

BAB 2

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Berikut adalah beberapa penelitian yang bisa dijadikan acuan dan referensi untuk digunakan oleh peneliti sehingga bisa menjadi pembanding dalam penelitian yang peneliti lakukan, yaitu penelitian dari Maria Braileanu, MD, Benjamin B. Risk, PhD, Nadja Kadom, MD, Mark E. Mullins, MD, PhD, FACR, Elizabeth A. Krupinski, PhD, Brent D. Weinberg, MD, PhD (2019) [14], Mohamed El Mohadab, Belaid Bouikhlane, Said Safi (2020) [15], Pradeep Kumar Roy, Sarabjeet Singh Chowdhary, Rocky Bhatia (2020) [16], Honorio Apaza Alanoca, Americo A. Rubin de Celis Vidal, Josimar Edinson Chire Saire (2020) [17], Sujit Amin, Nikita Jayakar, Sonia Sunny, Pheba Babu, M. Kiruthika, Ambarish Gurjar (2019) [12].

Penelitian Maria Braileanu, MD, Benjamin B. Risk, PhD, Nadja Kadom, MD, Mark E. Mullins, MD, PhD, FACR, Elizabeth A. Krupinski, PhD, Brent D. Weinberg, MD, PhD, pada tahun 2019 [14] dengan judul “*Structured curriculum vitae Scoring as a Standardized Tool for Selecting Interview Candidates for Academic Neuroradiology Faculty Position*”. Tujuan dari penelitian ini adalah untuk mengembangkan alat skrining CV, secara retrospektif dengan menerapkan rubrik identifikasi awal kandidat wawancara yang memenuhi syarat untuk posisi Fakultas Neuroradiologi akademik. Indikator penilaian standar CV sebagai alat skrining awal dalam proses *hiring* kedokteran akademik adalah layak. Terdapat 5 variabel yang dipilih sebagai indikator penilaian, yaitu peringkat sekolah kedokteran, jumlah publikasi *manuscript* dan nomor naskah, gelar tambahan, pengalaman mengajar, serta jumlah penghargaan atau penghargaan memiliki hubungan yang signifikan dalam seleksi wawancara. Metode yang digunakan yaitu dengan Regresi Logistik Univariat. Dengan hasil analisis ROC dapat membedakan

pilihan wawancara berdasarkan total penilaian CV adalah 0,69 (interval kepercayaan 95% 0,56-0,82). Model ini 82,4% sensitif, dan 54,1% spesifik.

Penelitian Mohamed El Mohadab, Belaid Bouikhlane, Said Safi pada tahun 2020, [15] dengan judul “*Automatic CV Processing for Research using Data Mining Algorithm*”. Penelitian ini bertujuan untuk meningkatkan sistem informasi, terutama penelitian ilmiah, sebuah studi tentang mengotomatisasi daftar riwayat hidup bagi para peneliti dari berbagai disiplin ilmu yang dimiliki. Penerapan pemrosesan bahasa alami dengan data *mining classifier* Decision Tree disajikan untuk memprediksi bidang pekerjaan. Penggunaan Data Mining pada penelitian ini menggambarkan pendekatan penambahan data dengan automata yang mampu belajar mengidentifikasi tipologi CV dengan hasil pengujian menunjukkan bahwa presisi, recall, dan F-Measures mendekati satu dengan hasil berikut : 0,943, 0,971 dan 0,957.

Penelitian Pradeep Kumar Roy, Sarabjeet Singh Chowdhary, Rocky Bhatia, pada tahun 2020 [16], dengan judul “*A Machine Learning Approach for Automation of Resume Recommendation System*” mengimplementasikan *machine learning* dalam membuat sistem rekomendasi *curriculum vitae* untuk memudahkan proses *screening* dan penentuan kandidat cocok sesuai dengan *job description* dengan menggunakan kesamaan *cosine* dan dengan menggunakan K-NN untuk mengidentifikasi CV. Model yang diusulkan bekerja dalam dua fase: pertama, mengklasifikasikan CV ke dalam kategori yang berbeda. Kedua, merekomendasikan CV berdasarkan indeks kesamaan dengan deskripsi pekerjaan yang diberikan. Pendekatan yang diusulkan secara efektif menangkap wawasan *resume*, semantik mereka dan menghasilkan akurasi 78,53% dengan pengklasifikasi Linear SVM.

Penelitian Honorio Apaza Alanoca, Americo A. Rubin de Celis Vidal, Josimar Edinson Chire Saire [17] pada tahun 2020, dengan judul “*Curriculum Vitae Recommendation Based on Text Mining*”. Penelitian tersebut membuat sistem rekomendasi *curriculum vitae* berbasis *text mining*. Penelitian ini bertujuan untuk mengoptimalkan proses seleksi, yang menghasilkan daftar rekomendasi

berdasarkan kolaborasi atau kesamaan konten. Indikator yang diproses yaitu pengalaman, pengetahuan, dan keterampilan. Metode yang digunakan teknik *text mining*, Natural Language Processing, dan TF-IDF untuk mengidentifikasi CV yang relevan dengan tawaran deskripsi pekerjaan pada situs web. Dengan hasil berupa nilai tertimbang yang dapat digunakan sebagai nilai kualifikasi dari *curriculum vitae* yang relevan untuk direkomendasikan.

Persamaan penelitian ini dengan penelitian-penelitian di atas terletak pada fokus penelitian yaitu mengoptimalkan proses seleksi pada tahap administrasi dengan cara *me-review* CV calon kandidat secara efisien dan efektif. Adapun perbedaan pada penelitian ini yaitu penggunaan metode algoritma yang diimplementasikan untuk membuat model *review* CV. Pada penelitian ini digunakan dengan pendekatan metode Deep Learning menggunakan algoritma Long Short Term Memory (LSTM).

Dari penjelasan di atas, ringkasan dari penelitian sebelumnya yang relevan ditunjukkan pada Tabel 2.1 di bawah ini :

Tabel 2.1 Tabel para penelitian sebelumnya

No	Judul	Penulis, Tahun	Masalah	Algoritma	Hasil	Perbedaan Penelitian yang dilakukan
11	<i>Structured Curriculum Vitae Scoring as a Standardized Tool for Selecting Interview Candidates for Academic Neuroradiology Faculty Position</i>	Maria Braileanu, MD, Benjamin B. Risk, PhD, Nadja Kadom, MD, Mark E. Mullins, MD, PhD, FACR, Elizabeth A. Krupinski, PhD, Brent D. Weinberg, MD, PhD, (2019) [18]	Pemilihan wawancara kandidat untuk posisi fakultas radiologi akademik bervariasi dan busur pada bias yang tidak disadari.	Regresi logistik univariat	hasil analisis ROC dapat membedakan pilihan wawancara berdasarkan total penilaian CV adalah 0,69 (interval kepercayaan 95%, 0,56-0,82). Akurasi model ini 82,4% sensitif, dan 54,1% spesifik.	Perbedaan keduanya terletak pada teknik yang diimplementasikan. Pada penelitian menggunakan teknik pemberian <i>scoring</i> menggunakan <i>Standardized Tool</i> [18]. Sedangkan penelitian ini dilakukan menggunakan teknik Deep Learning berbasis algoritma Long Short Term Memory (LSTM).
22	<i>Automatic CV Processing for Scientific Research using Data Mining Algorithm</i>	Mohamed El Mohadab, Belaid Bouikhalene, Said Safi, (2020) [19]	Layanan administrasi dan tata kelola dalam pengambilan keputusan masih kurang efektif.	Pemrosesan Bahasa Alami, Data mining, Decision tree.	Penggunaan Data Mining pada penelitian ini menggambarkan pendekatan penambangan data dengan automata yang mampu belajar mengidentifikasi tipologi	Perbedaan terletak pada penerapan metode yang digunakan, dimana penelitian Mohamed El Mohadab, Belaid Bouikhalene, Said Safi (2020) yang mengkombinasikan antara data mining dan NLP berbasis algoritma

No	Judul	Penulis, Tahun	Masalah	Algoritma	Hasil	Perbedaan Penelitian yang dilakukan
					CV dengan hasil pengujian menunjukkan bahwa presisi, recall, dan F-Measures mendekati satu dengan hasil berikut : 0,943, 0,971 dan 0,957.	Decission Tree dan Naïve Bayes. Sedangkan pada penelitian ini mengimplementasikan teknik Deep Learning menggunakan algoritma Long Short Term Memory (LSTM) untuk penentuan kelulusan CV dosen di Institut Teknologi Telkom Purwokerto.
33	<i>A Machine Learning approach for automation of Resume</i>	Pradeep Kumar Roy, Sarabjeet Singh Chowdhary, Rocky Bhatiab [23] (2020)	Proses mengklasifikasikan <i>resume</i> kandidat secara manual akan memakan waktu, dan pemborosan sumber daya.	<i>Machine Learning</i> dengan penerapan k-NN dan Linear SVM.	Model berbasis <i>machine learning</i> otomatis dapat merekomendasikan <i>resume</i> kandidat yang sesuai ke SDM berdasarkan deskripsi pekerjaan yang diberikan. Pendekatan yang diusulkan secara efektif menangkap wawasan <i>resume</i> , semantik mereka dan menghasilkan akurasi 78,53% dengan	Perbedaan pada penelitian ini yaitu membuat pengklasifikasian dan kecocokan disetiap parameter yang ada di <i>resume</i> dengan mengimplementasikan teknik machine learning yaitu k-NN dan SVM. Sedangkan pada penelitian ini dilakukan review CV dosen berdasarkan data latih CV kandidat dosen yang sudah lolos dan yang belum lolos, dan selanjutnya dilakukan pembuatan model dengan

No	Judul	Penulis, Tahun	Masalah	Algoritma	Hasil	Perbedaan Penelitian yang dilakukan
					pengklasifikasi Linear SVM.	arsitektur algoritma Recurrent Neural Network dan Heatmap untuk mengevaluasi dari CV kandidat.
44	<i>Curriculum Vitae Recommendation Based on Text Mining</i>	<u>Honorio Apaza Alanoca, Americo A. Rubin de Celis Vidal, Josimar Edinson Chire Saire, (2020)</u> [20]	Proses seleksi yang belum optimal sehingga membutuhkan sistem rekomendasi untuk menghasilkan daftar rekomendasi berdasarkan kolaborasi atau kesamaan konten.	<i>Text Mining</i> dan <i>Natural Language Processing</i> .	TF-IDF untuk mengidentifikasi CV yang relevan dengan tawaran deskripsi pekerjaan pada situs web. Dengan hasil berupa nilai tertimbang yang dapat digunakan sebagai nilai kualifikasi dari <i>curriculum vitae</i> yang relevan untuk direkomendasikan.	Perbedaan terletak pada teknik yang digunakan. Dimana pada penelitian ini menggunakan teknik Text Mining dengan pemrosesan menggunakan NLP. Sedangkan pada penelitian ini berbasis Recurrent Neural Network dan Heatmap.
55	<i>Web Application for Screening Resume</i>	Sujit Amin, Nikita Jayakar, Sonia Sunny, Pheba Babu, M. Kiruthika, Ambarish Gurjar, (2020) [21]	Perekrutan adalah proses yang membosankan di mana tugas pertama bagi setiap perekrut	<i>Machine Learning</i> dan <i>Natural Language Processing</i> .	Aplikasi berbasis <i>web</i> yang digunakan untuk menyaring <i>resume</i> atau <i>curriculum vitae</i> sesuai dengan postingan suatu pekerjaan tertentu.	perbedaan terletak pada tahap proses pembuatan model dan pengimplementasian algoritma yang diterapkan. menggunakan kombinasi dari Natural Language Processing (NLP) dan NER (Named

No	Judul	Penulis, Tahun	Masalah	Algoritma	Hasil	Perbedaan Penelitian yang dilakukan
			adalah menyaring <i>resume</i> .			Entity Recognition) untuk menyaring <i>resume</i> serta melihat kecocokan antara job data dengan <i>dataset resume</i> [21]. Sedangkan penelitian ini menggunakan algoritma Long Short Term Memory (LSTM).

2.2 Landasan Teori

2.2.1 *Curriculum Vitae*

Curriculum Vitae atau biasa disingkat dengan CV merupakan daftar riwayat hidup seseorang, dimana terdapat informasi riwayat pendidikan, biodata diri, *soft skills*, *hard skills*, dan prestasi, serta pencapaian-pencapaian lainnya [22]. CV digunakan sebagai alat promosi untuk melamar pekerjaan yang dibutuhkan pada tahap administrasi. Secara umum, *resume* yang baik dan menarik adalah salah satu kunci HRD untuk menerima kandidat pekerjaan [13].

2.2.2 Institut Teknologi Telkom Purwokerto

Institut Teknologi Telkom Purwokerto merupakan perguruan tinggi yang dikelola oleh Yayasan Pendidikan Telkom dan merupakan satu-satunya lembaga di Jawa Tengah yang fokus pada pengembangan ilmu pengetahuan berbasis teknologi informasi di bidang *Healthcare*, *Agro-Industry*, *Tourism*, dan *Small Medium Enterprise* (HATS). Terdapat tiga fakultas yaitu Fakultas Teknik Telekomunikasi dan Eektro (FTTE), Fakultas Informatika (FIF), dan Fakultas Rekayasa Industri dan Desain (FRID) [23].

2.2.3 *Preprocessing*

Preprocessing merupakan langkah yang dilakukan untuk menghilangkan masalah yang dapat mengganggu hasil pengolahan data. Dari semua data, teks adalah bentuk yang paling tidak terstruktur dan karenanya membutuhkan banyak pembersihan. Langkah *preprocessing* ini membantu untuk mengubah *noise* pada data untuk mendapatkan informasi seakurat mungkin dari teks [28].

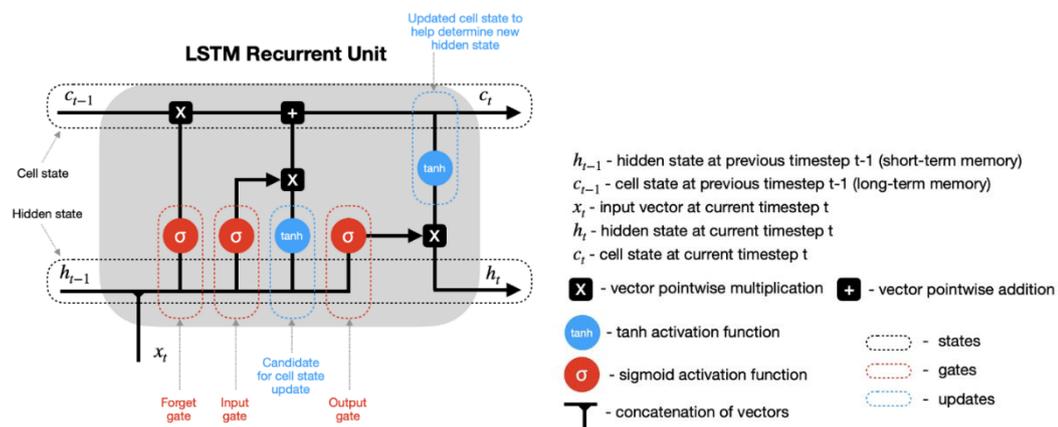
Case folding merupakan tahapan pertama dalam *preprocessing* [24]. *Case folding* dilakukan untuk mengubah huruf dalam suatu dokumen ke format standar atau menjadi huruf kecil atau besar sesuai format yang telah ditetapkan[25]. Namun, proses yang sering dilakukan adalah mengubah karakter dalam teks menjadi huruf kecil, dan hanya huruf "a" hingga "z" yang diproses, dan karakter lain dihilangkan, seperti titik tanda baca (.) atau koma (,) [26].

2.2.4 Deep Learning

Deep Learning adalah bagian dari spesialisasi kecerdasan buatan dalam model jaringan saraf skala besar yang mampu membuat keputusan berbasis data yang akurat, terutama untuk sejumlah besar data [27]. Kemajuan terbaru dalam pembelajaran mendalam telah menunjukkan penerapannya untuk berbagai aplikasi di berbagai bidang [28] yang juga mencakup pelamar untuk tugas genomik. [29].

2.2.5 Long Short Term Memory

Long Short Term Memory (LSTM) ditunjukkan pada Gambar 2.1. LSTM terkenal dengan memori panjang dan banyak digunakan untuk pemodelan sistem dinamis dan prediksi penggunaan LetVdenotes kosakata[30]. Namun, LSTM berisi unit berulang di lapisan tersembunyi, yang memungkinkan algoritma untuk memproses data urutan yang dilakukan dengan berulang kali melewati keadaan tersembunyi dari *timestep* sebelumnya dan menggabungkannya dengan *input* dari yang saat ini. Pelatihan LSTM pada penelitian ini didasarkan pada x dan y . Parameter termasuk *tuple* bentuk *input*, panjang urutan *output*, jumlah *dataset* CV dosen Institut Teknologi Telkom Purwokerto yang lolos dan jumlah *dataset* CV dosen Institut Teknologi Telkom Purwokerto yang belum lolos.



Gambar 2.1 LSTM Recurrent Unit [31]

Gambar 2.1 merupakan diagram yang disederhanakan (bobot dan bias tidak ditampilkan) untuk mempelajari bagaimana unit berulang LSTM memproses informasi.

- a. *Hidden state & new inputs* → keadaan tersembunyi dari *timestep* sebelumnya (h_{t-1}) dan *input* pada langkah waktu saat ini (x_t) digabungkan sebelum melewati salinannya melalui berbagai gerbang.
- b. *Forget gate* → gerbang ini mengontrol informasi apa yang harus dilupakan. Karena fungsi sigmoid berkisar antara 0 dan 1, menetapkan nilai mana dalam keadaan sel yang harus dibuang (dikalikan dengan 0), diingat (dikalikan dengan 1), atau sebagian diingat (dikalikan dengan beberapa nilai antara 0 dan 1).
- c. *Input gate* → membantu mengidentifikasi elemen penting yang perlu ditambahkan ke status sel. Perhatikan bahwa hasil gerbang *input* dikalikan dengan kandidat status sel, dengan hanya informasi yang dianggap penting oleh gerbang *input* yang ditambahkan ke status sel.
- d. *Update cell state* → pertama, keadaan sel sebelumnya (c_{t-1}) dikalikan dengan hasil gerbang lupa. Kemudian kami menambahkan informasi baru dari [gerbang *input* × kandidat negara sel] untuk mendapatkan status sel terbaru (c_t).
- e. *Update hidden state* → bagian terakhir adalah memperbarui status tersembunyi. Keadaan sel terbaru (c_t) dilewatkan melalui fungsi aktivasi tanh dan dikalikan dengan hasil gerbang *output*.
- f. *Final state* → keadaan sel terbaru (c_t) dan keadaan tersembunyi (h_t) kembali ke unit berulang, dan prosesnya berulang pada saat-saat berjalan $t + 1$. *Loop* berlanjut sampai mencapai akhir urutan.