
BAB II

LANDASAN TEORI

A. Data Mining

Data Mining atau Penambangan Data merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data. *Data mining* juga disebut *knowledge discovery in database (KDD)*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam *data set* yang berukuran besar, keluaran dari data mining dapat dipakai untuk memperbaiki pengambilan keputusan di masa depan. Secara sederhana, *data mining* atau penambangan data dapat diartikan sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya. Pengetahuan yang dihasilkan dari proses *data mining* harus baru, mudah dimengerti, dan bermanfaat. Dalam *data mining*, data disimpan secara elektronik dan diproses secara otomatis oleh komputer menggunakan teknik dan perhitungan tertentu [5].

Berikut ini merupakan teknik data mining yang sering diterapkan yaitu [6] :

1. *Clustering*

Analisis *Clustering* merupakan bagian dari proses mengidentifikasi *data set* yang mirip antara satu sama lain, untuk memahami perbedaan serta kesamaan data. Tujuan dari penggunaan teknik *Clustering* ini yaitu guna dibuat kelompok data atau objek ke dalam kumpulan cluster sehingga tiap *cluster* akan berisi data-data yang cukup mirip.

2. *Association Rule (Market Basket Analysis)*

Teknik ini dapat digunakan untuk memungkinkan penemuan hubungan yang menarik antara berbagai variabel dalam database yang besar. *Association rule learning* akan mengungkap pola yang belum

terlihat pada data yang dapat diterapkan untuk mengidentifikasi variabel dalam data dan kejadian bersama dari berbagai variabel yang muncul dengan frekuensi terbesar.

3. *Classification*

Analisis klasifikasi merupakan proses yang sistematis digunakan untuk memperoleh informasi dan sekelompok model yang relevan dari data set. Analisis klasifikasi akan membantu mengidentifikasi kategori atau kelas-kelas yang dimiliki data, sehingga analisis klasifikasi terkait erat dengan analisis *cluster* dikarenakan klasifikasi dapat digunakan untuk data *cluster*.

4. *Regression (Predictive)*

Analisis regresi merupakan teknik dalam mendefinisikan ketergantungan antar variabel. Analisis regresi dapat menunjukkan bahwa satu variabel tergantung pada variabel lain tetapi tidak sebaliknya.

B. Algoritma Iterative Dichotomiser Three (ID3)

Algoritma Iterative Dichotomiser Three (ID3) merupakan salah satu pendekatan klasifikasi dalam data mining dengan menciptakan pohon berdasarkan atribut yang ada untuk mengatasi suatu permasalahan. Pohon keputusan didefinisikan sebagai sebuah pohon dimana masing-masing cabang dari simpul merepresentasikan alternatif pilihan dan masing-masing ujung simpul atau *node* merepresentasikan keputusan [7]. Algoritma ID3 juga dapat diartikan sebagai sebuah metode yang digunakan untuk membangkitkan pohon keputusan yang akan menggali informasi pada data mining dengan cara mengklasifikasikan [10]. Algoritma ID3 mengambil konsep dan teori informasi dimana pemilihan atribut untuk membentuk pohon tersebut dilakukan dengan properti statistik yang disebut *information gain*, nilai *gain* digunakan untuk mengukur kualitas suatu atribut dalam memisahkan *training example* ke dalam kelas target [7].

Atribut dengan informasi tertinggi akan dipilih, dengan tujuan untuk mendefinisikan *gain* menggunakan ide dari teori informasi yang disebut

entropy [8]. *Entropy* adalah jumlah dari informasi yang terdapat pada atribut yang diperoleh dari rumus [7] :

$$Entropy(S) = -P_+ \log_2 P_+ - P_- \log_2 P_- \quad (1)$$

Keterangan [9] :

S : Ruang (data) sampel yang digunakan untuk *training*.

P_+ : Probabilitas/ jumlah yang bersolusi positif (mendukung) pada data sampel untuk kriteria tertentu

P_- : Probabilitas/ jumlah yang bersolusi negatif (tidak mendukung) pada data sampel untuk kriteria tertentu

Dalam penentuan akar (*root*) dan cabang, ID3 menggunakan nilai *information gain* terbesar dari atribut-atribut yang ada, *information gain* didapatkan dari rumus [7] :

$$Gain(S, A) = Entropy(S) - \sum \frac{|S_v|}{S} Entropy(S_v) \quad (2)$$

Keterangan [9] :

A : Atribut

V : Menyatakan suatu nilai yang mungkin untuk atribut

$|S_v|$: Jumlah sampel untuk nilai v

$|S|$: Jumlah seluruh sampel data

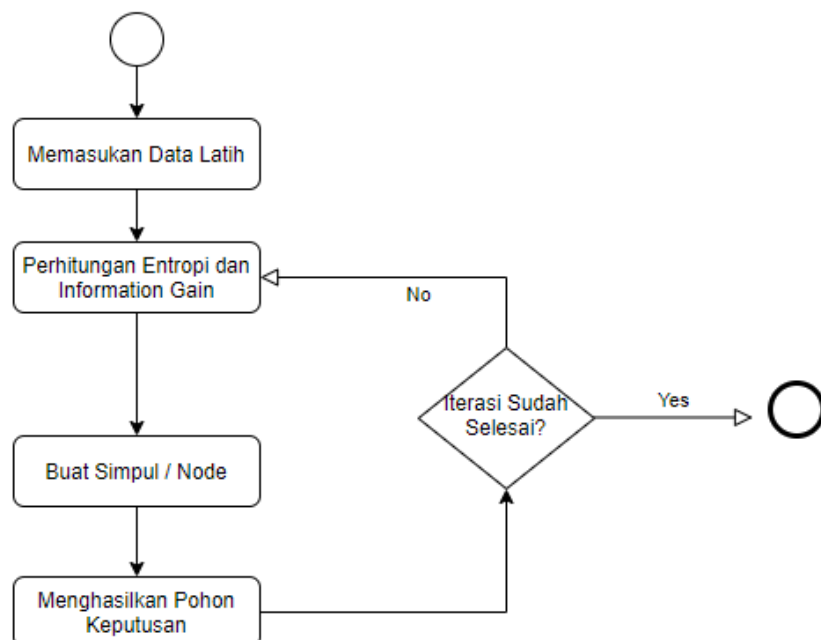
$Entropy(S_v)$: *Entropy* untuk sampel-sampel yang memiliki nilai v

Secara ringkas cara kerja algoritma ID3 dalam pemilihan atribut dengan menggunakan *information gain* dapat digambarkan sebagai berikut [8] :

- a. Pilih atribut dimana nilai *information gain* terbesar.
- b. Buat simpul yang berisi atribut tersebut.
- c. Proses perhitungan *information gain* akan terus dilaksanakan sampai semua data telah masuk ke dalam kelas yang sama. Atribut yang telah dipilih tidak diikuti lagi dalam perhitungan nilai *information gain*.

Langkah-langkah dari pembentukan pohon keputusan dengan *Algoritma ID3* sebagai berikut [9] :

1. Memasukkan data klasterisasi, ini adalah syarat yang paling penting dari pembentukan algoritma ID3. Jadi pola keputusan dari data klasterisasi yang akan menjadi acuan dalam pembentukan pohon keputusan.
2. Melakukan perhitungan *entropy* dan *information gain*, ini adalah tahap yang digunakan untuk menemukan simpul akar dari atribut yang ada sampai pohon berhenti terbentuk.
3. Setelah pohon sudah terbentuk, dibuatlah sebuah rule yang akan menjadi aturan dalam pohon keputusan.



Gambar 2.1 Diagram Alir *Algoritma ID3*

Hasil dari perhitungan entropy dan gain akan menghasilkan suatu pohon keputusan (*decision tree*). Konsep yang digunakan oleh pohon keputusan untuk mengubah data menjadi suatu keputusan dan aturan-aturan keputusan (*rule*) [10].

C. Pengaduan Masyarakat

Pengaduan Masyarakat merupakan pernyataan ketidakpuasan dan memiliki bentuk tertulis, lisan, ataupun melalui bahasa tubuh mengenai pelayanan, tindakan/kekurangan, tindakan yang dilakukan oleh instansi penyedia pelayanan dan dirasakan oleh masyarakat atau pengguna pelayanan. Atau pengaduan juga dapat diartikan sebagai ungkapan publik yang bisa timbul dikarenakan adanya ketidakpuasan publik atas suatu pelayanan atau produk yang dimiliki. Dan pengaduan semestinya diperuntukan untuk memperbaiki kualitas pelayanan masyarakat. Berdasarkan Peraturan Presiden nomor 76 tahun 2013 tentang Pengelolaan Pengaduan Pelayanan Publik pasal 1 ayat 8, Pengaduan adalah penyampaian keluhan yang disampaikan kepada pengadu kepada pengelola pengaduan pelayanan publik atas pelayanan pelaksana yang tidak sesuai dengan standar pelayanan, atau pengabaian kewajiban atau pelanggaran larangan oleh penyelenggara. Pengaduan sering dipandang sebagai hal buruk bagi penyelenggara pelayanan publik, sehingga banyak pihak berusaha untuk menutupi atau mengabaikannya, padahal pengaduan dapat dijadikan sebagai peringatan bermanfaat untuk meningkatkan kualitas kinerja penyelenggara pelayanan publik. Penanganan pengaduan dapat menjadi salah satu instrumen untuk memperbaiki hubungan antara pemerintah dengan masyarakat seperti transparansi, responsivitas, dan akuntabilitas [3].

D. Pelayanan Publik

Pelayanan publik menurut Undang-undang nomor 25 tahun 2009 tentang pelayanan publik, pasal 1 ayat 1 diartikan sebagai kegiatan atau rangkaian kegiatan dalam rangka pemenuhan kebutuhan pelayanan sesuai dengan peraturan perundang-undangan bagi setiap warga negara dan penduduk atas barang, jasa atau pelayanan administratif yang diselenggarakan oleh penyelenggara pelayanan publik. Oleh karena itu setiap warga negara berhak untuk mendapatkan pelayanan publik yang berkualitas, selain itu masyarakat juga dapat berpartisipasi dalam menjaga

dan meningkatkan kualitas pelayanan publik tersebut. Pelayanan yang berkualitas akan menjadi tolak ukur bagi kinerja instansi pemerintah, tolak ukur pelayanan yang berkualitas dapat dicermati melalui bukti langsung (*tangibles*), kehandalan (*reliability*), kepekaan/daya tanggap (*responsiveness*), jaminan (*assurance*), dan kemampuan untuk memahami kebutuhan pelanggan (*empathy*). Prinsip-prinsip tersebut menjadi faktor pembangunan dalam mewujudkan penyelenggaraan pelayanan publik. Pelayanan publik juga harus responsif terhadap perubahan-perubahan yang terjadi dan tuntutan masyarakat yang menjadi penting bagi peningkatan kualitas pelayanan publik [3].

E. Rapidminer

Rapidminer adalah platform perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif (*predictive analytics*). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan prototype dengan cepat, dan pengembangan aplikasi serta mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan [11]. Rapidminer dilengkapi dengan satu *processor* logika dan 10.000 baris data, dan berada dibawah lisensi AGPL [5]. Untuk mengoperasikan Rapidminer dapat berjalan di semua operating system [6].

Rapidminer ditujukan khusus untuk pengelolaan data dalam jumlah yang banyak, didalamnya sudah terdapat banyak model yang dapat digunakan seperti, Neural Network, ID3, Decision tree, dsb. Dan metode yang terdapat pada rapidminer cukup banyak, diantaranya, klustering, asosiasi, klasifikasi dll [6].