

BAB I

PENDAHULUAN

1.1. Latar Belakang

Hiperglikemia terjadi ketika kadar gula dalam darah (KGD) dalam diri penderita diabetes meningkat melebihi kadar normal (126 mg/dL (7.0 mmol/L)). Hiperglikemia sebagai manifestasi stres paling sering ditemukan tidak lama setelah masuk ICU, dan menghilang saat gangguan perusakan serat-serat otot yang mendasarinya berkurang. Risiko terjadinya hiperglikemia bervariasi, diantaranya diagnosis saat masuk ICU dengan kondisi klinis spesifik yang berbeda, serta kerusakan organ karena hiperglikemia (Falciglia, et al., 2009). Kondisi hiperglikemia pada pasien ICU menunjukkan adanya suatu keterkaitan terhadap kecacatan dan kematian pasien (Umpierrez, et al., 2002), sehingga pihak rumah sakit perlu membuat suatu kebijakan terkait pengobatan hiperglikemia.

Kebijakan tersebut dapat dirumuskan dengan melakukan analisa terhadap data pasien saat masuk ICU yang memiliki kondisi hiperglikemia. Pada dasarnya data merupakan suatu entitas yang tidak memiliki arti, meskipun kemungkinan memiliki nilai yang terkandung didalamnya. Kita hidup di era informasi dimana pengumpulan data mudah dilakukan dan penyimpanan data relatif tidak begitu mahal. Sayangnya, semakin banyak data yang terkumpul tidak diimbangi dengan kemampuan memahami dan memanfaatkan data tersebut. Analisa data secara tradisional memakan waktu lama dimana setiap hipotesis harus dirumuskan dan diuji secara individu.

Penggalian informasi dari data dilakukan dengan melakukan penggalian data (*data mining*). *Data mining* merupakan kegiatan ekstraksi untuk mendapatkan informasi penting yang sifatnya tersembunyi dan sebelumnya tidak diketahui dari sekumpulan data. Pada umumnya dalam sekelompok data rekam medis (dataset medis) didalamnya terdapat data yang berharga namun terlalu beragam sehingga menimbulkan kesulitan dalam pemahaman, hal ini disebabkan karena banyaknya data yang tidak lengkap (*missing values*), data tidak konsisten, dan banyaknya fitur-

fitur yang digunakan sehingga menimbulkan masalah tingginya dimensionalitas data (Krzysztof & Moore, 2002). *Data mining* dapat digunakan terhadap data medis. Terdapat paling tidak tiga tujuan utama dari diterapkannya teknik *data mining* terhadap data medis, yaitu: untuk lebih memahami tentang data tersebut, membantu para tenaga profesional kesehatan, dan mengembangkan metodologi analisis data yang sesuai dengan data klinis (Janvidrasana, et al., 2009).

Dalam *Data Mining* terdapat berbagai macam metode, dimana klasifikasi merupakan salah satu metode pada *Data Mining*. Klasifikasi *Data Mining* adalah suatu metode pembelajaran untuk memprediksi nilai dari sekelompok fitur dalam menggambarkan dan membedakan kelas. *Naïve Bayes* merupakan algoritma klasifikasi sederhana dimana setiap fitur dianggap bersifat berdiri sendiri (*independent*), *Naïve Bayes* menganggap bahwa ada atau tidaknya suatu fitur tidak berkaitan dengan keberadaan fitur lain. *Naïve Bayes* dikatakan sederhana karena *Naïve Bayes* menggunakan probabilitas fitur-fitur terhadap kelas (Murphy, 2006).

Dalam sekelompok data (dapat disebut dataset), semakin banyak data kejadian dan fitur yang terkumpul dapat menyebabkan masalah berupa tingginya dimensionalitas data. Dimensionalitas data merupakan salah satu faktor yang mempengaruhi kinerja tugas data mining, hal ini dapat terlihat dari banyaknya kejadian dan fitur dalam dataset tersebut. Semakin tinggi dimensionalitas data dapat menyebabkan data *overload*, data berlebihan atau data menyimpang, dan hal itu dapat mempengaruhi proses pembelajaran dari *Data Mining* (Danubianu, et al., 2012).

Oleh karena itu pada tugas akhir ini, diterapkan seleksi fitur yang bertujuan mengidentifikasi dan menghilangkan fitur yang dianggap menyimpang dan berlebihan. Dan setelah proses seleksi fitur tersebut akan memberikan beberapa keuntungan seperti meningkatnya hasil akurasi, data menjadi lebih padat, dan mengurangi proses eksekusi algoritma. Seleksi fitur dapat dilakukan salah satunya dengan metode pembungkus (*Wrapper*), dimana pada metode ini dimasukkan sebuah algoritma induksi untuk kemudian akan dihitung akurasi dari algoritma induksi terhadap subset fitur tersebut. Selanjutnya untuk pemilihan subset fitur

dapat dilakukan dengan metode pencarian *Greedy Stepwise* dengan arah pencarian maju (*forward*) yang bermula dengan tidak ada fitur sama sekali kemudian fitur akan ditambahkan satu-persatu, atau mundur (*backward*) yang bermula dengan semua fitur kemudian akan dikurangi satu-persatu. Dalam setiap iterasi pengurangan/penambahan fitur tersebut akan dilakukan evaluasi berdasarkan hasil akurasi dari algoritma induksi dalam metode *Wrapper*. Fitur yang memberikan peningkatan hasil akurasi kemudian akan dipilih dan dimasukkan ke dalam subset fitur akhir untuk selanjutnya dilakukan proses pembuatan model klasifikasi dalam tahap pelatihan.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah bagaimana penerapan seleksi fitur *Wrapper Greedy Stepwise* dan metode klasifikasi *Naïve Bayes* untuk meningkatkan hasil klasifikasi penerimaan pasien rawat inap di rumah sakit.

1.3. Batasan Masalah

Berdasarkan latar belakang dan rumusan masalah tersebut, diperlukan suatu batasan masalah agar lingkup dari penelitian ini tidak meluas ke permasalahan lain. Terdapat beberapa batasan masalah dalam tugas akhir ini:

1. Data set yang digunakan adalah data pasien diabetes pada 130 rumah sakit Amerika tahun 1999-2008 yang berasal dari repository *University California Irvine*.
2. Dari 101.766 data pasien, pada penelitian ini hanya akan digunakan sekitar 70.434 data saja, yang berisi pasien pada pertemuan pertama dan pasien dengan status masih hidup.
3. Proses klasifikasi, seleksi fitur, dan perhitungan akurasi menggunakan kelas *Naïve Bayes* dan *AttributeSelectedClassifier* dari Weka.
4. Fitur "*readmission_status*" akan dijadikan fitur kelas, fitur ini berisikan hasil pengelompokan nilai-nilai dari fitur "*readmitted*".

5. Proses pembelajaran mesin dan penggalian data menggunakan WEKA *java library*.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk meningkatkan hasil akurasi klasifikasi rasio pasien rawat inap dengan cara mengabaikan fitur-fitur yang tidak relevan dengan proses klasifikasi dalam dataset penerimaan pasien rawat inap.

1.5. Manfaat Penelitian

Dengan mengetahui fitur-fitur pilihan yang berkontribusi terhadap proses klasifikasi yang ditunjukkan dengan meningkatnya nilai akurasi dari klasifikasi, maka pihak rumah sakit dapat mengetahui perawatan dan pengobatan yang efektif sehingga diharapkan dapat mengurangi rasio pasien rawat inap.

1.6. Sistematika Penulisan

Sistematika penulisan penelitian ini terdiri dari beberapa bab yaitu BAB I Pendahuluan, berisi mengenai latar belakang masalah, rumusan masalah, batasan masalah, manfaat penelitian serta sistematika penelitian tugas akhir. BAB II Tinjauan Pustaka, berisi mengenai teori-teori yang dijadikan dasar utama penelitian ini seperti Hiperglikemia, *Data Mining*, Algoritma Klasifikasi *Naïve Bayes*, Metode Seleksi Fitur *Wrapper Greedy Stepwise*, *Confusion Matrix* serta mengenai Penelitian Terkait dan Rencana Penelitian. BAB III Metodologi Penelitian, berisi mengenai langkah-langkah yang dilakukan penulis dalam melakukan penelitian ini. BAB IV Hasil dan Pembahasan, memaparkan hasil dari penelitian. BAB V Kesimpulan dan Saran, berisi kesimpulan dan saran. Kesimpulan berisi jawaban dari rumusan masalah, dan saran berisi tindak lanjut dari hasil penelitian.