...	true Menggunakan ...	true Menggunakan ...	class precision
pred. Menggunakan ...	11	8	1	0	55.00%
pred. Menggunakan ...	6	9	0	1	56.25%
pred. Menggunakan ...	0	0	0	0	0.00%
pred. Menggunakan ...	0	0	0	0	0.00%
class recall	64.71%	52.94%	0.00%	0.00%	

Gambar 8. *Performace*

7. Graph View

Graph view menunjukkan hasil dari klasifikasi dengan percabangan yang dapat dihasilkan kesimpulan. Berikut adalah hasil dari *decision tree*.



Gambar 9. Hasil *Decision Tree*

Dari grafik tersebut diatas dapat diberikan penjelasan bahwa dari 4 jumlah kegiatan donasi, terdapat 2 tipe data yang ditampilkan dan mengindikasikan lebih banyak penggunaan *webbase* (Pangestu & Solichin, 2022).

Visualisasi grafik di tampilan hasil. Ini tidak hanya mencakup grafik pohon keputusan, tetapi juga pengelompokan, aturan asosiasi, dan lainnya. menampilkan informasi garis dalam Pohon Keputusan memiliki ketebalan berbeda sehingga Anda dapat segera mengidentifikasi cabang lebih besar yang membawa lebih banyak informasi.

4. Kesimpulan

Adapun dari hasil pembahasan tersebut diatas, maka diberikan kesimpulan antara lain; Memilih algoritma yang memiliki recall lebih tinggi atau true positif lebih baik dari pada false positif.

Hasil dari data yang sudah diolah menunjukkan bahwa keputusan berdonasi menggunakan *webbase* mendapatkan tempat dan efektif di masyarakat kabupaten kota DKI dan kepulauan seribu. Dimana dilihat dari perolehan point analisa di putaran kegiatan donasi berdasarkan atribut tipe donasi masih menjadi indikator keefektifan.

Model yang dihasilkan didasarkan pada pemungutan suara karena perbedaan, ia tidak terlalu rentan terhadap jumlah donasi.

Index ukuran ketidaksetaraan antara sebaran karakteristik label. Pemisahan pada Atribut yang dipilih menghasilkan pengurangan rata-rata indeks dari himpunan bagian yang dihasilkan.

5. References

- [1] Arif Rinaldi Dikananda, etc (2022). Comparison of Decision Tree Classification Methods and Gradient Boosted Trees. *TEM Journal* Vol.11 Issue 1, pages 316-322
- [2] Amna & Wahyudin s, etc (2023). *Data Mining*. PT. Global Eksekutif Teknologi, Anggota IKAPI No.033/SBA/2022
- [3] Bahri, S., & Lubis, A. (2020). Metode Klasifikasi Decision Tree Untuk Memprediksi Juara English Premier League. *Jurnal Sintaksis*, 2(1), 63–70.
- [4] Debora, R. M., Mahdiana, D., Rony, M. A., & Hasugian, H. (2023). *Application of C . 45 Algorithm for Potential Students Classification Drop Out at Budi Luhur University*. 2(1), 316–325.
- [5] Dewi Anggraeni, Rizaldi (2020). Rapidminer Implementation In Analyzing Student Mainstream Data. *Proceeding International Conference on Social, Science and Information Technology ICoSSIT hlm. 1-6*
- [6] Lutfi Nur Afifah, A. (2020). Analisis Prediksi dan Faktor Pendukung Kesuksesan Crowdfunding Berbasis Donasi Menggunakan Algoritma C.45 dan K-Nearest Neighbor (Studi Kasus : Kitabisa.com). *Corporate Governance (Bingley)*, 10(1), 54–75.
- [7] Mirka Saarela & Sussane Jauhiainen (2021). Comparison of Feature Importance Measures as Explanations for Classification Models. *Springer Natures Journal - SN Applied Sciences* Vol.3 No.272
- [8] Muningsih, E., Nur, H. M., Dwi Imaniawan, F. F., Saifudin, Handayani, V. R., & Endiarto, F. (2020). Comparative Analysis on Dimension Reduction Algorithm of Principal Component Analysis and Singular Value Decomposition for Clustering. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012101>
- [9] Robianto, M., Hotlan Sitorus, S., Ristian, U., Rekayasa Sistem Komputer, J., & MIPA Universitas Tanjungpura Jalan Hadari Nawawi Pontianak, F. H. (2021). *Penerapan Metode Decision Tree Untuk*. 09(01), 76–86.
- [10] Nurhayati, dkk (2021). Implementasi Metode Decision Tree Pada Tingkat Prestasi Belajar Siswa di SMK Swasta Anak Bangsa. *BEES : Bulletin Of Electrical and Electronics Engineering* Vol.1 No.3, pp 108-114
- [11] Omar Chamorro-Atalaya etc (2023). K-Fold Cross-validation through Identification of the Opinion Classification Algorithm for the Satisfaction of University Students. *iJOE International Journal of Online and Biomedical Engineering* Vol.19 No.11
- [12] Pangestu, R., & Solichin, A. (2022). Klasifikasi Serangan Jaringan Menggunakan Metode Decision Tree Berbasis Website. *Seminar Nasional Mahasiswa Fakultas Teknologi*

- Informasi (SENAFTI)*, September, 614–620.
<https://senafti.budiluhur.ac.id/index.php/>
- [13] Resky, D. P., & IAIN Parepare. (2021). Faktor-faktor Yang Mempengaruhi Keputusan Konsumen Berdonasi Pada Alfamart Parepare (Analisis Ekonomi Syariah). *Tesis*, 3(2), 6.
- [14] Samponu, Y. B., & Kusrini, K. (2018). Optimasi Algoritma Naive Bayes Menggunakan Metode Cross Validation Untuk Meningkatkan Akurasi Prediksi Tingkat Kelulusan Tepat Waktu. *Jurnal ELTIKOM*, 1(2), 56–63.
<https://doi.org/10.31961/eltikom.v1i2.29>
- [15] Syahancim Marzukhi, etc (2021). RapidMiner and Machine Learning Techniques for Classifying Aircraft Data. *Journal of Physics Asian Conference on Intelligent Computing and Data Sciences (ACIDS)* IOP Publishing