

topik utama dalam data mining atau machine learning, klasifikasi mengelompokkan data dimana data yang digunakan tersebut mempunyai kelas label atau target. *Klasifikasi* mempelajari pola-pola dari data historis (sekumpulan informasi –seperti ciri-ciri, variabel-variabel, fitur-fitur) dengan tujuan untuk menempatkan instans (object-object) baru (dengan label yang tak diketahui sebelumnya) ke dalam kelompok atau kelas masing-masing, dan teknik klasifikasi yang digunakan salah satunya adalah algoritma C4.5.

Manajemen kas yang optimal dan ketersediaan layanan merupakan salah satu faktor yang paling penting dalam bisnis layanan jaringan ATM [9]. Kebutuhan uang setiap mesin bisa berbeda-beda sesuai dengan lokasi mesin ATM, mesin ATM yang berada dipusat pembelanjaan akan lebih tinggi transaksinya dibandingkan ATM yang di tempatkan di lokasi yang bukan pusat pembelanjaan, selain itu disebabkan oleh faktor musiman misalnya pada bulan ramadhan, hari besar, hari libur, atau weekend sehingga banyak terjadi transaksi tunai yang melebihi dari transaksi yang terjadi pada hari biasa. Kebutuhan uang setiap ATM nantinya disesuaikan dengan tingkat fluktuatif penarikannya sehingga dapat meminimalisir banyaknya uang yang menganggur di ATM terlalu lama karena uang tersebut dapat diputar untuk proses bisnis yang lain oleh bank.

Algoritma C4.5 pada penelitian ini digunakan untuk memprediksi jumlah kas (uang) yang akan disediakan serta memprediksi waktu pengisian kas berikutnya untuk masing-masing ATM dengan mengambil sampel data beberapa mesin ATM di salah satu perbankan yang ada di Kota Padang. Penggunaan metode algoritma C4.5 ini diharapkan dapat membantu tim tambahan kas untuk memprediksi perkiraan waktu pengisian uang berikutnya sehingga dapat mengantisipasi uang *cashout* (uang habis) dalam caset/kas ATM dan diharapkan juga penelitian ini akan menghasilkan analisa untuk memprediksi jumlah kas ATM yang akan disediakan untuk pengisian uang berikutnya sehingga dapat meningkatkan kinerja dan kualitas pelayanan ATM.

2. Tinjauan Literatur

Dalam penulisan jurnal ini, beberapa hal-hal atau teori yang dikemukakan yang berkaitan dengan permasalahan dan ruang lingkup sebagai landasan dalam permbuat jurnal seperti : pemodelan dan UML,data minning, *knowledge discovery in Database* (KDD), pohon keputusan dan algoritma C4.5, yang bersumber dari buku, jurnal-jurnal penelitian yang telah lebih dahulu dilakukan oleh peneliti lain.

2.1. Pemodelalan dan UML

Unified Modeling Language (UML) merupakan salah satu alat bantu yang dapat digunakan dalam bahasa pemograman yang berorientasi objek. UML (*Unified Modeling Language*) yaitu keluarga notasi grafis yang didukung oleh meta-model tunggal, yang membantu pendeskripsian dan desain sistem perangkat lunak, khususnya sistem yang dibangun dengan menggunakan pemrograman berorientasi objek [4].

2.2. Data Mining

Data mining merupakan suatu alat yang memungkinkan para pengguna untuk mengakses secara cepat data dengan jumlah yang besar. Data Mining juga merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data dengan ukuran yang cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut. Data Mining dapat diterapkan pada berbagai bidang yang mempunyai sejumlah data. Menurut Daryl Pregibon bahwa “Data mining adalah campuran dari statistic, kecerdasan buatan, dan riset basis data” yang masih berkembang [1].

Data mining memungkinkan pemakai menemukan pengetahuan dalam data database yang tidak mungkin diketahui keberadaannya oleh pemakai, Istilah lain yang sering digunakan diantaranya knowledge discovery (mining) in databases (KDD).

2.3. Knowledge Discovery in Database

Knowledge Discovery in Database (KDD) adalah proses menentukan informasi yang berguna, Informasi ini terkandung dalam basis data yang berukuran besar yang sebelumnya tidak diketahui namun memiliki potensial dan bermanfaat, dengan proses yang bantu oleh komputer untuk menggali dan menganalisis sejumlah besar himpunan data dan mengekstrak informasi sehingga diperoleh pengetahuan yang berguna.

Pada proses KDD terdapat 5 tahapan (Nofriansyah dan Nurcahyo,2017)[1], yaitu sebagai berikut :

1. Seleksi Data (*Data selection*) : merupakan tahap awal dalam KDD yaitu penyeleksian sekumpulan dari data yang digunakan untuk proses data mining, yang nantinya data tersebut akan dikumpulkan dalam suatu berkas, tapi data tersebut terpisah dari basis data operasional.
2. Pemilihan Data (*Pre-processing/Cleaning*) : tahapan ini merupakan proses data *cleaning* meliputi yaitu pembuangan data-data yang redundan atau pembuangan duplikasi data, data-data yang tidak konsisten juga akan diperiksa pada tahap ini, serta kesalahan yang terdapat pada data diperbaiki ditahap ini seperti kesalahan ceta (*tipografi*). Selain itu dilakukan proses *enrichment* yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD seperti data atau informasi *eksternal*.
3. Transformasi (*Transformation*) : tahap ini data diubah kedalam bentuk yang sesuai untuk *mining coding* adalah proses transformasi pada data yang telah dipilih. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
4. *Data mining* : proses mencari pola atau informasi menarik dalam data yang terplih dengan menggunakan teknik atau metode tertentu. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dari proses KDD secara keseluruhan.
5. *Interpretation/Evaluation* : tahap ini mencakup pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

2.4. Pohon Keputusan

Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Menurut [1] pohon keputusan (*decision tree*) dapat membagi kumpulan data yang besar (*learning dataset*) menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan (*rules*), sehingga anggota himpunan hasil menjadi mirip satu dengan lainnya

dengan memperhatikan pada variabel tujuannya (*target variable*). *Decision tree* merupakan salah satu metoda klasifikasi *data mining* yang populer digunakan karena mudah diinterpretasikan oleh manusia dengan konsep dasar mengubah data menjadi pohon keputusan dan aturan-aturan keputusan [10].

Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan.



Gambar 1. Konsep Pohon Keputusan/*Decision Tree*

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan tree.

2.5. Algoritma C.45

Algoritma C4.5 atau disebut juga sebagai algoritma *decision tree* merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Algoritma ini termasuk metode *Data Mining*, yang merupakan proses menemukan pola dengan memilah-milah sejumlah data yang besar menggunakan teknologi pengenalan pola (Kamagi et.al, 2014).. Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama

Untuk menghitung algoritma *C4.5*, maka harus mencari nilai *entropy* dan *information gain*. Dimana nilai *entropy* adalah pengukuran ketidakpastian, di mana semakin tinggi *entropy*, maka semakin tinggi ketidakpastian[11]. Entropy (S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S. Entropy dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Entropy digunakan untuk mengukur ketidakaslisan S.

Rumus *entropy* dapat dilihat pada persamaan 1.

$$Entropi (R) = \sum_{i=0}^a - k_i * \log_2 k_i$$

Di mana:

- R = himpunan kasus.
- a = jumlah partisi R .
- k_i = proporsi R pada partisi ke- i

setelah menghitung *entropy* suatu kumpulan data, selanjutnya kita dapat mengukur seberapa efektifnya atribut dari data tersebut dalam mengklasifikasikan data, yang disebut juga dengan *information gain* [11]. Untuk menentukan nilai *gain* dapat dilihat dari persamaan 2 berikut ini :

$$\text{Gain}(T, W) = \text{Entropy}(R) \sum_{i=0}^a - \frac{|R_i|}{|R|} * \text{Entropy}(k_i)$$

Di mana :

W = atribut.

R = himpunan kasus.

$|R_i|$ = jumlah kasus pada partisi ke i

$|R|$ = jumlah kasus dalam S

3. Metodologi Penelitian

Metodologi penelitian yang digunakan meliputi Studi literatur dengan teknik mengumpulkan data atau informasi dari bagian terkait di perbankan, kemudian diolah dengan menggunakan konsep yang ada pada algoritma c.45, adapun tahapan seperti :

1. Tahap mengklasifikasikan data
Pada tahap ini setiap data dikelompokkan, dimana transaksi penarikan uang disetiap mesin ATM yang sudah ditotalkan dikelompokkan berdasarkan kriteria waktu dan jumlah penarikan.
2. Menghitung nilai entropi
Setelah memperoleh data analisis ketersediaan uang dimasing-masing ATM berdasarkan report penarikan uang secara tunai, langkah selanjutnya adalah menghitung nilai *entropy*, yang mana nilai *entropy* ini digunakan untuk mencari nilai *gain*.
3. Menghitung Nilai *Gain*
Setelah mendapat nilai *entropy* selanjutnya dicari nilai *gain*, nilai *gain* ini nantinya digunakan untuk menentukan atribut akar untuk pembuatan *decision tree*.
4. Menentukan Atribut *Root*
Setelah didapat nilai *gain* untuk setiap kriteria, selanjutnya akan ditentukan kriteria yang akan menjadi atribut *root* pada pembentukan *tree* dengan ketentuan kriteria dengan nilai *gain* yang tertinggi yang menjadi atribut *root*.
5. Membuat *Decision Tree*
Setelah didapat atribut *root*, kemudian dibuatlah pohon keputusan untuk kriteria yang dengan nilai *gain* tertinggi. Untuk mendapatkan pohon keputusan dari penentuan dari transaksi di mesin ATM yang telah dikelompokkan bersarkan kritria penarikan uang, maka harus dicari kembali nilai *entropy* dan *gain* dari setiap atribut.
6. Membuat cabang setiap kasus
Tahapan ini dilakukan dengan mengulangi proses perhitungan *Entropy* dan *Gain* untuk mencari cabang sampai semua kasus pada cabang memiliki kelas yang sama yaitu pada saat semua variabel telah menjadi bagian pohon keputusan (daun).
7. Membuat *rule* berdasarkan pohon keputusan
Setelah dibuat pohon keputusan sehingga semua atribut-atribut yang ada pada setiap kasus sudah tergambar akan tampak *rule* yang akan kita gunakan untuk pengambilan keputusan dalam penentuan persediaan uang di setian mesin ATM
8. Implementasi
Sistem yang dirancang menggunakan bahasa pemograman dengan metode-metode yang mengacu pada algoritma c.45

4. Hasil dan Diskusi

Setelah dilakukan tahapan analisis dan perancangan, untuk menampilkan hasil analisa dari penelitian ini apakah sesuai dengan tujuan, maka perlu dilakukannya implementasi dan pengujian pada penelitian ini, di mana mengimplementasikan hasil yang telah dianalisis dan dirancang sebelumnya. Adapun tahapan-tahapan implementasi tersebut berupa spesifikasi implementasi perangkat keras dan spesifikasi pengujian perangkat lunak.

Data pada penelitian ini telah dianalisa yang berasal dari sampel dari beberapa mesin ATM (12 mesin ATM) dari sebuah industri jasa keuangan yang ada di Provinsi Sumatera barat. Diketahui data set dapat dilihat dari tabel dibawah ini :

Tabel 2 : Dataset Mesin ATM

| No | TID | Pusat Keramaian | Tingkat Kerusakan | Listrik Padam | trx > 300/hari | Sering Cashout (< 2 hari) | Tambahan Kas ATM (< 2 hari) |
|----|-------|-----------------|-------------------|---------------|----------------|---------------------------|-----------------------------|
| 1 | 12045 | Ya | Jarang | Jarang | Tidak | Tidak | Ya |
| 2 | 12056 | Ya | Sering | Jarang | Tidak | Ya | Tidak |
| 3 | 12067 | Ya | Jarang | Jarang | Ya | Tidak | Ya |
| 4 | 12068 | Ya | Jarang | Sering | Ya | Tidak | Ya |
| 5 | 12069 | Ya | Jarang | tidak ada | Ya | Tidak | Tidak |
| 6 | 12071 | Tidak | Jarang | jarang | Tidak | Tidak | Ya |
| 7 | 12070 | Tidak | Sering | Sering | Ya | Ya | Tidak |
| 8 | 12078 | Ya | Jarang | Jarang | Ya | Tidak | Ya |
| 9 | 12080 | Tidak | Sering | Jarang | Ya | Ya | Tidak |
| 10 | 12077 | Ya | Sering | Sering | Ya | Ya | Tidak |
| 11 | 12089 | Ya | Jarang | Jarang | Tidak | Tidak | Ya |
| 12 | 12087 | Tidak | Jarang | Jarang | Tidak | Tidak | Ya |

Dari tabel diatas terdapat 7 atribut (tid, pusat keramaian, tingkat kerusakan, listrik padam, trx > 300/hari, cashout dan tambahan kas). Kelas Tambahan kas < 2 hari terdiri dari record “Ya” dan “tidak”, jumlah data sebanyak 12 terdiri dari Ya=7 dan tidak=5.

1. Entropy Total Dataset

$$S = \left\{ -\frac{7}{12} * (\log_2 \left(\frac{7}{12} \right)) \right\} + \left\{ -\frac{5}{12} * (\log_2 \left(\frac{5}{12} \right)) \right\} = 0,9798687107$$

2. Entropy dan gain tiap atribut

$$S_{[\text{Pusat keramaian - Ya}]} = \left\{ -\frac{5}{8} * (\log_2 \left(\frac{5}{8} \right)) \right\} + \left\{ -\frac{3}{8} * (\log_2 \left(\frac{3}{8} \right)) \right\} = 0,954434003$$

$$S_{[\text{Pusat keramaian- Tidak}]} = \left\{ -\frac{2}{4} * (\log_2 \left(\frac{2}{4} \right)) \right\} + \left\{ -\frac{2}{4} * (\log_2 \left(\frac{2}{4} \right)) \right\} = 1$$

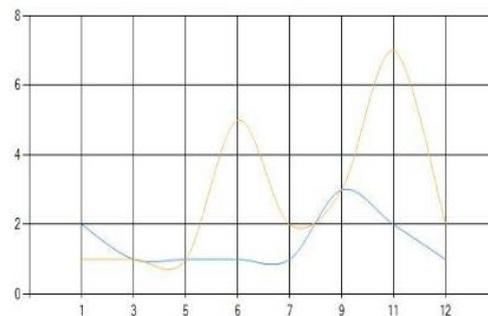
$$\text{Gain [Pusat keramaian]} = 0,9798687107 - \left\{ \frac{8}{12} * 0,954434003 \right\} + \left\{ \frac{4}{12} * 1 \right\} = 0,0102460421$$

$$\begin{aligned} &\text{Entropy [Listrik padam – Rutin]} \\ S &= \left\{ -\frac{6}{8} * (\log_2 \left(\frac{6}{8} \right)) \right\} + \left\{ -\frac{2}{8} * (\log_2 \left(\frac{2}{8} \right)) \right\} = 0,8112781243 \\ &\text{Entropy [Listrik padam – Jarang]} \\ S &= \left\{ -\frac{1}{3} * (\log_2 \left(\frac{1}{3} \right)) \right\} + \left\{ -\frac{2}{3} * (\log_2 \left(\frac{2}{3} \right)) \right\} = 0,9182958338 \\ &\text{Entropy [Listrik padam – Tidak ada]} = 0 \\ \text{Gain [Listrik]} &= 0,9798687107 - \left\{ \frac{8}{12} * 0,8112781243 \right\} + \left\{ \frac{3}{12} * 0,9182958338 \right\} + \left\{ \frac{1}{12} * 0 \right\} = 0,2094410656 \\ \\ &\text{Entropy [trx >300/ hari – Ya]} \\ S &= \left\{ -\frac{4}{5} * (\log_2 \left(\frac{4}{5} \right)) \right\} + \left\{ -\frac{1}{5} * (\log_2 \left(\frac{1}{5} \right)) \right\} = 0,7219280947 \\ &\text{Entropy [trx >300/ hari – Tidak]} \\ S &= \left\{ -\frac{3}{7} * (\log_2 \left(\frac{3}{7} \right)) \right\} + \left\{ -\frac{4}{7} * (\log_2 \left(\frac{4}{7} \right)) \right\} = 0,9852281359 \\ \text{Gain [trx >300/ hari]} &= 0,9798687107 - \left\{ \frac{5}{12} * 0,7219280947 \right\} + \left\{ \frac{7}{12} * 0,9852281359 \right\} = 0,0102460421 \\ \\ &\text{Entropy [Cash out – Ya]} \\ S &= \left\{ -\frac{7}{8} * (\log_2 \left(\frac{7}{8} \right)) \right\} + \left\{ -\frac{1}{8} * (\log_2 \left(\frac{1}{8} \right)) \right\} = 0,5435644433 \\ &\text{Entropy [Cash out – Tidak]} = 0 \\ \text{Gain [Cash out]} &= 0,9798687107 - \left\{ \frac{8}{12} * 0,5435644433 \right\} + \left\{ \frac{4}{12} * 0 \right\} = 0,6174924154 \end{aligned}$$

Hitung total jumlah data untuk setiap atribut nilai kelas =
 Cashout rendah = 8
 Cashout tinggi = 4

TID :12071
 Cashout rendah = Ya
 Cashout tinggi = Tidak
 trx > 300 = tidak
 Pusat keramaian = TI...
 Tambahan kas ATM <2 hari = Tid...
 Caset kosong dalam <2= 1

SIMPAN KELUAR CETAK



Gambar 6 : Form prediksi

5. Kesimpulan

Kesimpulan yang dapat dikemukakan dari analisa ketersediaan uang di mesin ATM menggunakan Algoritma C4.5 berdasarkan hasil analisis data dan pengujian dapat ditarik beberapa kesimpulan sebagai berikut :

- a. Analisa ketersediaan uang menggunakan algoritma C4.5 dapat membantu tim tambahan kas untuk memprediksi perkiraan waktu untuk pengisian uang berikutnya, sehingga dapat mengantisipasi uang cashout (uang habis) dalam caset/kas ATM.
- b. Dengan adanya analisa ketersediaan uang dengan menggunakan algoritma C4.5 tim tambahan kas dapat memprediksi jumlah kas ATM yang akan disediakan untuk pengisian uang berikutnya.

Referensi

- [1] Nofriansyah, D., And Nurcahyo, G, W. (2017). “*Algoritma Data Mining Dan Pengujian*”.
- [2] Gorunescu, F. (2011). *Data Mining : Concepts, Models and Techniques*. New York: Springer-Verlag.
- [3] Tampubolon, K., Saragih, H., Reza, B., Epicentrum, K., Asosiasi, A., & Apriori, A. (2013). Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan. *Informasi Dan Teknologi Ilmiah*, 93.
- [4] Fowler, Martin, ” *UML Distilled : Panduan Singkat Bahasa Pemodelan Objek Standar.*” Yogyakarta: Andi,2005.
- [5] Tampubolon, K., Saragih, H., Reza, B., Epicentrum, K., Asosiasi, A., & Apriori, A. (2013). Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan. *Informasi Dan Teknologi Ilmiah*, 93–106. <https://doi.org/2339-210X>.
- [6] Witten, I. H., Frank, E., Hall, M. A., 2011, *Data Mining Practical Machine Learning Tools and Techniques* (3rd ed). USA: Elsevier.
- [7] Berry, Michael J.A dan Gordon S. Linoff., 2004, *Data Mining Techniques for Marketing, Sales, Customer Relationship Management*, Second Edition, Willey Publishing, Inc.
- [8] Kamagi, et.al, 2014 *Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa*. Jurnal ULTIMATICS – Vol. VI, No. 1, ISSN : 2085-4552.
- [9] D. Alverina, A. R. Chrismanto, and R. G. Santosa. (2018), "*Perbandingan Akurasi Algoritma C4.5 dan CART dalam Memprediksi Kategori Indeks Prestasi Mahasiswa*," Jurnal Teknologi dan Sistem Komputer, vol. 6, no. 2, Apr. 2018. Doi: 10.14710/jtsiskom.6.2.2018.76-83.