

## BAB 1

### PENDAHULUAN

#### 1.1 Latar Belakang

Metode yang sering digunakan dalam menganalisis suatu data penelitian adalah metode pengelompokan (*clustering*) dan klasifikasi (*classification*). *Clustering* merupakan pengelompokan data tanpa berdasarkan kelas data tertentu, sehingga dapat pula dipakai untuk memberikan label pada kelas data yang belum diketahui. Prinsip *clustering* adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimalkan kesamaan antar kelas atau *cluster*. Algoritma yang dapat digunakan dalam melakukan proses *clustering*, diantaranya: algoritma *K-Means*, *Expectation Maximization*, *Lingo* (Sumathi, 2006).

Klasifikasi (*classification*) adalah proses dari pembangunan terhadap suatu model yang mengklasifikasikan suatu objek sesuai dengan atribut-atributnya (Elly, Mira, & Alfian, 2015). Pohon klasifikasi merupakan suatu metode yang digunakan untuk mengklasifikasikan data. Pohon klasifikasi yang dibentuk berasal dari hasil penyekatan data secara berulang, dimana kelas dan nilai peubah penjelas dari setiap sampel pada data sudah diketahui (Loh dan Shih, 1997).

Klasifikasi banyak digunakan di berbagai bidang, misalnya di bidang pemasaran dan bidang kesehatan. Klasifikasi dibedakan menjadi dua kelompok, yaitu: parametrik dan non parametrik. Adapun kelompok parametrik diantaranya adalah regresi logistik dan diskriminan yang memerlukan asumsi. Sedangkan kelompok non parametrik diantaranya adalah *Exhaustive CHAID* dan *QUEST* (Schowengerdt, 2012).

## IR – PERPUSTAKAAN UNIVERSITAS AIRLANGGA

Ada beberapa algoritma yang dapat digunakan untuk melakukan proses klasifikasi, diantaranya adalah pohon keputusan (*Decision Tree*), *Naive Bayes*, *K-NN*, *Support Vector Machine*. Keempat metode ini dapat digunakan untuk klasifikasi, tetapi berasal dari rumpun yang berbeda (Prasetyo, 2013).

Pohon keputusan (*Decision Tree*) merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) dimana setiap *node* mempresentasikan atribut, cabangnya mempresentasikan nilai dari atribut, dan daun mempresentasikan kelas. *Node* yang paling atas dari *Decision Tree* disebut *root*. Algoritma yang dapat digunakan untuk melakukan proses *Decision Tree*, diantaranya: ID3, C4.5, C5.0, CHAID, CART, *Exhaustive* CHAID dan QUEST (Prasetyo, 2013).

Metode klasifikasi *Decision Tree* memiliki keunggulan diantaranya:

1. Daerah pengambilan keputusan pada saat analisis yang sebelumnya kompleks dan sangat global, dapat diubah menjadi lebih simpel dan spesifik.
2. Dapat mengeliminasi perhitungan-perhitungan yang tidak diperlukan, karena ketika menggunakan metode *Decision Tree*, maka sampel diuji hanya berdasarkan kriteria atau kelas tertentu.
3. Lebih fleksibel dalam memilih fitur dari *internal node* yang berbeda, sehingga dapat meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode perhitungan satu tahap yang lebih konvensional.
4. Metode *Decision Tree* dapat menghindari munculnya permasalahan mengestimasi baik distribusi dimensi tinggi maupun parameter tertentu dari distribusi kelas dianalisis multivariat dengan kriteria dan kelas yang sangat

## IR – PERPUSTAKAAN UNIVERSITAS AIRLANGGA

banyak. Cara *Decision Tree* dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap *internal node* tanpa mengurangi kualitas keputusan yang dihasilkan (Sumathi, 2006).

Regresi logistik merupakan metode yang dapat digunakan untuk melihat hubungan antara variabel dependen (Y) yang bersifat kategorik (kualitatif) dan dapat berbentuk *dikotomus* atau *polychotomous* dengan variabel-variabel independen (X) yang bersifat nominal atau ordinal (kualitatif) maupun interval atau rasio (kuantitatif). Variabel independen yang bersifat kualitatif digunakan variabel *dummy*, sedangkan yang bersifat kuantitatif dapat dijabarkan secara langsung (Hosmer dan Lemeshow, 2000).

Analisis diskriminan merupakan salah satu teknik statistik yang bisa digunakan dalam bentuk dependensi (hubungan antar variabel dimana sudah dapat dibedakan variabel dependen yang bersifat kategorik dan variabel independen yang bersifat numerik). Analisis diskriminan bertujuan untuk mengklasifikasikan suatu individu atau observasi ke dalam kelompok yang saling bebas (*mutually exclusive*) (Malhotra, 2005).

*Exhaustive CHAID* merupakan metode eksplorasi untuk mengklasifikasikan data dengan cara membangun pohon klasifikasi yang dapat memberikan informasi berupa variabel independen yang berpengaruh secara signifikan terhadap variabel dependen (Bagozzi, 1994). Sedangkan *QUEST* merupakan salah satu bentuk metode pohon klasifikasi yang menghasilkan pohon biner yang diterapkan pada data dengan variabel respon (Y) berupa kategorik dan variabel prediktor (X) berupa kategorik maupun numerik (Loh dan Shih, 1997).

## IR – PERPUSTAKAAN UNIVERSITAS AIRLANGGA

Metode *Exhaustive* CHAID dan QUEST dapat digunakan untuk data dengan variabel berskala kategorik, sehingga secara tidak langsung kedua metode ini dapat dibandingkan. Untuk membandingkan kedua metode tersebut, digunakan hasil pengobatan data kasus TBC paru di Puskesmas Jati.

TBC paru merupakan penyakit menular yang disebabkan oleh *Mycobacterium tuberculosis*, yang dapat menyerang paru dan organ lainnya. TBC paru sampai dengan saat ini masih merupakan salah satu masalah kesehatan masyarakat di dunia walaupun upaya penanggulangan TBC paru telah dilaksanakan di banyak negara sejak tahun 1995. Target program Penanggulangan TBC paru nasional yaitu eliminasi pada tahun 2035 dan Indonesia bebas TBC paru tahun 2050 (Permenkes, 2016).

Data dari WHO (2017) menyebutkan bahwa setiap tahun telah ditemukan jutaan orang di dunia yang terinfeksi penyakit TBC paru. Salah satu negara dengan penderita TBC paru terbanyak di dunia adalah Indonesia. Di Indonesia, angka insiden TBC paru sebanyak 391 per 100.000 penduduk dan angka kematian 42 per 100.000 penduduk.

Di Indonesia sebesar 360.565 kasus TBC paru telah ditemukan pada tahun 2016 dan kasus tersebut meningkat menjadi 425.089 kasus pada tahun 2017. Jumlah penduduk yang dilaporkan dengan kasus tertinggi (43% dari seluruh jumlah kasus TBC paru di Indonesia) yaitu terdapat di Provinsi Jawa Barat, Jawa Tengah dan Jawa Timur (Kemenkes RI, 2018).

Angka notifikasi semua kasus TBC paru penduduk Provinsi Jawa Tengah tahun 2017 sebesar 132,9 per 100.000 penduduk, hal ini menunjukkan bahwa

## IR – PERPUSTAKAAN UNIVERSITAS AIRLANGGA

penemuan kasus TBC paru di Jawa Tengah mengalami peningkatan dibandingkan tahun 2016 sebesar 118 per 100.000 penduduk (Dinkes Jateng, 2018).

Data kasus TBC paru yang diperoleh dari 19 Puskesmas di wilayah Kabupaten Kudus pada tahun 2014 menyebutkan bahwa sebanyak 495 orang terinfeksi penyakit TBC paru (Dinkes Kab. Kudus, 2016).

Di wilayah Puskesmas Jati data kasus TBC paru pada tahun 2013 sebanyak 54 pasien meliputi sebanyak 49 pasien dengan BTA positif dan sebanyak 5 pasien dengan BTA negatif; tahun 2014 sebanyak 39 pasien meliputi sebanyak 31 pasien dengan BTA positif dan sebanyak 8 pasien dengan BTA negatif; tahun 2015 sebanyak 52 pasien meliputi sebanyak 45 pasien dengan BTA positif dan sebanyak 7 pasien dengan BTA negatif; tahun 2016 sebanyak 46 pasien meliputi sebanyak 42 pasien dengan BTA positif dan sebanyak 4 pasien dengan BTA negatif; tahun 2017 sebanyak 64 pasien meliputi sebanyak 56 pasien dengan BTA positif dan sebanyak 8 pasien dengan BTA negatif; dan tahun 2018 sebanyak 58 pasien meliputi sebanyak 44 pasien dengan BTA positif dan sebanyak 14 pasien dengan BTA negatif. Dilihat dari data tersebut, kejadian TBC paru di wilayah Puskesmas Jati masih mengalami kenaikan dan penurunan (UPK Puskesmas Jati).

Penelitian ini menggunakan metode *Exhaustive* CHAID dan QUEST yang akan diterapkan pada hasil pengobatan data kasus TBC paru di wilayah Puskesmas Jati Kudus, kemudian hasil kedua metode yang digunakan akan dibandingkan untuk mengetahui metode mana yang lebih tepat dalam mengklasifikasikan data penelitian.

## 1.2 Kajian Masalah

Klasifikasi adalah proses pengelompokan data menjadi kelas-kelas sesuai kriteria yang telah ditentukan. Algoritma yang dapat digunakan untuk melakukan proses klasifikasi, diantaranya adalah pohon keputusan (*Decision Tree*), *naive Bayes*, *K-NN*, *Support Vector Machine* (Prasetyo, 2013).

Pohon keputusan (*Decision Tree*) merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) dimana setiap *node* mempresentasikan atribut, cabangnya mempresentasikan nilai dari atribut, dan daun mempresentasikan kelas. Algoritma yang dapat digunakan untuk melakukan proses *Decision Tree*, diantaranya: ID3, C4.5, C5.0, CHAID, CART, *Exhaustive CHAID* dan QUEST (Prasetyo, 2013).

Metode yang digunakan dalam penelitian ini adalah metode *Exhaustive CHAID* dan QUEST. *Exhaustive CHAID* mempunyai tiga tahapan, yaitu: tahap penggabungan (*merging*), tahap pemisahan (*splitting*), dan tahap penghentian (*stopping*). Diagram pohon pada *Exhaustive CHAID* dimulai dari simpul akar (*root simpul*) melalui tahap penggabungan, pemisahan, dan penghentian pada setiap simpul yang terbentuk dan langkah tersebut dilakukan secara berulang (Bagozzi, 1994). QUEST juga mempunyai tiga tahapan, yaitu: pemilihan variabel, penentuan simpul penyekat, dan penghentian pembentukan pohon (Abdurrahman, dkk 2014).

Konsep dasar *Exhaustive CHAID* adalah melakukan penggabungan variabel independen kategori yang tidak signifikan terhadap variabel dependennya, menggunakan uji independensi *Chi Square* yang menghasilkan nilai statistik *Chi Square* dan *p-value*. Kemudian *p-value* dikalikan dengan koreksi *Bonferroni* untuk

## IR – PERPUSTAKAAN UNIVERSITAS AIRLANGGA

mendapatkan *p-value* terkoreksi *Bonferroni*. Koreksi *Bonferroni* pada *Exhaustive CHAID* digunakan untuk penyesuaian *p-value* dari variabel prediktor yang telah terjadi pengurangan kategori. Penyesuaian ini dilakukan dengan melakukan pengalihan antara *p-value* sebelum terjadi pengurangan dengan pengali *Bonferroni* (Sholehudin, 2014).

Konsep dasar QUEST adalah memilih *p-value* terkecil dari hasil uji setiap variabel prediktor terhadap variabel respon, kemudian dibandingkan dengan koreksi *Bonferroni*. Koreksi *Bonferroni* adalah suatu koreksi yang menggunakan tingkat kepercayaan/tingkat signifikansi lebih kecil ketika beberapa uji dilakukan. Ketika  $n$  uji dilakukan, maka masing-masing uji menggunakan tingkat kepercayaan/tingkat signifikansi  $\frac{\alpha}{n}$ , dimana  $\alpha$  adalah tingkat signifikansi/tingkat kepercayaan (Grabczewski, 2014).

Penelitian yang menggunakan metode *Exhaustive CHAID*, antara lain: penelitian yang disusun oleh Suyanto (2014) menyimpulkan bahwa faktor-faktor yang berpengaruh secara signifikan terhadap keputusan nasabah untuk menggunakan layanan internet banking berdasarkan metode klasifikasi *Exhaustive CHAID* diantaranya yaitu: saya terbiasa menggunakan *internet banking* (IB); internet banking (IB) dapat meningkatkan produktivitas kerja saya; menggunakan *internet banking* (IB) tidak beresiko kebocoran data; saya merasa senang dapat menggunakan *internet banking* (IB); *internet banking* (IB) membantu dalam menyelesaikan pekerjaan saya.

Penelitian lain dengan menggunakan metode *Exhaustive CHAID* yang disusun oleh Ahmadi dan Herlina (2017), menyimpulkan bahwa terdapat pengaruh

## IR – PERPUSTAKAAN UNIVERSITAS AIRLANGGA

antara segmentasi demografis ( $X_1$ ) terhadap keputusan pembelian ( $Y$ ), terdapat pengaruh antara segmentasi psikografis ( $X_2$ ) terhadap keputusan pembelian ( $Y$ ), dan produk yang diteliti banyak digunakan oleh konsumen dengan status pekerjaan sebagai pelajar, mahasiswa, pegawai swasta yang memiliki gaya hidup lebih suka di ruangan dan suka berolahraga dengan usia 24-30 tahun.

Penelitian sebelumnya dengan menggunakan metode QUEST yang disusun oleh Abdurrahman, dkk (2014), menyimpulkan bahwa dari pohon yang terbentuk, variabel yang penting dalam mengklasifikasikan seorang terkena penyakit liver atau tidak, yaitu variabel  $X_4$  (DB),  $X_5$  (FA),  $X_7$  (SGOT), dan  $X_2$  (umur) yang selanjutnya variabel (hasil tes fungsi hati) tersebut dapat disebut sebagai ciri-ciri pasien liver. Tingkat akurasi pohon klasifikasi dalam mengklasifikasikan data sebesar 73,4%.

Penelitian lain dengan metode QUEST yang disusun oleh Fikri, dkk menyimpulkan bahwa faktor yang paling berpengaruh terhadap tingkatan kemiskinan adalah kepemilikan *personal computer* dan banyaknya anggota rumah tangga.

*Exhaustive CHAID* dan QUEST mempunyai kesamaan karakteristik, yaitu: sama-sama diterapkan pada data dengan variabel berskala kategorik, penggunaan uji *Chi Square* dalam pemilihan variabel penyekat dan hasilnya sama-sama dibandingkan dengan koreksi *Bonferroni*.

Kedua metode tersebut akan diterapkan pada hasil pengobatan data kasus TBC paru yang diperoleh dari Puskesmas Jati Kabupaten Kudus. Pada penelitian ini data diklasifikasikan berdasarkan tiga variabel independen, yaitu: umur, jenis

## IR – PERPUSTAKAAN UNIVERSITAS AIRLANGGA

kelamin, dan tipe pasien. Di Indonesia 75% penderita TBC paru adalah kelompok umur 15-50 tahun. Kelompok umur ini masuk dalam penduduk umur produktif, dimana seseorang yang tgermasuk dalam umur produktif banyak melakukan kegiatan seperti bekerja, belajar, maupun kegiatan lainnya. Seseorang yang melakukan banyak akativitas akan sering berinteraksi dengan orang lain dan lingkungan. Interaksi tersebut dapat memungkinkan terjadinya penularan TBC paru (Kemenkes RI, 2014).

Tipe pasien TBC paru ditentukan berdasarkan riwayat pengobatan sebelumnya, ada beberapa tipe pasien TBC paru, antara lain: baru, kambuh, putus berobat, gagal, pindah, dan lain-lain (Depkes RI, 2008).

Hasil pengobatan dilaksanakan dengan pemeriksaan dahak secara mikroskopis karena pemeriksaan dengan cara ini lebih baik dibandingkan dengan pemeriksaan *rontgen*, pemeriksaan *specimen* dahak dilakukan dua kali. Kategori hasil pengobatan, antara lain: sembuh, pengobatan lengkap, meninggal, pindah, putus berobat, dan gagal (Depkes RI, 2008).

Kasus TBC paru dipengaruhi oleh beberapa faktor, sehingga perlu adanya penelitian untuk menemukan faktor yang sangat berpengaruh terhadap kasus TBC paru. Pada sisi lain, belum banyak penelitian yang membandingkan langsung antara *Exhaustive CHAID* dan *QUEST*. Oleh karena itu, dengan melakukan klasifikasi menggunakan metode *Exhaustive CHAID* dan *QUEST*, diharapkan dapat menentukan metode klasifikasi yang tepat dalam mengklasifikasikan hasil pengobatan data kasus TBC paru di wilayah Puskesmas Jati.

### 1.3 Rumusan Masalah

Berdasarkan latar belakang dan kajian masalah yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah bagaimana perbandingan metode *Exhaustive* CHAID dan QUEST dalam mengklasifikasikan hasil pengobatan data kasus TBC paru di wilayah Puskesmas Jati Kabupaten Kudus tahun 2013 - 2018?

### 1.4 Tujuan Penelitian

#### 1.4.1 Tujuan Umum

Membandingkan metode *Exhaustive* CHAID dan QUEST dalam mengklasifikasikan hasil pengobatan data kasus TBC paru di Wilayah Puskesmas Jati Kabupaten Kudus.

#### 1.4.2 Tujuan Khusus

1. Mengidentifikasi pasien TBC paru di wilayah Puskesmas Jati berdasarkan umur, jenis kelamin, tipe penderita, dan hasil pengobatan dengan menggunakan tabel distribusi frekuensi.
2. Mengklasifikasikan hasil pengobatan data kasus TBC paru di wilayah Puskesmas Jati menurut kelompok hasil pengobatan, jenis kelamin, umur, dan tipe penderita dengan menggunakan *Exhaustive* CHAID dan QUEST.
3. Membandingkan ketepatan klasifikasi antara metode *Exhaustive* CHAID dan QUEST dengan melihat nilai ketepatan klasifikasi yang paling tinggi.

## 1.5 Manfaat Penelitian

### 1.5.1 Teoritis

Hasil penelitian ini diharapkan dapat memberikan masukan untuk penerapan ilmu biostatistika yang berhubungan dengan pengklasifikasian dengan menggunakan metode *Exhaustive* CHAID dan QUEST, serta dapat mengetahui perbandingan besarnya kesalahan klasifikasi dari kedua metode tersebut.

### 1.5.2 Praktis

#### 1.5.2.1 Tempat Penelitian

Hasil penelitian ini diharapkan dapat menjadi masukan dan evaluasi untuk Puskesmas Jati Kudus dalam mengklasifikasikan hasil pengobatan data kasus TBC paru misalnya menggunakan metode *Exhaustive* CHAID dan QUEST, sehingga capaian kinerja menjadi lebih baik pada tahun mendatang.

#### 1.5.2.2 Akademisi

Hasil penelitian ini diharapkan menjadi tambahan informasi yang dapat digunakan pada penelitian selanjutnya, khususnya penelitian yang berkaitan dengan klasifikasi, seperti *Exhaustive* CHAID dan QUEST.

#### 1.5.2.3 Mahasiswa/Peneliti

Hasil penelitian ini diharapkan dapat menjadi sarana belajar mahasiswa untuk meningkatkan *hard skill* dan *soft skill* mahasiswa dalam proses penelitian baik itu penelitian yang membuktikan teori

## IR – PERPUSTAKAAN UNIVERSITAS AIRLANGGA

tertentu maupun mengembangkan serta melatih kepekaan terhadap berbagai permasalahan di masyarakat yang berhubungan dengan penerapan metode statistik seperti metode *Exhaustive* CHAID dan QUEST yang digunakan dalam proses pengklasifikasian di bidang kesehatan baik secara langsung maupun tidak langsung yang dapat mempengaruhi permasalahan kesehatan masyarakat.