

BAB 1

PENDAHULUAN

1.1 *Latar Belakang*

Di era modernisasi seperti sekarang, teknologi merupakan salah satu dari banyaknya hal yang dibutuhkan seseorang untuk menunjang aktivitas harian mereka baik dalam hal pekerjaan, pendidikan hingga penelitian. Salah satu jenis teknologi yang paling banyak digunakan dalam kehidupan sehari-hari adalah teknologi dalam bidang informasi dan komunikasi. Pesatnya perkembangan teknologi informasi yang ada membuat banyak instansi-instansi baik dari lembaga pemerintahan maupun swasta memiliki sistem informasi mereka sendiri, sehingga memungkinkan masyarakat luas untuk mendapatkan informasi dengan mudah dan cepat.

Secara umum, sistem informasi merupakan suatu sistem yang di dalamnya memuat berbagai macam informasi yang terkait dengan operasional suatu organisasi meliputi jaringan komunikasi, transaksi rutin, manajemen dan informasi yang dibutuhkan oleh pihak internal maupun eksternal (Sirait, 2017). Situs atau laman *web* merupakan salah satu bentuk atau bagian dari sistem informasi.

Pada hakikatnya, laman *web* merupakan kumpulan informasi yang diberikan suatu individu, kelompok, atau organisasi. Sebuah laman *web* biasanya ditempatkan setidaknya pada suatu server *web* yang dapat diakses melalui jaringan seperti internet atau jaringan area lokal (LAN) melalui alamat internet yang dikenal sebagai *Uniform Resource Locator* (URL) (Silviana and Thalib, 2019).

Saat ini banyak laman *web* yang bermunculan dengan menyediakan berita seputar kesehatan sehingga dapat dimanfaatkan oleh masyarakat luas untuk pertolongan pertama dalam menjaga dan mencegah tumbuhnya penyakit pada tubuh mereka. Pada dasarnya, sebuah laman *web* kesehatan biasanya dapat menampilkan satu hingga jutaan informasi atau berita kesehatan dalam bentuk konten *web*. Berita tersebut umumnya diurutkan berdasarkan kategori yang beragam, seperti ibu hamil, kanker, dan alergi. Koleksi berita dalam jumlah besar sering mengandung banyak data ilmiah penting yang membutuhkan metode otomatis untuk mengatur dokumen-dokumen tersebut (Puri and Kumar, 2017). Beberapa laporan berita mungkin masuk dalam topik tertentu, sehingga

dalam dunia jurnalisme ada banyak editor yang beralih untuk mengatur berita tersebut ke topik khusus.

Sebuah kategori atau topik di dalamnya memiliki komponen-komponen tertentu dalam melakukan pengkategorian, sehingga hampir tidak mungkin untuk melacak seluruh komponen-komponen yang ada pada suatu berita dari topik tertentu. Kumpulan konten *web* dapat berisi banyak topik dan karenanya, mungkin termasuk dalam beberapa kategori yang berbeda. Pekerjaan dengan metode ini membutuhkan banyak tenaga dan sumber daya serta memiliki beban kerja yang berat, tetapi hasilnya tidak cukup praktis. Selain itu, Indonesia saat ini masih belum memiliki *web* atau situs kesehatan yang dapat merepresentasikan konten berita secara otomatis berdasarkan pada topik yang terkandung di dalam setiap konten berita. Oleh karena itu, metode mendeteksi dan melacak topik pada sebuah konten *web* yang relevan dan secara otomatis disarankan.

LDA merupakan model Bayesian hirarki tiga tingkat, di mana setiap *item* koleksi dokumen dimodelkan sebagai campuran terbatas atas serangkaian set topik (Blei, Ng and Jordan, 2003). LDA adalah teknologi *machine learning* yang termasuk dalam kelompok *unsupervised learning* yang dapat digunakan untuk mengidentifikasi informasi subjek tersembunyi dalam kumpulan dokumen besar atau korpus. LDA menggunakan metode *bag of words*. Metode ini memperlakukan setiap dokumen sebagai *frequency vector* kata sehingga mengubah dari informasi teks menjadi informasi digital. Metode *bag of words* tidak memperhitungkan urutan kata-kata yang dapat menyederhanakan kerumitan masalah dan bisa lebih cepat serta nyaman untuk melakukan proses *stemming* pada suatu dokumen teks (Wang, Hong and Shi, 2018).

Berdasarkan pada definisi dari LDA, model ini dapat diterapkan untuk mendeteksi topik yang terkandung pada konten *web* kesehatan dengan melakukan pengecekan serangkaian komponen topik pada topik yang telah terbentuk. Pada penelitian kali ini penulis menggunakan metode LDA untuk membentuk topik berdasarkan hasil yang diperoleh dari ekstraksi *bag of word* dalam konten *web* kesehatan. Dalam penerapannya, dilakukan penentuan topik yang terkandung pada suatu laman *web* dengan memanfaatkan nilai probabilitas yang diperoleh pada topik-topik yang dibangun oleh model LDA. Dalam prosesnya, konten *web* akan melalui

proses pengolahan kata *stopwords removal*, *stemming* dan *tokenizing* karena pada prosesnya setiap teks dan jumlah kemunculannya sangat berpengaruh.

Selain melakukan analisis dan representasi konten pada situs kesehatan Indonesia menggunakan metode LDA, skripsi ini juga akan meninjau penelitian tersebut menurut pandangan Agama Islam, dilihat dari boleh atau tidak bolehnya dalam Islam. Dalam sudut pandang Islam, sudah suatu kelaziman dalam berbagai bidang keahlian maupun produk tertentu harus memenuhi standarnya termasuk topik pada berita kesehatan, sehingga keabsahan, kualitas dan validasinya terjamin dan dapat dipertanggungjawabkan (Putra, 2012).

إِنَّ الدِّينَ يُسْرٌ، وَلَنْ يُثَادَّ الدِّينَ إِلَّا غَلْبُهُ

Artinya:

“Sesungguhnya agama ini sangat mudah. Dan tiada seseorang yang mencoba mempersulit diri dalam agama ini melainkan ia pasti kalah” (HR. Bukhari)

وَهُوَ الَّذِي يُنَزِّلُ الْغَيْثَ مِنْ بَعْدِ مَا قَنَطُوا وَيَنْشُرُ رَحْمَتَهُ وَهُوَ الْوَلِيُّ الْحَمِيدُ

٢٨

Artinya:

“Dan Dialah Yang menurunkan hujan sesudah mereka berputus asa dan menyebarkan rahmat-Nya. Dan Dialah Yang Maha Pelindung lagi Maha Terpuji.” (QS. Asy-Syuura(42): 28).

Berdasarkan ayat di atas dapat dijadikan dasar untuk menyimpulkan apakah Analisis dan Representasi Konten pada Situs Kesehatan Indonesia diperbolehkan dalam ajaran Agama Islam, baik dari kitab suci Al-Qur'an maupun Hadits. Hal ini dikarenakan Islam telah menganjurkan kepada umat-Nya untuk mencari, mempelajari, serta mengamalkan ilmu yang sudah mereka miliki sehingga ilmu tersebut dapat bermanfaat (Kahfi, 2006).

1.3 *Identifikasi Masalah*

Berdasarkan latar belakang di atas, maka dapat diidentifikasi beberapa masalah yang ada sebagai berikut:

1. Banyak bermunculan situs kesehatan berbahasa Indonesia di mana pengelompokan konten kesehatan yang ada masih dilakukan secara manual sehingga memakan waktu yang relatif lama, membutuhkan banyak tenaga dan sumber daya serta memiliki beban kerja yang berat.
2. Belum ada korpus kesehatan berbahasa Indonesia.
3. Setiap konten berita memiliki kata kunci yang saling berhubungan sehingga perlu untuk dilakukan ekstraksi dan dibentuk model topik yang menggambarkan distribusi kata dalam suatu kategori.

1.4 *Rumusan Masalah*

Beberapa permasalahan yang akan diteliti dapat dirumuskan sebagai berikut:

1. Bagaimana cara menghasilkan korpus berita kesehatan sebagai bahan referensi penelitian?
2. Bagaimana cara mengidentifikasi topik yang terkandung pada konten berita *web* kesehatan secara otomatis?
3. Bagaimana tinjauan menurut Agama Islam mengenai sistem Representasi Konten Pada Laman *Web* secara otomatis?

1.5 *Batasan Masalah*

Agar penelitian yang dilakukan lebih terarah, fokus serta tidak meluas dari pembahasan, dalam hal ini penulis memberi batasan pada ruang lingkup penelitian sebagai berikut:

1. Data yang diteliti berupa konten dari beberapa situs kesehatan di Indonesia, yaitu *hellosehat.com*, *doktersehat.com*, *halodoc.com* dan *klikdokter.com*.
2. Konten *web* yang diambil hanya berupa berita seputar kesehatan.
3. Konten *web* yang diteliti berbentuk teks dalam bahasa Indonesia.
4. Konten *web* yang diambil adalah berita yang diterbitkan dari tahun 2016 hingga 2019.
5. Konten berita yang diproses memiliki format *html*.
6. Karakteristik dalam menentukan nama topik berdasarkan pada asumsi pakar di bidang kesehatan.

1.6 Tujuan dan Manfaat Penelitian

1.6.1 Tujuan Penelitian

Adapun tujuan dari dilakukannya penelitian ini adalah:

1. Membuat korpus berita kesehatan berdasarkan data yang diambil pada situs kesehatan sehingga dapat digunakan untuk membangun model topik dengan *Latent Dirichlet Allocation* (LDA).
2. Membangun LDA terbaik sehingga dapat melakukan identifikasi topik secara otomatis.
3. Meninjau dari sudut pandangan Agama Islam tentang pengembangan prototipe Analisis dan Representasi Konten Pada Laman *Web* Kesehatan Dengan Menggunakan Metode *Latent Dirichlet Allocation* (LDA).

1.6.2 Manfaat Penelitian

Penelitian ini dilakukan dengan harapan dapat memberikan manfaat antara lain:

1. Memudahkan pihak editor jurnalis pada situs kesehatan untuk mengelompokkan berita berdasarkan topik tertentu.
2. Sebagai acuan untuk merepresentasikan dokumen yang akan digunakan pada pengukuran kredibilitas pada dokumen tersebut.
3. Korpus dan kamus dari penelitian yang dihasilkan dapat digunakan pada penelitian di masa yang akan datang.