

Prediksi Jumlah Retribusi Sampah Perbulan menggunakan Algoritma Decition Tree (C4.5) pada Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur

Yupi Kuspandi Putra¹, Yeni Kurniatul Mashun²
Fakultas Teknik Universitas Hamzanwadi
yupi.putra@gmail.com¹, yenikurnia390@gmail.com²

Abstrak

Seiring perkembangan zaman dan pertumbuhan penduduk yang semakin tinggi, kebutuhan manusia pun semakin meningkat. Setiap manusia memiliki kebutuhan yang beragam. Apapun yang manusia butuhkan akan menghasilkan sesuatu yang harus dibuang pada akhirnya dan menjadi sampah. Atas jasa pengangkutan sampah tersebut, Kantor Dinas Lingkungan Hidup dan Kebersihan akan memungut bayaran dari pihak-pihak terkait sesuai jumlah yang telah ditentukan. Pemungutan bayaran ini dikenal dengan istilah retribusi. Guna membantu pihak kantor, diperlukan alat bantu untuk memprediksi pendapatan retribusi sampah berdasarkan data retribusi hariannya. Model algoritma yang digunakan adalah Decition Tree (C4.5) untuk mendapatkan rule dalam memprediksi jumlah retribusi sampah perbulan pada Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur. Setelah dilakukan pengujian sebanyak tiga kali, ketiganya menunjukkan hasil akurasi dan nilai AUC yang tinggi. Hasil yang paling unggul yaitu dengan menggunakan number of validation 3, dengan nilai akurasi yang diperoleh sebesar 96.51% dan nilai AUC sebesar 0.976% dengan tingkat diagnosa nilai akurasi excellent classification.

Kata kunci : Prediksi jumlah retribusi sampah perbulan, Algoritma Decision Tree (C4.5)

Abstract

Along with the times and growing population growth, human needs are increasing. Every human being has diverse needs. Whatever humans need will produce something that must be discarded in the end and become waste. For the waste transportation services, the Environmental Service Office and the Hygiene Agency will collect fees from the relevant parties according to the specified amount. This fee collection is known as retribution. In order to help the office, a tool is needed to predict waste retribution revenue based on the daily retribution data. The algorithm model used is Decition Tree (C4.5) to get the rule in predicting the amount of monthly waste retribution at the Environmental and Hygiene Office of Selong District, East Lombok Regency. After three tests, all three showed high accuracy and AUC values. The most superior result is by using number of validation 3, with the accuracy value obtained is 96.51% and AUC value is 0.976% with a diagnosis level of excellent classification accuracy.

Keywords: Prediction of the amount of monthly waste retribution, Decision Tree Algorithm (C4.5)

1. Pendahuluan

Setiap manusia memiliki kebutuhan yang beragam. Apapun yang manusia butuhkan akan menghasilkan sesuatu yang harus dibuang pada akhirnya dan menjadi sampah. Bisa dibayangkan

berapa banyak sampah yang dihasilkan oleh seseorang dalam setiap harinya. Maka dari itu, perusahaan-perusahaan atau dinas-dinas yang bekerja untuk masyarakat dalam menangani sampah tersebut sangat dibutuhkan. Salah satu

dinas yang bergerak dalam bidang ini adalah Dinas Lingkungan Hidup dan Kebersihan.

Setiap harinya, staf yang bertugas akan mengangkut sampah-sampah yang dihasilkan masyarakat. Atas jasa pengangkutan sampah tersebut, Kantor Dinas Lingkungan Hidup dan Kebersihan akan memungut bayaran dari pihak-pihak terkait sesuai jumlah yang telah ditentukan. Pemungutan bayaran ini dikenal dengan istilah retribusi. Pada setiap bulannya Kantor Dinas Lingkungan Hidup dan Kebersihan membuat laporan hasil retribusi sampah untuk mengetahui jumlah retribusi sampah yang didapat. Hasil dari retribusi tersebut kemudian menjadi salah satu sumber Pendapatan Asli Daerah (PAD). Oleh karena itu, prediksi jumlah retribusi sampah perbulan harus diketahui. Hal ini diperlukan untuk mengetahui prediksi jumlah retribusi yang didapatkan setiap bulannya. Informasi ini dibutuhkan oleh Kepala Dinas dan Bendahara Penerimaan. Untuk menyelesaikan permasalahan di atas, salah satu metode yang dapat digunakan adalah metode pohon keputusan dengan menggunakan algoritma *decision tree* (C4.5). Algoritma *decision tree* (C4.5) merupakan salah satu jenis algoritma pengelompokan data berdasarkan kriteria tertentu.

2. Tinjauan Pustaka

2.1. Landasan Teori

1) Retribusi Daerah

Menurut Pasal 1 Undang-Undang Nomor 28 Tahun 2009 tentang pajak daerah dan retribusi daerah, pengertian retribusi daerah, yang selanjutnya disebut retribusi adalah pungutan daerah sebagai pembayaran atas jasa atau pemberian izin tertentu yang khusus disediakan dan/atau diberikan oleh pemerintah daerah untuk kepentingan orang pribadi atau badan.

Wajib retribusi adalah orang pribadi atau badan yang menurut peraturan perundang-undangan retribusi diwajibkan untuk melakukan pembayaran retribusi, termasuk pemungut atau pemotong retribusi tertentu. Besarnya retribusi yang terutang oleh orang pribadi atau badan yang menggunakan jasa atau perizinan tertentu dihitung dengan cara mengalikan tarif retribusi dengan penggunaan jasa (Pasal 1 Undang-Undang Nomor 28 Tahun 2009).

2) Data Mining

Secara sederhana, Data Mining adalah penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004). Data Mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2006). Data Mining sering juga disebut Knowledge Discovery in Database (KDD). KDD merupakan suatu

kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2007). Data Mining memiliki beberapa karakteristik, diantaranya adalah:

- Data Mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang sebelumnya tidak diketahui.
- Data Mining biasanya menggunakan data yang sangat besar agar hasilnya lebih dipercaya.
- Data Mining bisa digunakan untuk membuat keputusan yang kritis, terutama dalam strategi (Davies, 2004)

3) Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (Iterative Dichotomiser 3) yang ditemukan oleh J Ross Quinlan. Dalam pohon keputusan sangat berhubungan dengan algoritma C4.5, karena dasar algoritma C4.5 adalah pohon keputusan.

Dalam algoritma C4.5 pemilihan atribut dilakukan dengan menggunakan gain ratio, dengan mencari nilai entropy. Algoritma C4.5 sendiri menggunakan pendekatan induksi dimana dalam pendekatan ini, algoritma C4.5 membagi data berdasarkan kriteria yang dipilih untuk membuat sebuah pohon keputusan yang menggunakan pendekatan secara top-down.

Algoritma C4.5 menggunakan konsep information gain atau entropy reduction untuk

memilih pembagian yang optimal. Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5 yaitu:

- Mempersiapkan data training, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
- Menentukan akar dari pohon dengan menghitung nilai gain yang tertinggi dari masing-masing atribut atau berdasarkan nilai index entropy terendah.

3. Metode Penelitian

3.1. Pengumpulan Data

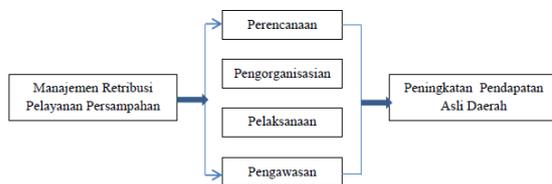
Dalam penelitian ini data yang digunakan adalah data sekunder karena diperoleh dari database yang dimiliki Kantor DLHK yaitu melalui salah satu staf kantor yang berkedudukan sebagai bendahara penerima. Data yang dikumpulkan adalah data retribusi periode 2016. Data terkumpul sebanyak 2.552 data, dengan atribut tanggal, nomor bukti, uraian dan nilai.

3.2. Model Yang Diusulkan

Model yang diusulkan dalam penelitian ini adalah menggunakan algoritma Decision Tree (C4.5). Model ini melakukan pengolahan dataset dalam melakukan klasifikasi terhadap kelas yang telah ditentukan. Dalam pengujian ini, penulis melakukan 3 kali uji coba dengan menggunakan K-Fold Cross Validation.

3.3. Kerangka Pemikiran

Untuk menjalankan roda pemerintahan sebagai otonomi daerah harus memanfaatkan semua potensi yang ada salah satunya adalah retribusi persampahan untuk meningkatkan pendapatan asli daerah (PAD). Untuk itu pemerintah daerah harus mengetahui manajemen retribusi pelayanan persampahan di Kota Selong. Sehingga dapat mengidentifikasi setiap permasalahan dalam retribusi persampahan yang dalam pengelolaannya di bedakan menjadi 4 bagian sesuai konsep dari **George R. Terry** yaitu perencanaan, pengorganisasian, pelaksanaan dan pengawasan, yang mengarah pada peningkatan Pendapatan Asli Daerah (PAD). Dapat dilihat pada gambar 1.



Gambar 1. Kerangka Pemikiran

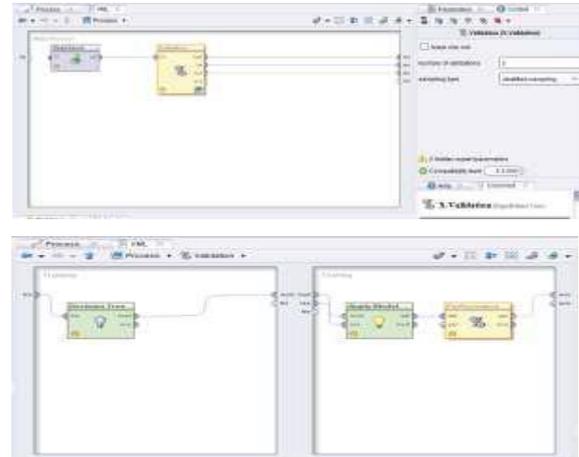
4. Hasil Penelitian Dan Pembahasan

Data yang akan diuji adalah data retribusi persampahan/kebersihan yang berjumlah 173 data. Data tersebut kemudian akan dibagi menjadi dua bagian yaitu data training dan data testing. Pada uji coba pertama ini penulis menggunakan *number of validation* 3.

a. Evaluasi model dengan *Confusion Matrix*

Model *confusion matrix* akan membentuk matrix yang terdiri dari *true positif* dan *true negatif*, kemudian masukkan data testing yang telah

disiapkan sebelumnya ke dalam *confusion matrix* sehingga mendapatkan hasil. Berikut gambar area kerja RapidMiner untuk mendapatkan hasil akurasi.



Gambar 2. Lembar kerja RapidMiner untuk menentukan hasil *accuracy*

Hasil akurasi yang diperoleh dalam pengujian adalah 96.51%. Hasilnya dapat dilihat pada gambar berikut :

	true SEMPIT	true LUAS	class precision
pred SEMPIT	122	0	100.00%
pred LUAS	6	45	88.24%
class recall	95.97%	100.00%	

Gambar 3. Hasil *accuracy* uji coba pertama

Untuk memperjelas hasil *accuracy* di atas, maka dibuat dalam bentuk tabel, seperti tabel di bawah ini.

Tabel 1. Hasil *accuracy* uji coba pertama

<i>Predicate Class</i>		
<i>Observed Class</i>	Sempit	Luas
Sempit	122	0
Luas	6	45

Jumlah *true positif* (TP) adalah 122 *record* yang diklasifikasikan sebagai titik pengangkutan yang SEMPIT dan *false negatif* (FN) sebanyak 0 *record* diklasifikasikan sebagai titik

pengangkutan yang SEMPIT namun ternyata LUAS. Selanjutnya 6 record untuk *true negatif* (TN) diklasifikasikan sebagai titik pengangkutan LUAS dan 45 record *false positif* (FP) diklasifikasikan sebagai titik pengangkutan LUAS namun ternyata SEMPIT. Berdasarkan tabel 4.1 di atas, tingkat akurasi yang diperoleh menggunakan algoritma Decision Tree (C4.5) dengan *K-Fold Cross Validation* dengan *number of validation* 3 adalah sebesar 96.51%. Nilai *accuracy*, *sensitivity*, *specitivity*, *ppv*, dan *npv* dapat dicari dengan menggunakan persamaan dibawah ini.

$$\begin{aligned} \textit{Accuracy} &= \frac{tp+tn}{tp+tn+fp+fn} \\ &= \frac{122+45}{122+45+6+0} = 96.51\% \end{aligned}$$

Dengan menggunakan *K-Fold Cross Validation* dengan *number of validation* 3, sebanyak 173 data untuk diprediksi yang terdiri dari data training dan data testing diperoleh hasil akurasi 96.51%, ini artinya dari 173 data yang di uji, 167 data yang dapat diprediksi dan 6 data sisanya tidak dapat diprediksi. Sehingga tingkat kesalahan yang diperoleh adalah 4%.

$$\begin{aligned} \textit{Sensitivity} &= \frac{tp}{tp+fn} \\ &= \frac{122}{122+0} = 100\% \end{aligned}$$

Dengan pengujian *K-Fold Cross Validation* dengan *number of validation* 3, dari 173 data untuk diprediksi, dimana data tersebut terdiri dari data training dan data testing, diperoleh hasil *true positif* (TP) yaitu hasil dari titik pengangkutan yang sempit sebanyak 100%, hal

ini berarti semua data diprediksi sebagai titik pengangkutan yang sempit.

$$\begin{aligned} \textit{Specitivity} &= \frac{tn}{tn+fp} \\ &= \frac{45}{45+6} = 88.24\% \end{aligned}$$

Dengan pengujian *K-Fold Cross Validation* dengan *number of validation* 3, sebanyak 173 data untuk diprediksi yang terdiri dari data training dan data testing diperoleh hasil *true negative* (TN) yaitu hasil dari titik pengangkutan yang luas adalah 88.24%, ini berarti 153 data yang diprediksi titik pengangkutan luas dan 20 data sisanya tidak diprediksi sebagai titik pengangkutan luas. Sehingga tingkat kesalahan yang diperoleh adalah 12%.

$$\begin{aligned} \textit{Ppv} &= \frac{tp}{tp+fp} \\ &= \frac{122}{122+6} = 95.31\% \end{aligned}$$

Dengan pengujian *K-Fold Cross Validation* dengan *number of validation* 3, sebanyak 173 data untuk diprediksi yang terdiri dari data training dan data testing diperoleh hasil *false positif* (FP) yaitu hasil dari titik pengangkutan yang luas namun ternyata sempit adalah 95.31%, hal ini berarti 164 data yang diprediksi titik pengangkutan yang luas namun ternyata sempit dan 9 data sisanya tidak diprediksi demikian. Sehingga tingkat kesalahan yang diperoleh adalah 5%.

$$\begin{aligned} \textit{Npv} &= \frac{tn}{tn+fn} \\ &= \frac{45}{45+0} = 100\% \end{aligned}$$

Dengan pengujian *K-Fold Cross Validation* dengan *number of validation* 3, sebanyak 173 data untuk diprediksi yang terdiri dari data training dan data testing diperoleh hasil false positif (FN) yaitu hasil dari titik pengangkutan yang sempit namun ternyata luas adalah 100%, hal ini berarti semua data diprediksi sebagai titik pengangkutan yang sempit namun ternyata luas.

Tabel 2. Tabel nilai *accuracy*, *sensitify*, *specitivity*, *ppv*, *npv*

Nilai(%)	
Accuracy	96.51%
Sensitify	100%
Specitivity	88.24%
Ppv	95.31%
Npv	100%

b. Evaluasi *ROC curve*

Dari tabel 4.2 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*). Sebesar 0.976% dengan tingkat diagnosa nilai akurasi excellent classification.



Gambar 4. Nilai AUC dalam grafik ROC uji coba pertama

Penelitian ini bertujuan untuk menguji keakuratan analisa prediksi retribusi persampahan/kebersihan dengan menggunakan Algoritma Decition Tree (C4.5). Data yang

dianalisa adalah data retribusi persampahan/kebersihan. Data tersebut diambil di Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur dan telah disetujui oleh semua pihak dalam kantor tersebut. Dari implementasi di atas, ditemukan jumlah akurasi dari algoritma Deciton Tree (C4.5) dengan *number of validation* 3 adalah 96.51%, dengan *number of validation* 5 adalah 94.77%, dan dengan *number of validation* 10 diperoleh nilai akurasi sebesar 95.33%.

Dari hasil pengujian di atas dengan dilakukan evaluasi secara *confusion matrix* ternyata terbukti bahwa algoritma Decition Tree (C4.5) dengan *number of validation* berbeda akan menghasilkan nilai akurasi yang berbeda. Berikut table hasil perbandingan akurasi dari masing-masing uji coba tersebut tersebut.

Tabel 4. Perbandingan akurasi

	Uji coba pertama	Uji coba kedua	Uji coba ketiga
Accuracy	96.51%	94.77%	95.33%

Dari tabel di atas terbukti bahwa hasil pengujian algoritma Deciton Tree (C4.5) dengan *number of validation* 3 memiliki akurasi tertinggi yaitu sebesar 96.51% dan hasil akurasi terendah ialah dengan menggunakan *number of validation* 5 dengan nilai akurasi 94.77%. Selisih nilai akurasi dari kedua algoritma tersebut adalah sekitar 2%.

5. Kesimpulan

Dari pengujian yang telah dilakukan sebelumnya dengan algoritma Decision Tree dengan *number of validation* yang berbeda, hasil yang diperoleh menunjukkan bahwa algoritma Decision Tree (C4.5) memperoleh hasil yang unggul. Dari tiga kali uji coba, ketiganya menunjukkan hasil akurasi dan nilai AUC yang tinggi. Hasil yang paling unggul yaitu dengan menggunakan *number of validation* 3, dengan nilai akurasi yang diperoleh sebesar 96.51% dan nilai AUC sebesar 0.976% dengan tingkat diagnosa nilai akurasi excellent classification. Dengan demikian dapat disimpulkan bahwa algoritma Decision Tree (C4.5) terbukti akurat dalam prediksi retribusi persampahan/kebersihan.

Berdasarkan pengujian yang telah dilakukan dan kesimpulan di atas, algoritma Decision Tree (C4.5) yang diterapkan sudah memberikan hasil yang baik. Namun ada beberapa hal yang dapat dicoba atau ditambahkan untuk laporan selanjutnya agar algoritma Decision Tree (C4.5) dapat menghasilkan model yang lebih baik lagi.

1. Menambahkan jumlah data yang lebih besar dan atribut yang lebih banyak, sehingga hasil pengukuran yang akan didapatkan lebih baik lagi.
2. Dapat menerapkan Particle Swarm Optimization (PSO) untuk mengoptimasi algoritma agar model yang didapat lebih akurat dan nilai akurasi lebih meningkat.

3. Dengan adanya laporan ini diharapkan bisa membantu Kantor Dinas Lingkungan Hidup dan Kebersihan dalam memprediksi jumlah retribusi persampahan/kebersihan setiap bulannya agar retribusi tetap lancar atau stabil.

Daftar Pustaka

- [1]. Eviciensia (2012). Penerapan Algoritma C4.5 berbasis Particle Swarm Optimization untuk Prediksi Hasil Pemilihan Umum Legislatif DPRD DKI Jakarta.
- [2]. Firmansyah (2011). Penerapan Algoritma Klasifikasi C4.5 untuk Penentuan Kelayakan Pemberian Kredit Koperasi
- [3]. <http://fairuzelsaid.com/konsep-data-mining/>
- [4]. Siti Masripah (2011). Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization untuk Evaluasi Penentuan Kelayakan Kredit Koperasi Syariah
- [5]. Khafi Heryandi Suradiradja (2012). Deteksi Transaksi Pencucian Uang Dengan Algoritma Klasifikasi C4.5
- [6]. Sigit Abdullah (2012). Penerapan Algoritma Decision Tree C4.5 untuk Diagnosa Penyakit Stroke dengan Klasifikasi Data Mining pada Rumah Sakit Santa Maria Pematang
- [7]. Khoirul Muarif (2013). Komparasi Pemodelan Data Menggunakan C4.5 dan C4.5 berbasis Particle Swarm Optimization untuk Memprediksi Kelulusan Mahasiswa