

BAB II

LANDASAN TEORI

2.1 Anamnesis Keluhan Pasien

Anamnesis atau keluhan pasien merupakan wawancara medis yang dilakukan oleh dokter terhadap pasiennya untuk memperoleh informasi mengenai kondisi yang sedang dialami oleh pasien agar dokter dapat menyimpulkan diagnosis penyakit dari pasien tersebut (Markum, 2000). Menurut Aspects (1996), penting bagi pasien untuk menggambarkan secara jelas mengenai gejala penyakit yang sedang dialaminya dengan bahasanya sendiri dan keluhan pasien harus didokumentasi dengan lengkap dari awal pemeriksaan. Tujuan dari anamnesis adalah untuk mendapatkan informasi yang menyeluruh mengenai kesehatan pasien dan menjaga hubungan komunikasi yang baik antara dokter dan pasien secara profesional agar dokter dapat mengekspresikan empati terhadap pasiennya dan sebaliknya (Markum, 2000). Berkomunikasi secara empatik termasuk salah satu aspek penting dalam interaksi antara dokter dengan pasiennya, karena dapat memberikan kepuasan tersendiri bagi pasien (Ohm *et al.*, 2013). Namun, pada kenyataannya wawancara medis juga merupakan salah satu keterampilan yang paling sulit untuk dikuasai oleh dokter (Lichstein, 1988).

Anamnesis masih dianggap sangat penting untuk pengambilan keputusan klinis, menurut Bernard Lown mengatakan bahwa data yang diperoleh dari anamnesis dapat memberikan informasi sekitar 75% untuk membuat diagnosis penyakit pasien sebelum dokter melakukan pemeriksaan fisik (Ohm *et al.*, 2013). Pengambilan keputusan pada diagnosis medis ditentukan dari tiga hal yaitu, anamnesis, pemeriksaan fisik, dan hasil investigasi dari laboratorium (Hampton *et al.*, 1975).

Anamnesis dibagi menjadi dua jenis, yaitu autoanamnesis dan alloanamnesis. Autoanamnesis adalah wawancara medis yang dilakukan secara langsung antara dokter dan pasien itu sendiri, sedangkan alloanamnesis dilakukan oleh dokter dengan keluarga pasien yang membawa pasien tersebut ke dokter (Markum, 2000). Alloanamnesis sangat dibutuhkan jika berhubungan dengan anak kecil atau bayi, orang tua lansia, dan pasien sakit jiwa.

Menurut penelitian Markum (2000), data anamnesis dikelompokkan menjadi enam bagian data penting, yaitu identitas pasien, riwayat penyakit sekarang (didahului dengan keluhan utama), riwayat penyakit dahulu, anamnesis sistem, riwayat kesehatan keluarga, dan riwayat

pribadi terkait sosial, ekonomi, dan budaya. Data identitas pasien berisi nama, umur, jenis kelamin, pekerjaan, alamat, status perkawinan, agama, dan suku bangsa. Riwayat Penyakit Sekarang (RPS) dimulai dari akhir masa sehat secara kronologis waktu, lalu dicatat setiap keluhan pasien dengan mendeskripsikan perjalanan penyakitnya, dan data yang ditulis sebaiknya menggunakan kata-kata atau bahasa dari pasien itu sendiri (Markum, 2000).

Pada tahun 1982, Engel menjelaskan bahwa karakteristik dari suatu gejala yang dialami pasien dapat dilihat dari tujuh dimensi atau lebih dikenal dengan istilah “*The Sacred Seven*” berdasarkan aspek fisik dan emosi (Lichstein, 1988), yaitu sebagai berikut:

a. *Chronology*

Deskripsi dari kronologis ini menjelaskan tentang perjalanan penyakit, dengan itu dokter harus mendapatkan laporan kronologis dengan menanyakan kapan sakit pertama kali dirasakan dan dilanjutkan dengan pertanyaan lainnya secara spesifik. Kronologi juga mencakup durasi simtomatik, periodisitas dan apakah gejalanya menjadi lebih baik atau lebih buruk dari waktu ke waktu.

b. *Bodily location*

Lokasi sakitnya harus didefinisikan seakurat mungkin dengan cara menunjukkan lokasi nyeri yang dirasakan oleh pasien dengan menggunakan gerakan tangannya. Perlu diingat bagi dokter bahwa pasien mungkin memiliki lebih dari satu penyakit dan rasa sakit itu dapat mengindikasikan berbagai proses penyakit. Untuk itu, dokter dapat meminta pasien agar dapat membedakan dan mengkarakterisasikan masing-masing penyakitnya.

c. *Quality*

Sebagian besar pasien menggambarkan kualitas dari rasa sakitnya menggunakan analogi. Beberapa pasien juga menggunakan bahasa yang deskriptif atau emosional seperti, “Rasanya seperti seseorang menikam saya dengan pisau”, dan pasien sulit untuk menemukan bahasa yang sedang dideskripsikannya.

d. *Quantity*

Intensitas nyeri yang dirasakan pasien dapat diperkirakan dengan menggunakan skala 1 sampai 10 atau dapat dibandingkan dengan nyeri yang lainnya. Selain itu, *volume* juga termasuk dalam salah satu contoh kuantitas. Misalnya, jumlah dahak yang dikeluarkan dalam sehari.

e. *Setting*

Pengaturan ini menjelaskan tentang gejala, di mana, apa, dan dengan siapa pasien pada saat merasakan sakit itu. Hal itu merupakan pertanyaan yang bagus untuk digunakan di awal wawancara.

f. *Aggravating or alleviating factors*

Dokter mengumpulkan data mengenai hal-hal apa saja yang telah dilakukan oleh pasien. Misalnya, pasien telah minum obat sebelumnya. Hasil dari data tersebut, dokter dapat menganalisis apa yang telah membuat gejala lebih buruk atau menjadi lebih baik.

g. *Associated manifestations*

Gejala jarang terjadi dengan sendirinya, maka dari itu dokter harus mendengarkan gejala-gejala lainnya yang dapat memberikan informasi diagnostik tentang patologi dan organ yang terlibat.

Penelitian yang dilakukan oleh Ratnasari *et al.* (2014) pada sistem perumusan keluhan pasien, memetakan Riwayat Penyakit Sekarang (RPS) ke dalam sebelas *slot/field*. Sebelas *slot/field* tersebut yaitu keluhan utama, onset, keluhan lain, keterangan, frekuensi serangan, sifat serangan, durasi, lokasi, perjalanan penyakit, riwayat pengobatan sebelumnya, dan akibat gangguan yang timbul. Dari klasifikasi tersebut, tujuh di antaranya termasuk dalam *The Sacred Seven* anamnesis.

2.2 Ekstraksi Informasi

Ekstraksi informasi adalah sebuah proses untuk mencari informasi yang terstruktur dari teks-teks yang tidak terstruktur atau semi terstruktur dan struktur informasi yang akan diekstraksi tergantung pada kebutuhan aplikasi tersebut (Jiang, 2012). Sedangkan menurut Kaiser dan Miksch (2005), ekstraksi informasi adalah teknik yang digunakan untuk mendeteksi informasi dalam dokumen yang lebih besar secara relevan dan menampilkannya dalam format terstruktur. Tujuan dari ekstraksi informasi adalah mengubah teks menjadi format yang terstruktur dengan menampilkan informasi inti yang ada dalam dokumen menjadi struktur tabel (Eikvil, 1999).

Jiang (2012) mengatakan bahwa, “Ekstraksi informasi mengasumsikan bahwa struktur yang akan diekstraksi harus didefinisikan dengan baik, misalnya tipe nama entitas, relasi, dan *slot template*. Sebagai contoh, dari sebuah artikel berita gempa bumi, mungkin dari sebagian orang menginginkan secara otomatis dapat menemukan tanggal, waktu, pusat gempa, besar

kekuatan gempa, dan jumlah korban sebagai bagian terpenting dari informasi dalam artikel tersebut”. Menurut Kaiser dan Miksch (2005), mengatakan bahwa ekstraksi informasi digunakan untuk menganalisis teks dan menemukan informasi tertentu yang ada dalam teks, tetapi bukan untuk memahami teks. Ekstraksi informasi telah dipelajari secara luas untuk berbagai penelitian termasuk pemrosesan bahasa alami, pencarian informasi, dan *web mining* (Jiang, 2012). Pada awalnya, ekstraksi informasi digunakan untuk memperoleh informasi yang spesifik dari dokumen bahasa alami untuk pemrosesan bahasa alami dan telah dikembangkan baik untuk teks yang terstruktur maupun teks bebas (Eikvil, 1999). Yang dimaksud dari bahasa alami itu sendiri adalah bahasa yang biasa diucapkan atau digunakan oleh manusia untuk berkomunikasi atau berinteraksi dan mempunyai gaya bahasa tersendiri (Sari & Wicaksana, 2011).

Menurut penelitian dari Abdelmagid, Ahmed, dan Himmat (2015), pendekatan pada ekstraksi informasi dikategorikan dalam empat jenis, yaitu *Supervised Learning Systems*, *Semi-supervised Learning Systems*, *Unsupervised Learning Systems*, dan *Hybrid NER (Named Entity Recognition)*. Pendekatan pada *hybrid NER* sendiri merupakan kombinasi antara metode berbasis kamus dan statistik (*Conditional Random*). Penelitian oleh Kushmerick adalah mengklasifikasikan alat ekstraksi informasi ke dalam dua kategori yang berbeda, yaitu *finite-state* dan pembelajaran relasional. Aturan ekstraksi pada *finite-state* secara formal sama dengan tata bahasa atau automata, misalnya WIEN, *SoftMealy*, dan STALKER, sedangkan aturan ekstraksi pada alat pembelajaran relasional pada dasarnya berbentuk program logika seperti Prolog, misalnya SRV, *Crystal*, *WebFoot*, *Rapier* dan *Pinocchio* (Kaiser & Miksch, 2005).

Ekstraksi informasi merupakan salah satu cabang dari Pemrosesan Bahasa Alami. Metode yang dapat digunakan dalam ekstraksi informasi, yaitu *rule-based methods*, *statistical-based methods*, dan *knowledge-based methods* (Ismaya, 2014). *Rule-based* merupakan metode yang menggunakan aturan yang telah dibuat dan mengidentifikasi nama entitas dengan mencocokkan kata dengan aturan yang ditentukan, *statistical-based* merupakan metode dengan korpus yang menentukan peluang menggunakan nilai *threshold* dengan kata sebagai nama entitas, sedangkan *knowledge-based* merupakan metode campuran antara *rule-based* dan *statistical-method* (Pramiyati et al., 2015).

2.3 TF-IDF

Term Frequency Invers Document Frequency (TF-IDF) adalah salah satu metode dari *features extraction* yang digunakan untuk menentukan seberapa jauh hubungan suatu kata (*term*) terhadap dokumen dengan memberikan bobot pada setiap kata (Herwijayanti, Ratnawati, & Muflikhah, 2018). *Term frequency* merupakan metode untuk memberikan bobot pada kata, sedangkan *document frequency* digunakan untuk memperhatikan kata (*term*) yang muncul pada dokumen. Dalam perhitungan tf-idf, dihitung terlebih dahulu nilai tf perkata dan diberi bobot 1 pada setiap kata. Nilai tf-idf dihitung berdasarkan perkalian antara tf dan idf.

2.4 Metode Statistik

Ekstraksi informasi merupakan teknik yang dapat mengekstraksi informasi dari data yang tidak terstruktur menjadi terstruktur. Selain metode berbasis aturan, salah satu metode yang juga dapat digunakan dalam ekstraksi informasi adalah metode pembelajaran statistik seperti *Hidden Markov Model (HMM)*, *Conditional Random Fields (CRF)*, dan *Support Vector Machine (SVM)* (Chen, Feng, Christopher, & Wang, 2012). Metode statistik pertama kali diperkenalkan pada tahun 1960-an sampai tahun 1990-an, metode ini berisi analisis teoritis mengenai masalah estimasi fungsi dari kumpulan data yang diberikan.

Menurut penelitian dari Wang dan Cohen (2014), “Pembelajaran pada metode statistik terdiri dari dua langkah, yaitu membangun basis pengetahuan pada teks-teks, kemudian dilanjutkan dengan melakukan tugas-tugas yang terdapat dalam pembelajaran statistik menggunakan logika probabilistik atau peluang. Akan tetapi, yang jadi masalahnya adalah apabila ada kesalahan yang terjadi pada ekstraksi informasi dari teks dapat mempengaruhi kualitas pada basis pengetahuan dan probabilitas tersebut.”

Jika dibandingkan dengan metode berbasis aturan, metode statistik dapat menangkap pola yang lebih kompleks. Semakin berkembangnya teknologi program aplikasi ekstraksi informasi, membuat metode statistik banyak diaplikasikan di *Alibaba*, *Microsoft's Academic Search*, *Cite Seer*, *Kylin*, dan *YAGO* (Chen *et al.*, 2012).

Pada dasarnya, metode statistik memanfaatkan metode pembelajaran mesin. Metode statistik dikategorikan menjadi tiga kelompok yang berbeda, yaitu (Allahyari *et al.*, 2017):

- a. Metode berbasis klasifikasi, adalah sebuah metode yang mengubah tugas *Named Entity Recognition (NER)* menjadi klasifikasi yang dapat digunakan untuk kata atau frasa.

Metode yang biasa digunakan untuk biomedis adalah *Naive Bayes* dan *Support Vector Machine* (SVM).

- b. Metode berbasis urutan, adalah sebuah metode yang memprediksi *tag* yang paling mungkin untuk urutan kata. Metode yang biasa digunakan adalah *Hidden Markov Model* (HMM) dan *Conditional Random Field* (CRF).
- c. Metode *hybrid*, metode ini menggunakan dua pendekatan yaitu berbasis aturan dan statistik

2.5 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah salah satu jenis pendekatan pembelajaran mesin dari teori pembelajaran statistik (Cai *et al.*, 2002). *Support Vector Machine* (SVM) menjadi salah satu metode klasifikasi yang sangat populer karena telah banyak digunakan dalam berbagai aplikasi seperti klasifikasi teks, pengenalan wajah, analisis wajah, dan lain sebagainya (Fradkin & Muchnik, 2014).

Pada tahun 1990-an, ditemukan algoritma pembelajaran baru yaitu *Support Vector Machine* (SVM) yang telah dikembangkan dari teori sebelumnya (Vapnik, 1999). *Support Vector Machines* (SVM) pertama kali ditemukan oleh V. Vapnik, metode ini termasuk pembelajaran klasifikasi berdasarkan prinsip minimalisasi risiko struktural dari pembelajaran komputasi yang digunakan untuk menemukan hipotesis agar dapat mengurangi kesalahan yang tidak terlihat (Joachims, 1998).

Support Vector Machines (SVM) merupakan algoritma klasifikasi linear yang menentukan keputusan klasifikasi berdasarkan nilai kombinasi linear dan dokumen teks, klasifikasi linear banyak digunakan pada masalah klasifikasi teks karena dapat membuat keputusan berdasarkan nilai kombinasi linear (Allahyari *et al.*, 2017). Selain untuk klasifikasi teks, *Support Vector Machines* (SVM) juga digunakan dalam pengenalan gambar dan desain obat (Cai *et al.*, 2002)

Menurut penelitian dari (Liliana *at al.*, 2011), ”*Support Vector Machine* (SVM) adalah sistem pembelajaran mesin yang menggunakan hipotesis fungsi linear dalam fitur dimensi yang dilatih dengan pembelajaran algoritma pada teori optimalisasi berdasarkan teori pembelajaran statistik. Tujuan dari metode *Support Vector Machine* (SVM) adalah menemukan fungsi pemisahan yang optimal atau fungsi linear yang akan digunakan untuk memisahkan kelas-kelas atau klasifikasi”.

Support Vector Machine (SVM) mempunyai nilai akurasi klasifikasi yang tinggi dalam pengenalan entitas bernama apabila dibandingkan dengan ketiga metode lainnya, seperti kernel kuadrat, entropi maksimal, dan pembelajaran berbasis aturan (Li *et al.*, 2005). Jika SVM dibandingkan dengan teks konvensional, maka klasifikasi SVM terbukti lebih menunjukkan keuntungan dengan kinerja yang sangat baik (Joachims, 1998).

Support Vector Machine (SVM) memiliki beberapa fitur yang bermanfaat, yaitu (Vapnik, 1999):

- a. Dapat memberikan solusi yang unik dalam masalah optimasi untuk membangun *Support Vector Machine* (SVM).
- b. Proses pembelajaran untuk membangun *Support Vector Machine* (SVM) lebih cepat.
- c. Set vektor dukungan dapat diperoleh dengan membuat aturan keputusan.
- d. Implementasi pada fungsi keputusan yang baru dapat dilakukan hanya dengan mengubah satu fungsi.

Pada penelitian yang dilakukan (Joachims, 1998), memberikan penjelasan teoritis bahwa *Support Vector Machine* (SVM) yang baik untuk klasifikasi teks dapat dilihat berdasarkan masalah berikut:

- a. Ruang *input* yang berdimensi tinggi
Support Vector Machine (SVM) mampu memproses *input* yang berdimensi tinggi secara efektif karena klasifikasi teks mempunyai fitur yang sangat banyak sekitar lebih dari 10.000, metode *Support Vector Machine* (SVM) tidak tergantung dengan berapa banyak jumlah fiturnya.
- b. Fitur yang tidak relevan
 Masalah ruang *input* yang berdimensi tinggi dapat diatasi dengan mengurangi fitur-fitur yang tidak relevan dan pemilihan fitur yang salah dapat menghilangkan informasi karena klasifikasi yang baik harus menggabungkan banyak fitur.
- c. Masalah pada klasifikasi teks dapat dipisahkan secara linear
 Salah satu tugas dari *Support Vector Machine* (SVM) adalah untuk menemukan pemisah linear, karena semua kategori dari data OHSUMED dapat dipisahkan secara linear. OHSUMED merupakan koleksi jurnal medis yang dapat dilihat di <ftp://medir.ohsu.edu> pada direktori </pub/ohsumed>. Jurnal tersebut dikelola oleh *National Library of Medicine*.