

PENERAPAN C4.5 DALAM PREDIKSI PENIPUAN PEMENANG UNDIAN BERHADIAH MENGGUNAKAN RAPID MINER

Ricky Aurelius Nurtanto Diaz

Program Studi Sistem Komputer, STMIK STIKOM BALI Jln.
Raya Puputan Renon No. 86 Telp. (0361) 244445 Denpasar
ricky.aurelius@gmail.com

Abstrak

Banyak masyarakat yang terjebak dengan berbagai undian berhadiah yang sebenarnya tidak pernah terjadi. Bahkan ada cara-cara tertentu dari pelaku penipuan yang terkesan memaksa korban untuk mengambil hadiah dari undian tersebut. Salah satu alasan banyaknya masyarakat yang masih terjebak dengan berbagai undian berhadiah yang palsu tersebut adalah kurangnya pengetahuan akan faktor-faktor penunjang atau petunjuk yang sebenarnya menjelaskan bahwa undian tersebut adalah palsu. Rapid Miner dengan dukungan metode Decision Tree (C4.5) dapat digunakan untuk memberikan gambaran bagaimana proses prediksi dapat dilakukan dengan berdasar pada kasus-kasus terdahulu dengan variabel penentu yang telah ditetapkan sebelumnya.

Kata kunci : C4.5 , Prediksi, Penipuan, Rapid Miner

1. Pendahuluan

1.1 Latar Belakang

Undian adalah suatu bentuk pengambilan hadiah dimana aturan utamanya adalah dengan cara menentukan suatu keputusan dengan cara pemilihan acak. Sebagai contoh adalah undian di mana peserta harus membeli sepotong tiket yang diberi nomor. Nomor tiket-tiket ini lantas secara acak ditarik dan nomor yang ditarik adalah nomor pemenang. Pemegang tiket dengan nomor pemenang ini berhak atas hadiah tertentu.

Masih banyak masyarakat yang terjebak dengan berbagai undian berhadiah yang sebenarnya tidak pernah terjadi. Bahkan ada cara-cara tertentu dari pelaku penipuan yang terkesan memaksa korban untuk mengambil hadiah dari undian tersebut. Salah satu alasan banyaknya masyarakat yang masih terjebak dengan berbagai undian berhadiah yang palsu tersebut adalah kurangnya pengetahuan akan faktor-faktor penunjang atau petunjuk yang sebenarnya menjelaskan bahwa undian tersebut adalah palsu.

Perkembangan teknologi informasi dimana salah satunya adalah metode data mining yang banyak digunakan oleh perusahaan-perusahaan besar, lambat laun semakin mudah ditemukan teknik dan aplikasinya bagi masyarakat kecil dan menengah.

Melihat segala kondisi tersebut, dirasa perlu melakukan penelitian untuk memprediksi penipuan berhadiah dengan memperhatikan beberapa faktor pendukung terjadinya penipuan

tersebut. Dengan demikian dipilihlah suatu metode yang dapat memberikan alternatif untuk menangani permasalahan yang telah dijabarkan. Untuk itu, maka dibuatlah sebuah Sistem Prediksi Penipuan Pemenang Undian Berhadiah Menggunakan Rapid Miner dengan menerapkan metode Decision Tree (C4.5).

1.2 Rumusan Masalah

Pengetahuan akan faktor-faktor penjelas adanya sebuah tindakan penipuan undian berhadiah adalah sebuah intisari penting bagi masyarakat. Rumusan masalah dalam penelitian ini adalah kebutuhan masyarakat untuk mengetahui faktor-faktor penting yang menunjukkan bahwa sebuah informasi yang diperoleh adalah informasi yang sebenarnya atau informasi yang dilatarbelakangi tindak penipuan.

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Membantu masyarakat dalam mengetahui atau memprediksi apakah sebuah informasi undian berhadiah tersebut benar atau tidak.
2. Menerapkan algoritma C4.5 dalam proses klasifikasi dengan memanfaatkan *tools* yaitu Rapid Miner.

1.4 Metodologi Penelitian

Dalam melakukan penelitian ini, maka metode yang digunakan adalah :

1. Studi Pustaka
2. Metode Observasi
Melihat dan mempelajari permasalahan yang ada dan erat kaitannya dengan objek yang diteliti yaitu informasi mengenai penipuan undian berhadiah.
3. Analisis Data
Melihat kemungkinan data yang akan dijadikan data training.

2. Tinjauan Pustaka

2.1 Data Mining

Data mining merupakan suatu istilah yang digunakan untuk kegiatan menemukan informasi atau pengetahuan di dalam basis data. *Data mining* adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan tiruan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Menurut Gartner Group, *data mining* adalah suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika

Terdapat banyak fungsi dan algoritma dalam data mining seperti metode statistik, pohon keputusan, *jaringan saraf*, *teori himpunan*, *regresi linier dan non linier dan sebagainya*. Algoritma data mining merupakan hal penting. Ketika berhadapan dengan data perilaku pelanggan, algoritma harus mampu secara efektif menghadapi data yang memiliki dimensional tinggi. Algoritma juga harus mampu bekerja dengan batasan dan aturan bisnis. [1]

2.2 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

- a. Pilih atribut sebagai akar.
- b. Buat cabang untuk tiap-tiap nilai.
- c. Bagi kasus dalam cabang.
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dan atribut-atribut yang ada. [2]

3. Hasil dan Pembahasan

3.1 Analisis Kebutuhan Input

Sebelum masuk pada tahap pembentukan tree, terlebih dahulu akan ditentukan variabel-variabel dasar dalam proses training, yaitu :

1. Provider
2. Waktu
3. Biaya
4. Cara menghubungi
5. Link Web Gratis
6. Contact Person
7. Jenis Hadiah

Sedangkan untuk variabel keputusan terdapat satu variabel yaitu : Penipuan.

Pada tabel 3.1 menunjukkan detail tujuh variabel dasar dan satu variabel keputusan yang digunakan :

Tabel 3.1 Variabel Training

No.	Nama Variabel	Sifat	Keterangan
1	Provider	Polynomial (Provider I, Provider X, Provider S, Provider E, Provider A)	Merupakan provider dari nomor orang yang menghubungi
2	Waktu	Polynomial 00.00– 06.00 06.00– 12.00 12.00 –18.00 18.00 –00.00	Merupakan waktu korban dihubungi
3	Biaya	Binomial 1. Ada 2. Tidak	Merupakan biaya yang disyaratkan untuk mengambil hadiah
4	Cara Menghubungi	Binomial 1. Telepon 2. SMS	Merupakan cara pelaku menghubungi calon korban
5	Link Web Gratis	Binomial 1. Ada 2. Tidak	Merupakan informasi tambahan pelaku penipuan dengan menyediakan Website tertentu
6	Contact Person	Binomial 1. Ada 2. Tidak	Merupakan nomor tertentu yang dapat digunakan calon korban untuk komunikasi dengan pelaku
7	Jenis Hadiah	Binomial 1. Uang 2. Kendaraan	Merupakan jenis hadiah yang diberikan
8	Penipuan	Binomial	Merupakan

		1. Ya 2. Tidak	variabel keputusan akhir
--	--	-------------------	--------------------------

3.2 Data Training

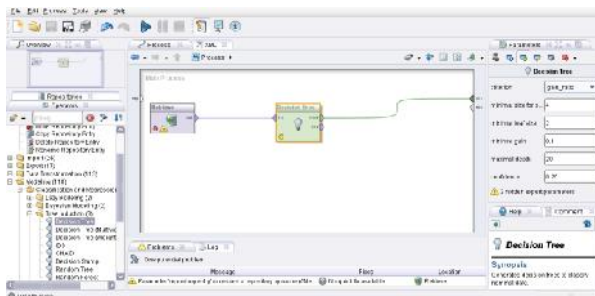
Setelah penentuan input, tahap berikutnya adalah mempersiapkan data *training* yang akan digunakan dalam pembentukan tree. Pada tahap ini digunakan 75 data *training*. Tabel 3.2 menampilkan contoh data *training* yang digunakan :

Tabel 3.2 *Data Training*

Provider	Waktu	biaya	Cara Menghubungi	Link Web Gratis	Contact Person	Jenis Hadiah	Penipuan
I	06.00 - 12.00	ada	sms	ada	ada	uang	ya
X	18.00 - 18.00	tidak	sms	tidak	ada	kendaraan	ya
S	00.00 - 00.00	tidak	telepon	ada	ada	kendaraan	ya
E	06.00 - 12.00	ada	telepon	tidak	ada	kendaraan	ya
A	18.00	ada	telepon	tidak	tidak	uang	tidak

3.3 Pembentukan Tree

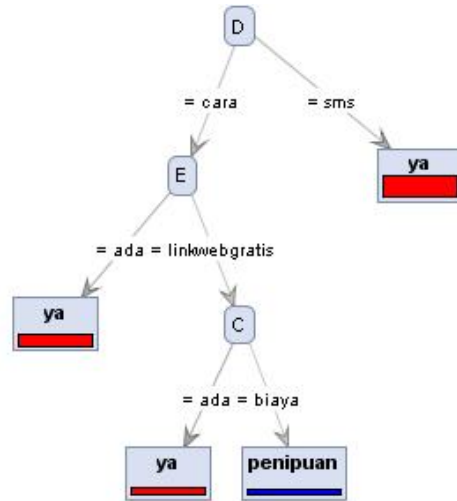
Proses pembentukan *tree* pada *Rapid Miner* dimulai dengan *import* data *training* dan dilanjutkan dengan proses modeling dengan cara memilih operator modeling yaitu C4.5 yang terdapat dalam direktori *Tree Induction*. Proses modeling dapat dilihat pada gambar 3.1 :



Gambar 3.1 *Proses Modelling*

3.4 Evaluasi Tree

Setelah data *Training* dijalankan menggunakan model algoritma C4.5 maka hasil terakhir yang diperoleh adalah sebut *Tree* yang dapat digunakan untuk melakukan prediksi penipuan undian berhadiah seperti yang terlihat pada gambar 3.2 :



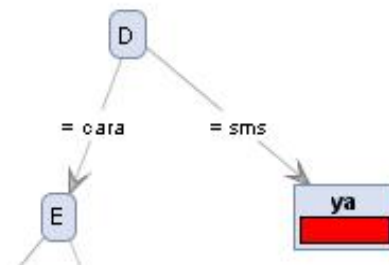
Gambar 3.2 *Tree Hasil Modelling*

Dari hasil pembentukan Tree diperoleh informasi sebagai berikut :

- Dari 7 atribut yang digunakan untuk pembentukan Tree, hanya terdapat 3 atribut yang berfungsi dalam *Decision Tree* yaitu : Cara Menghubungi, *Link Web Gratis* dan Biaya Tambahan
- Pada penerapannya, jika terdapat sebuah data Test :

- Provider** : I
- Waktu** : 06.00 - 12.00
- Biaya** : Ada
- Cara Menghubungi** : SMS
- Link Web** : Ada
- Contact Person** : Ada
- Hadiah** : Kendaraan
- Penipuan** : ?

Maka, proses penentuan prediksi hanya akan melihat 3 atribut diatas dimulai dari cara menghubungi. Untuk contoh kasus ini, karena cara menghubungi melalui SMS maka dapat dipastikan bahwa itu adalah sebuah penipuan seperti yang terlihat pada gambar 3.3.



4. Kesimpulan dan Saran

4.1 Kesimpulan

Dari penelitian ini dapat disimpulkan beberapa hal sebagai berikut :

1. Algoritma C4.5 cocok digunakan untuk membantu pihak-pihak yang berkepentingan dalam pengambilan keputusan salah satunya berupa teknik prediksi.
 2. Rapid Miner dapat digunakan untuk membentuk satu model pengolahan data dengan teknik atau algoritma tertentu.
 3. Atribut keputusan yang digunakan dapat disesuaikan dengan kebutuhan yang lain.
 4. Metode ini dapat diterapkan pada proses prediksi yang lain dengan menyesuaikan kriteria-kriteria penilaian yang akan digunakan.
2. Perangkat lunak pendukung dapat dikembangkan bagi pengguna yang ingin mengakses sistem menggunakan perangkat mobile, sehingga proses prediksi dapat berjalan dengan lebih efisien.

Daftar Pustaka:

- [1] Turban, 2005, *Decision Support Systems and Intelligent Systems*, Yogyakarta, Andi.
- [2] Sunjana, 2010, *Klasifikasi Data Nasabah Sebuah Asuransi Menggunakan Algoritma C4.5*. Seminar Nasional Aplikasi Teknologi Informasi 2010.
- [3] Khan Fahad Shahbaz, 2008, Data Mining in Oral Medicine Using Decision Trees, *International Journal of Biological and Life Sciences* 4:3.

4.2 Saran

Adapun saran yang mungkin dapat membantu dalam penelitian selanjutnya adalah :

1. Prediksi dapat dikembangkan menjadi lebih baik dengan menggunakan beberapa metode yang digabungkan dengan algoritma C4.5 untuk mencapai hasil yang akurat.