

Klasifikasi Pelanggan pada *Customer Churn Prediction* Menggunakan *Decision Tree*

Aldi Nurzahputra, Afifah Ratna Safitri, Much Aziz Muslim

FMIPA, Universitas Negeri Semarang, Semarang

aldinurzah96@gmail.com

Abstrak

Kompetisi dalam dunia perdagangan semakin meningkat. Perusahaan membutuhkan klasifikasi pelanggan untuk mengetahui tingkat loyalitas pelanggan. Data-data pelanggan yang tersedia dapat dimanfaatkan untuk mengklasifikasikan loyalitas pelanggan. Klasifikasi pelanggan dapat dilakukan dengan *Customer Churn Prediction*. Penentuan proses klasifikasi data dalam menemukan informasi dan pengetahuan dapat dilakukan dengan data mining. Salah satu metode yang digunakan pada *data mining* untuk prediksi adalah *decision tree*, representasi sederhana dari teknik klasifikasi. Dalam penelitian ini algoritma yang digunakan untuk membangun *decision tree* adalah algoritma C4.5. Data yang digunakan berjumlah 996 *data set sample* pelanggan dengan atribut gender, umur, metode pembayaran, dan jumlah transaksi. Tujuan dari penelitian ini adalah mengklasifikasi pelanggan dengan menggunakan *decision tree*. Klasifikasi dilakukan berdasarkan nilai *confidence* pelanggan termasuk loyal atau *churn*. Dari hasil klasifikasi, menunjukkan 61 pelanggan loyal dan 35 pelanggan *churn*, kemudian diakumulasikan total pelanggan loyal menjadi 636 pelanggan dan pelanggan *churn* 357 pelanggan.

Kata Kunci: *Customer Churn Prediction, Decision Tree, Algoritma C4.5*

PENDAHULUAN

Saat ini, iklim kompetisi dalam dunia perdagangan semakin meningkat. Di sisi lain perubahan lingkungan yang demikian pesat semakin mendukung kompetisi yang sedang terjadi saat ini. Salah satu tujuan utama aktivitas pemasaran seringkali dilihat dari pencapaian loyalitas pelanggan melalui strategi pemasaran. Loyalitas pelanggan merupakan bagian terpenting pada pengulangan pembelian pada pelanggan.

Loyalitas pelanggan memiliki korelasi yang positif dengan performa bisnis. Loyalitas pelanggan tidak hanya meningkatkan nilai dalam bisnis, tetapi juga dapat menarik pelanggan baru, (Aryani & Rosinta, 2010). Kepuasan pelanggan merupakan kunci dalam menciptakan loyalitas pelanggan.

Data-data pelanggan yang berjumlah besar yang tersimpan secara elektronik pada basis data perusahaan dapat dimanfaatkan untuk menemukan pola-pola perilaku dan karakteristik konsumen. Pengolahan yang tepat pada data-data tersebut dapat menghasilkan pengetahuan- pengetahuan yang bermanfaat untuk memprediksi pelanggan mana yang *churn* dan pelanggan mana yang *loyal*. Proses analisis data untuk menemukan informasi dan pengetahuan pada data yang sangat besar disebut *data mining*. Govindaraju *et al.* (2008). Salah satu metode yang digunakan pada *data mining* untuk prediksi adalah *decision tree*. *Decision Tree* merupakan representasi sederhana dari teknik klasifikasi yang merupakan proses pembelajaran suatu fungsi tujuan yang memetakan tiap himpunan atribut ke satu dari kelas yang didefinisikan sebelumnya.

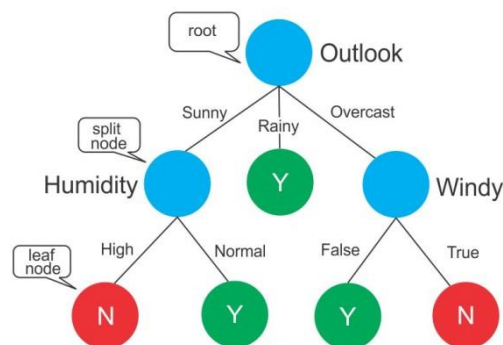
Rumusan permasalahan pada penelitian ini adalah bagaimana mengklasifikasi pelanggan pada *customer churn* menggunakan *Decision Tree*? Sedangkan Tujuan penelitian ini adalah dapat ditemukan pola pengklasifikasian pelanggan menggunakan *Decision Tree*.

METODE

Decision Tree

Data Mining merupakan proses yang diulang dan memerlukan interaksi manusia dalam prosesnya untuk menemukan pola atau model baru yang dapat digeneralisasi untuk masa yang akan datang, dan bermanfaat jika digunakan untuk melakukan suatu tindakan. Data mining berisi pencarian pola yang diinginkan dalam database besar untuk membantu dalam pengambilan keputusan diwaktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang kemudian dapat dipelajari dengan teliti dan bermanfaat. Dan salah satu teknik dari bagian data mining adalah klasifikasi, yaitu menentukan sebuah record data baru ke salah satu dari beberapa kategori atau klas yang telah difenisikan sebelumnya. (Azmi & Dahria, 2013)

Klasifikasi data mining adalah penempatan objek-objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Klasifikasi banyak digunakan untuk memprediksi kelas pada suatu label tertentu, yaitu dengan mengklasifikasi data (membangun model) berdasarkan training set dan nilai-nilai (label kelas) dalam mengklasifikasikan atribut tertentu dan menggunakannya dalam mengklasifikasikan data yang baru. Pohon keputusan biasanya digunakan untuk mendapatkan informasi untuk tujuan pengambilan sebuah keputusan. Pohon keputusan dimulai dengan sebuah root node (titik awal) yang digunakan oleh user untuk mengambil tindakan. Berdasarkan node root ini, user memecahkan leaf node sesuai dengan algoritma decision tree. (Meilina, 2015).



Gambar 1. Decision Tree

Decision Tree merupakan salah satu metode klasifikasi dan prediksi yang sangat kuat dan terkenal dalam penerapan *data mining*. Pada dasarnya *Decision Tree* mengubah data menjadi pohon keputusan (*decision tree*) dan aturan-aturan keputusan (*rule*). Keuntungan dalam metode ini adalah efektif dalam menganalisis sejumlah besar atribut dari data yang ada dan mudah dipahami oleh pengguna akhir.

Algoritma C4.5

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, salah satunya adalah algoritma C4.5. Algoritma C4.5 membuat pohon keputusan dari atas ke bawah, di mana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun. Keuntungan dalam metode ini adalah efektif dalam menganalisis sejumlah besar atribut dari data yang ada dan mudah dipahami oleh pengguna akhir.

Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

- a. Pilih atribut sebagai akar

- b. Buat cabang untuk masing-masing nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama. (Kamagi & Hansun, 2014)

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai Gain adalah dengan mencari nilai Entropi. Entropi digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar dari Entropi adalah sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n - pi * \log_2 pi \quad (2)$$

Keterangan:

S : Himpunan Kasus

n : Jumlah partisi S

pi: Proporsi dari Si terhadap S

HASIL DAN PEMBAHASAN

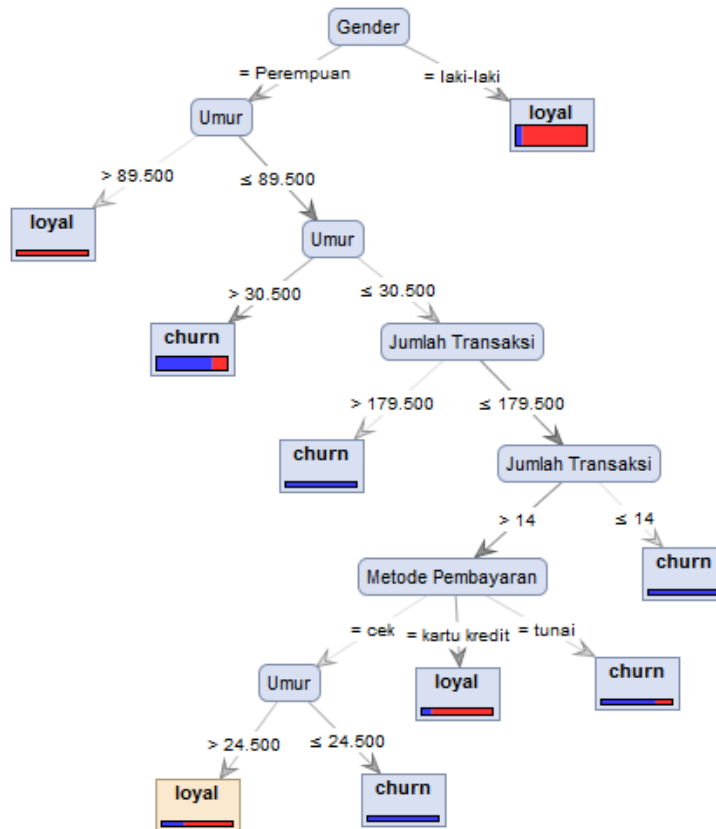
Dalam penelitian ini, menggunakan data sample set pelanggan terkait *customer churn* di suatu market yang berjumlah 996 pelanggan. Atribut yang digunakan adalah gender, umur, metode pembayaran, dan jumlah transaksi. Berdasarkan data set, klasifikasi pelanggan dilakukan berdasarkan data klasifikasi pelanggan yang telah tersedia. Dengan jumlah data yang besar, diperlukan prediksi untuk klasifikasi pelanggan.

Customer Churn	Jumlah Pelanggan
Loyal	578
Churn	322
Undefined	96
Total	996

Dalam mengklasifikasi pelanggan yang termasuk undefined, maka diperlukan pengenalan pola untuk melakukan prediksi klasifikasi pelanggan. Dengan menggunakan algoritma C4.5 dapat dibangun *decision tree* untuk data set tersebut.

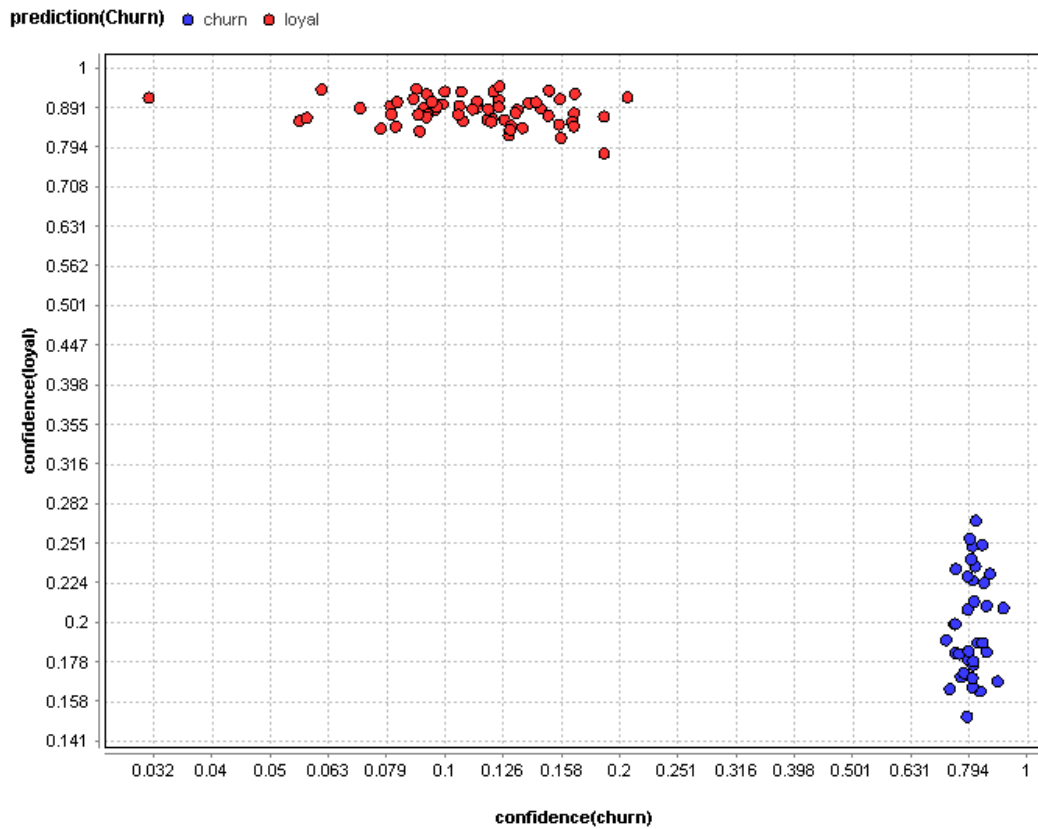
Pada penelitian ini, *decision tree* dibangun berdasarkan data customer yang telah ada dan digunakan untuk menentukan pelanggan yang masuk dalam klasifikasi undefined untuk diprediksi klasifikasi ke dalam loyal dan churn. Untuk membangun *decision tree*

dapat menggunakan tool data mining, yaitu RapidMiner Studio. Decision tree yang terbentuk dari data set tersebut dapat ditunjukkan pada Gambar 2.



Gambar 2. *Decision Tree Customer Churn*

Dengan menggunakan decision tree tersebut, maka dapat diprediksi untuk mengklasifikasikan pelanggan berdasarkan tingkat perbandingan *confidence* antara prediksi loyal dan churn. Klasifikasi loyal jika *confidence* loyal lebih tinggi daripada *confidence* churn dan klasifikasi churn jika *confidence* churn lebih tinggi daripada *confidence* loyal.



Gambar 3. Grafik *Confidence loyal* dan *churn* pelanggan

Sehingga hasil dari prediksi klasifikasi pelanggan berdasarkan decision tree tersebut diperoleh 61 pelanggan loyal dan 35 pelanggan churn.



Gambar 4. Klasifikasi Pelanggan

Kemudian dapat diakumulasikan dengan klasifikasi pelanggan yang telah ada, dapat di sajikan pada gambar 5.

Tabel 2. Klasifikasi Pelanggan Keseluruhan

Customer Churn	Jumlah Pelanggan
Loyal	639
Churn	357
Total	996

SIMPULAN

Berdasarkan hasil dan pembahasan, dengan menggunakan metode Decision Tree dan algoritma C4.5 dapat diprediksi klasifikasi pelanggan ke dalam pelanggan loyal atau pelanggan churn. Dalam mengklasifikasi, klasifikasi loyal jika confidence loyal lebih besar daripada confidence churn sedangkan klasifikasi churn jika confidence churn lebih besar daripada confidence loyal. Dari hasil klasifikasi, menunjukkan 61 pelanggan loyal dan 35 pelanggan churn, kemudian dapat diakumulasikan total pelanggan loyal adalah 636 pelanggan dan pelanggan churn 357 pelanggan.

DAFTAR PUSTAKA

- Aryani, Dwi., & Rosinta, Febrina., 2010. Pengaruh Kualitas Layanan terhadap Kepuasan Pelanggan dalam Membentuk Loyalitas Pelanggan. *Jurnal Ilmu Administrasi dan Organisasi*, 17(2), 114-126.
- Govindaraju, R., Simatapung, T., & Samadhi, A, T., 2008. Perancangan Sistem Prediksi Churn Pelanggan PT. Telekomunikasi Seluler Dengan Memanfaatkan Proses Data Mining. *Jurnal Informatika*. 9(1), 33-42.
- Azmi, Zulfian., & Dahria, Muhammad., 2013. *Decision Tree* Berbasis Algoritma untuk Pengambilan Keputusan. *Jurnal SAINTIKOM*. 12(3), 157-164.
- Meilina, Popy., 2015. Penerapan Data Mining dengan Metode Klasifikasi Menggunakan Decision Tree dan Regresi. *Jurnal Teknologi*. 7(1), 12-30.
- Kamagi, D. H., & Hansun, S., 2014. Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *ULTIMATICS*. 6(1), 15-18.