

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Pada tahun 2021 Elsevier menerbitkan lebih dari 600.000 artikel setiap tahun di lebih dari 2.700 jurnal (RELX, 2022). Literatur utama telah berkembang secara eksponensial selama 20 tahun terakhir, seperti halnya publikasi tinjauan sistematis (Borah *et al.*, 2017). Jumlah penelitian yang diterbitkan dalam jumlah besar dan terus bertambah, serta tingkat publikasinya yang meningkat, tahap skrining artikel biasanya membutuhkan waktu yang lama dan upaya yang signifikan karena melibatkan sekelompok ahli untuk mengevaluasi ribuan artikel untuk menemukan penelitian yang relevan. Otomatisasi *Systematic Literature Review* dapat menjadi solusi potensial untuk menggabungkan sejumlah besar informasi tentang satu masalah lintas disiplin ilmu. Tinjauan sistematis adalah hasil dari proses yang memakan waktu di mana peninjau dengan hati-hati mengevaluasi referensi artikel untuk relevansi dengan tinjauan. Ribuan, bahkan puluhan ribu, abstrak dari artikel yang mempunyai potensi diperiksa secara manual sebagai bagian dari proses ini.

*Systematic Literature Review* atau tinjauan pustaka sistematis adalah istilah yang mengacu pada proses penelitian atau proyek penelitian dan pengembangan berbasis bukti yang digunakan untuk menjawab pertanyaan penelitian, mengumpulkan dan menganalisis studi terkait yang relevan pada masalah tertentu (Lusiana and Suryani, 2014).

Melakukan tinjauan pustaka atas perkembangan ilmu pengetahuan melalui kajian literatur adalah salah satu bentuk potensi manusia yang diberikan oleh Allah Swt. Manusia memiliki potensi otak yang sangat besar sebagai makhluk yang dapat berpikir, sebagaimana tertuang dalam Surah Az-Zariyat ayat 20-21. Allah Swt berfirman:

وَفِي الْأَرْضِ آيَاتٌ لِّلْمُؤْمِنِينَ ۗ وَفِي أَنفُسِكُمْ ۗ أَفَلَا تُبْصِرُونَ

Artinya: “Di bumi terdapat tanda-tanda (kebesaran Allah) bagi orang-orang yang yakin. (Begitu juga ada tanda-tanda kebesaran-Nya) pada dirimu sendiri. Maka, apakah kamu tidak memperhatikan?” (Q.S Az- Zariyat [51]: 20-21) (Kementrian Agama Republik Indonesia, 2019).

Berdasarkan tafsir Al-Muyassar, dapat disimpulkan bahwa di bumi ada pelajaran dan bukti yang jelas yang menunjukkan Kuasa Allah bagi orang-orang yang yakin dan dalam

penciptaan manusia juga terkandung bukti atas Kuasa Allah dan pelajaran yang mengisyaratkan kepada keesaan Allah, bahwa Allah adalah satu-satunya Tuhan yang berhak disembah (*Tafsir Al-Muyassar*, no date).

Potensi yang ada di sekitar manusia menjadikan mereka makhluk yang istimewa. Setiap orang memiliki sejumlah potensi yang jika dikembangkan lebih lanjut, akan memungkinkan mereka memiliki ciri khas dan bahkan menciptakan nilai tambah.

Perkiraan waktu kerja yang diperlukan untuk membuat satu SLR berkisar dari minimal 6 bulan ketika satu peninjau menghabiskan 10-20 jam per minggu hingga, rata-rata, 16 bulan ketika lima rekan penulis terlibat. (Michelson and Reuter, 2019). Untuk melakukan tinjauan pustaka sistematis dibutuhkan tim yang terdiri dari dua orang atau lebih (*Systematic Reviews & Other Review Types*, 2022). Otomatisasi SLR membawa manfaat karena secara drastis mengurangi jumlah waktu dan upaya yang diperlukan. (van Dinter, Tekinerdogan and Catal, 2021). Upaya untuk mengotomatisasi tahap penyaringan literatur telah dilakukan oleh para peneliti sebelumnya.

Kajiwara et.al merancang rencana perawatan secara otomatis dari temuan yang disertakan dalam sertifikat medis yang ditulis oleh dokter gigi menggunakan metode BoW (Bag-of-Words), OoK (One-of-K), dan *Universal Sentence Encoder* (USE). Hasil eksperimen menggunakan 990 sertifikat dengan *F-1 Score* 58.5% (Kajiwara et al., 2019).

Ebraheem et.al mengusulkan dua metode untuk mempelajari representasi catatan. Metode pertama menggunakan model *GloVe word embedding* untuk mengubah kata menjadi vektor *embedding*. Metode kedua menggunakan *Recurrent Neural Networks* (RNNs) *uni*- dan *bi-directional* dengan *hidden units Long Short Term Memory* (LSTM) untuk menyusun vektor kata yang diambil dari *GloVe* menjadi vektor tingkat atribut secara semantik (Ebraheem et al., 2018).

Model *pre-trained* BERT menyediakan representasi kalimat bergantung konteks yang kuat dan dapat digunakan untuk berbagai tugas target, dan cocok untuk tugas klasifikasi teks (Bao et al., 2019; Soares et al., 2020).

## 1.2 Perumusan Masalah

Diberikan input adalah teks artikel ilmiah yang diambil dari korpus jurnal ilmiah bidang kedokteran gigi menggunakan kata kunci spesifik. Input  $X = [x_1, x_2, x_3, \dots, x_n]$  terdiri

dari  $n = 5$  fitur yaitu judul, penulis, abstrak, dan teks lengkap, dan referensi. Diberikan output  $y$  adalah hasil prediksi kelas artikel. Output  $y = \{1|0\}$  bernilai 1 jika artikel memenuhi kriteria inklusi dan 0 jika tidak.

Rumusan masalah pada penelitian ini adalah:

1. Bagaimana membuat model klasifikasi  $f(X) = y$ ? Model ini menerima input features  $X$  dan memberikan hasil prediksi kelas  $y$ ?
2. Bagaimana tanggapan Islam terhadap implementasi machine learning untuk tinjauan pustaka sistematis?

### 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan dan mengoptimalkan algoritma *transfer learning* pada proses tinjauan pustaka sistematis dari penelitian yang sudah pernah dilakukan sebelumnya..
2. Memberi kajian dari sudut pandang Islam terhadap implementasi *machine learning* untuk tinjauan pustaka sistematis.

### 1.4 Research Gap

Kajiwara et.al merancang rencana perawatan secara otomatis dari temuan yang disertakan dalam sertifikat medis yang ditulis oleh dokter gigi menggunakan metode BoW (Bag-of-Words), OoK (One-of-K), dan *Universal Sentence Encoder* (USE). Hasil eksperimen menggunakan 990 sertifikat dengan F-1 Score 58.5% (Kajiwara et al., 2019).

Ebraheem et.al mengusulkan dua metode untuk mempelajari representasi catatan. Metode pertama menggunakan model *GloVe word embedding* untuk mengubah kata menjadi vektor *embedding*. Metode kedua menggunakan *Recurrent Neural Networks* (RNNs) *uni-* dan *bi-directional* dengan *hidden units Long Short Term Memory* (LSTM) untuk menyusun vektor kata yang diambil dari *GloVe* menjadi vektor tingkat atribut secara semantik (Ebraheem et al., 2018).

Howard et.al mengembangkan model klasifikasi log-linear reguler yang memanfaatkan dua jenis fitur, yaitu BoW (Bags of Words) yang dibobot oleh TF-IDF dan fitur berbasis topik yang diekstraksi oleh *Latent Dirichlet Allocation* (LDA). Eksperimen di dua puluh tinjauan sistematis dilakukan, menunjukkan presisi dan recall sebesar 95% (Howard et al., 2016).

Namun demikian, meskipun penerapan AI di berbagai bidang sudah banyak dilakukan, penerapannya masih belum banyak diimplementasikan di bidang kedokteran gigi, khususnya tinjauan pustaka sistematis. Berdasarkan fakta tersebut, penulis terinspirasi untuk melakukan penelitian ini.

### **1.5 Metode yang Diajukan**

Eksperimen dimulai dengan mengekstrak teks dari dokumen artikel ilmiah yang telah diklasifikasikan secara manual ke dalam *dataset*. Data selanjutnya melalui tahap *text processing*, yaitu pembersihan teks agar mudah diolah. Tahap ini termasuk mengubah teks menjadi huruf kecil, menghilangkan *stopwords*, menghilangkan tanda baca, dan lematisasi. Representasi data juga diubah menggunakan pendekatan sesuai kebutuhan, yaitu *pre-trained word embedding*. Data selanjutnya dilatih menggunakan algoritma *transfer learning*.

### **1.6 Hasil Evaluasi**

Dataset yang digunakan pada penelitian ini berisi 96 baris data hasil ekstraksi artikel jurnal ilmiah. Terdapat masing-masing 48 baris data pada kelas *included* dan *excluded*. Evaluasi algoritma dilakukan dengan teknik *K-Fold Cross Validation* dengan nilai  $k = 4$ . Hasil evaluasi menunjukkan bahwa optimalisasi dari model *transfer learning* menggunakan *Universal Sentence Encoder* berhasil mendapatkan akurasi tertinggi dengan nilai 90.62% pada skrining Judul-Abstrak dan 85.41% pada skrining Teks-lengkap.

### **1.7 Kontribusi**

Bentuk kontribusi yang dihasilkan dari penelitian ini adalah sebagai berikut:

1. Dalam melakukan skrining tinjauan pustaka secara sistematis, metode alternatif berupa model *machine learning* yang efisien dan dapat meminimalkan beban kerja peneliti dan praktisi di bidang kedokteran gigi.
2. Model *machine learning* untuk skrining tinjauan pustaka sistematis dengan algoritma yang sudah di optimalkan.
3. Kontribusi kepada komunitas dalam bentuk publikasi *source code* dan artikel.

### **1.8 Sistematika Penulisan**

Penulisan skripsi dipecah menjadi enam bagian. Bab I meliputi latar belakang, rumusan masalah, tujuan penelitian, dan ringkasan dari penelitian ini. Bab II mencakup tinjauan

pustaka dari penelitian sebelumnya yang relevan. Bab III menguraikan metodologi penelitian, mulai dari data yang digunakan, desain studi, langkah-langkah pemrosesan data, dan cara mengoptimalkan model, dan metode evaluasi. Hasil evaluasi dan analisis proses yang dilakukan pada Bab III dijelaskan pada bab selanjutnya, Bab IV. Pada Bab V, penulis membahas manfaat penelitian yang dilakukannya dalam sudut pandang islam. Kesimpulan dan saran penulis akan disampaikan pada bab terakhir yaitu Bab VI.