BAB 2
LANDASAN TEORI

Bab ini menguraikan landasan teori dan tinjauan pustaka yang menjadi fondasi ilmiah dalam penelitian ini. Pembahasan secara khusus difokuskan pada konsep, definisi, dan model matematika yang secara langsung berkaitan dengan pengembangan sistem rekomendasi pekerjaan berbasis keterampilan bagi individu tanpa pendidikan formal. Teori utama yang dijabarkan meliputi pendekatan rekrutmen berbasis keahlian (*Skill-Based Recruitment*), konsep sistem rekomendasi *Content-Based Filtering*, tahapan pra-pemrosesan teks (*text preprocessing*), hingga perumusan matematis untuk ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) dan perhitungan kemiripan vektor menggunakan *Cosine Similarity*. Selain itu, bab ini juga menyajikan perumusan metrik evaluasi pemeringkatan yang mencakup *Precision*, *Recall*, *F1-Score*, *Mean Reciprocal Rank (MRR)*, dan *Hit Rate* guna mengukur performa rekomendasi sistem. Seluruh literatur dan teori yang disajikan berfokus pada substansi pemecahan masalah dan dirujuk dari berbagai publikasi ilmiah terkini.

2.1 Tinjauan Pustaka

Penelitian mengenai sistem rekomendasi pekerjaan berbasis pengolahan teks telah banyak dikembangkan untuk mengatasi permasalahan *information overload* pada platform rekrutmen. Beberapa penelitian terdahulu yang menggunakan metode serupa menjadi acuan sekaligus titik tolak untuk melihat kebaruan (*novelty*) dari penelitian ini.

Penelitian pertama dilakukan oleh Nadar dan Athulya (2020) yang mengusulkan sistem rekomendasi pekerjaan menggunakan algoritma *TF-IDF* dan *Cosine Similarity*. Sistem tersebut terbukti efektif dalam mencocokkan teks pada resume pelamar dengan deskripsi pekerjaan untuk mempercepat proses screening oleh rekruter [7]. Meskipun model ini menghasilkan tingkat akurasi yang baik, fitur yang digunakan masih sangat bergantung pada struktur resume konvensional yang secara implisit menempatkan latar belakang pendidikan sebagai parameter pembobotan yang tinggi.

Penelitian kedua dilakukan oleh Fitria dkk. (2024) yang menerapkan metode *Content-Based Filtering* dengan algoritma *TF-IDF* dan *Cosine Similarity*

untuk merekomendasikan lowongan pekerjaan di pasar kerja Indonesia. Penelitian tersebut membuktikan bahwa TF-IDF sangat efektif dalam mengekstraksi fitur penting dari teks deskriptif profil pengguna untuk dicocokkan dengan deskripsi lowongan [9]. Namun, sistem yang dikembangkan tersebut masih bersifat umum dan belum secara khusus memecah atribut portofolio seperti spesifikasi alat (*tools*) maupun jumlah proyek (*project count*) yang krusial sebagai bukti kompetensi bagi pekerja nonformal tanpa ijazah.

Penelitian ketiga diterbitkan dalam Journal of Computer Science and Research (2023), yang mengkaji pembangunan sistem rekomendasi pekerjaan di bidang IT menggunakan algoritma *Content-Based Filtering*. Penelitian ini secara spesifik mengekstraksi atribut *skills* pengguna untuk dicocokkan dengan *job description* menggunakan metode statistik *TF-IDF* [10]. Meskipun penelitian ini berhasil mengidentifikasi *area skill* yang perlu ditingkatkan oleh pencari kerja, sistem tersebut belum mengintegrasikan variabel tingkatan kemahiran (*skill level*) serta pembuktian proyek sebagai pendorong utama bobot relevansi.

Penelitian keempat dipublikasikan dalam Jurnal Techno.Com (2025), yang berfokus pada analisis kesesuaian lowongan pekerjaan dan kemampuan pencari kerja menggunakan *Cosine Similarity* dan ekstraksi *TF-IDF*. Penelitian ini mengevaluasi sejauh mana resume pencari kerja dari dataset publik (Kaggle) sesuai dengan deskripsi pekerjaan, serta memberikan rekomendasi *skill* tambahan [11]. Akan tetapi, penelitian ini masih bergantung pada pengolahan *resume* formal secara keseluruhan, sehingga potensi bias terhadap kualifikasi non-keterampilan (seperti riwayat akademik formal) belum sepenuhnya tereliminasi.

Di sisi lain, laporan global dari World Economic Forum (2023) menegaskan adanya pergeseran paradigma rekrutmen menuju pendekatan *skills-first*. Laporan tersebut menyoroti bahwa industri kini semakin memprioritaskan validasi keterampilan nyata (seperti portofolio dan penguasaan *tools*) dibandingkan kredensial akademik [12]. Hal ini menciptakan celah penelitian (*research gap*) pada sistem rekrutmen saat ini yang masih kaku terhadap ijazah.

Berdasarkan tinjauan pustaka tersebut, penelitian ini mengisi celah dengan mengembangkan sistem rekomendasi yang secara eksklusif membuang fitur pendidikan formal. Pembobotan teks sepenuhnya difokuskan pada ekstraksi keterampilan utama (*main skill*), perangkat yang dikuasai (*tools*), dan riwayat proyek. Pendekatan ini dirancang untuk menciptakan pencocokan kerja yang lebih inklusif bagi individu tanpa latar belakang pendidikan tinggi.

Tabel 2.1. Perbandingan Penelitian Terdahulu

Peneliti (Tahun)	**Metode Utama**	**Fitur yang Diekstraksi**	**Perbedaan Pendekatan**
Nadar & Athulya (2020) [7]	*TF-IDF, Cosine Similarity*.	Edukasi, Pengalaman Kerja, Keterampilan dasar.	Menggunakan metode yang sama namun membuang fitur edukasi; murni berbasis *skill* dan *tools*.
Fitria dkk. (2024) [9]	*Content-Based Filtering, TF-IDF, Cosine Similarity.*	Teks deskriptif profil pengguna dan lowongan kerja secara umum.	Menambahkan pembobotan spesifik pada spesifikasi *tools* dan level proyek sebagai proksi kompetensi pekerja nonformal.
Jurnal JCSR (2023) [10]	*Content-Based Filtering, TF-IDF.*	Spesifikasi skills di bidang IT dan Job Description.	Menambahkan parameter *skill level* (tingkat kemahiran) dan jumlah proyek.
Jurnal Techno.Com (2025) [11]	*TF-IDF, Cosine Similarity, Analisis Teks.*	*Resume* utuh dari kandidat dan deskripsi pekerjaan.	Tidak mengekstraksi *resume* utuh, melainkan hanya poin inti keahlian praktis nonformal.
Penelitian Ini.	*Hybrid Content-Based Filtering (TF-IDF, Cosine Similarity, Metadata Weighting).*	*Similarity Score, Main Skill, Tools, Stability Score.*	Fokus pada inklusivitas pendekatan *skills-first*, menerapkan pembobotan hybrid (penggabungan kemiripan teks portofolio dan pencocokan persis pada *skill/level*) untuk meningkatkan presisi rekomendasi.

Mengacu pada celah penelitian (*research gap*) yang telah diuraikan pada Tabel 2.1, perancangan sistem rekomendasi ini juga sangat didorong oleh kondisi faktual dan urgensi ketenagakerjaan saat ini. Berikut adalah penjabaran data statistik dan landasan mengapa penelitian ini difokuskan pada individu tanpa pendidikan formal:

2.1.1 Statistika Pengangguran dan Fenomena *Mismatch* Pendidikan-Pekerjaan

Berdasarkan data Badan Pusat Statistik (BPS) pada awal tahun 2025, angka pengangguran di Indonesia masih berada di angka jutaan jiwa [13]. Paradoksnya,

banyak lowongan di industri tidak terisi secara optimal atau pekerja justru terjebak dalam *horizontal mismatch*, yaitu kondisi di mana mereka bekerja di bidang yang sama sekali tidak relevan dengan latar belakang pendidikan formalnya [14]. Fenomena ini diperkuat oleh temuan dalam Jurnal Ketenagakerjaan (J-Naker) yang menyatakan bahwa ketidaksesuaian pendidikan dan pekerjaan merupakan tantangan struktural yang signifikan bagi pemuda Indonesia di pasar kerja [15].

Lebih lanjut, laporan BPS bertajuk Mismatch Pendidikan–Pekerjaan Pemuda Indonesia: Implikasi Bagi Bonus Demografi mengungkapkan bahwa masalah ini tidak hanya terjadi secara horizontal, tetapi juga vertikal. Tercatat sebanyak 35,36% pemuda di Indonesia mengalami *vertical mismatch*, yang berarti mereka bekerja pada posisi yang tidak sesuai dengan tingkat pendidikannya [16]. Sebagaimana dikutip oleh Kumparan, angka tersebut terdiri dari 22,36% kelompok *overeducated* (pendidikan lebih tinggi dari syarat jabatan) dan 13% kelompok *undereducated* (pendidikan lebih rendah dari syarat jabatan) [17].

Kondisi ini menegaskan urgensi pengembangan sistem rekrutmen yang tidak lagi bergantung pada ijazah semata. Dengan tingginya angka *vertical mismatch*, pendekatan berbasis keterampilan (*skills-first*) menjadi solusi krusial untuk memvalidasi kemampuan praktis individu, terutama bagi mereka yang memiliki kompetensi namun terhalang oleh batasan administratif gelar formal.

2.1.2 Ketimpangan Kompetensi pada Sektor TIK dan Ekonomi Kreatif

Penelitian ini secara khusus mengakomodasi pencocokan untuk tujuh bidang keahlian utama, yaitu *programming, design, art, animation, music, writing,* dan *editing.* Ketujuh keahlian ini secara garis besar terbagi ke dalam dua sektor industri krusial di Indonesia, yakni sektor Teknologi Informasi dan Komunikasi (TIK) serta sektor Ekonomi Kreatif. Kedua sektor tersebut saat ini menghadapi tantangan serius berupa *horizontal mismatch* dan *skill gaps* (kesenjangan keterampilan).

Pada ranah keahlian teknis seperti *programming* dan *design*, riset mengenai angkatan kerja digital di Indonesia menunjukkan bahwa lulusan pendidikan vokasi dengan jurusan Komputer dan Informatika menyumbang angka pengangguran tertinggi kedua, yakni lebih dari 246.000 jiwa [18]. Hal ini mengindikasikan terjadinya *horizontal mismatch* yang masif, di mana individu yang memiliki dasar keterampilan teknologi tidak berhasil terserap oleh industri TIK dan terpaksa bekerja di bidang yang sama sekali tidak relevan. Di sisi lain, bagi kandidat yang berhasil direkrut pada posisi linier, masalah kompetensi tetap menjadi ancaman.

Laporan dari The SMERU Research Institute (2024) menyoroti fenomena pada institusi pemerintahan dan sektor digital, di mana banyak pekerja pada posisi fungsional TIK dan sistem informasi justru masuk dalam kategori *unqualified* atau kurang mumpuni [19]. Mereka direkrut berdasarkan persyaratan administratif gelar formal, namun gagal mengimbangi kebutuhan kompetensi teknis tingkat lanjut seperti penggunaan perangkat (*tools*) dan bahasa pemrograman modern. Fenomena ini merujuk pada definisi *Skill Gaps* dari ILO, yaitu kondisi di mana pekerja mengisi suatu jabatan namun tidak memiliki keterampilan aktual yang dibutuhkan pasar kerja.

Pola serupa juga terjadi pada ranah Ekonomi Kreatif yang menaungi keahlian *design, art, animation, music, writing,* dan *editing*. Tingginya tingkat *mismatch* yang telah dijabarkan sebelumnya [16] turut berdampak pada sektor ini, di mana ranah jasa dan kesenian banyak diisi oleh pekerja lintas jurusan. Hal ini terjadi akibat sulitnya talenta mumpuni menembus sistem rekrutmen yang kaku tanpa adanya deteksi portofolio yang memadai. Kebutuhan akan karya kreatif seringkali diisi oleh pekerja yang memiliki ijazah relevan (misalnya sarjana sastra untuk penulisan), namun minim keterampilan praktis terkait *tools* industri standar. Akibatnya, kualitas *output* tidak memenuhi standar industri kreatif, dan masuk ke dalam kategori *under-skilled*. Di saat yang sama, banyak talenta otodidak mumpuni di bidang seni, animasi, hingga pengeditan justru tidak tersaring oleh *Applicant Tracking System (ATS)* konvensional dan berujung mengalami *underemployed* di sektor lain. Fakta-fakta ini secara kuat menjustifikasi mengapa sistem rekomendasi pekerjaan harus beralih dari parameter pendidikan formal menuju pendekatan *skills-first* yang menitikberatkan pada pembuktian alat bantu (*tools*) dan portofolio proyek.

2.1.3 Talenta Jalur Nonformal yang Tidak Diakui (*Unacknowledged*)

Merespons tingginya kasus *mismatch* pada jalur formal, penelitian ini mengambil fokus pada individu yang membangun keahlian teknisnya secara mandiri. Di era digital saat ini, banyak talenta memperoleh kemampuan teknis yang mumpuni melalui jalur alternatif seperti *bootcamp*, kursus daring, magang, maupun belajar otodidak. Namun, tanpa adanya ijazah atau gelar akademik formal yang linier, kemampuan riil mereka seringkali tidak diakui (*unacknowledged*) oleh sistem *Applicant Tracking System (ATS)* konvensional. Laporan dari Harvard Business School menyoroti bahwa filter otomatis pada ATS yang terlalu kaku terhadap

syarat gelar seringkali menciptakan ”pekerja tersembunyi” (*hidden workers*), di mana kandidat yang sebenarnya memiliki keterampilan praktis justru tereliminasi di tahap awal [20]. Kelompok ini rentan menjadi *underemployed* (bekerja di bawah kapasitas) padahal keahlian mereka sangat dibutuhkan industri.

2.1.4 Validasi Keterampilan melalui Portofolio

Tantangan utama dalam merekrut kandidat tanpa pendidikan formal yang linier adalah cara memastikan validitas dari keterampilan yang mereka klaim. Berdasarkan literatur dari International Labour Organization (ILO), keahlian teknis praktis dapat dibuktikan dan divalidasi secara kuat melalui bukti rekam jejak pembelajaran atau portofolio hasil karya [21]. Oleh karena itu, dalam penelitian ini, parameter seperti perangkat (*tools*) yang dikuasai, tingkat keahlian (*skill level*), dan riwayat jumlah proyek (*project count*) diekstrak sebagai representasi konkret dari portofolio. Teks dari portofolio ini memberikan bukti objektif yang kemudian diolah menggunakan algoritma TF-IDF untuk mencocokkan kemampuan praktis kandidat secara langsung dengan kebutuhan spesifik perekrut.

2.1.5 Dampak Pendekatan Berbasis Keterampilan

Penerapan sistem rekomendasi berbasis keterampilan memiliki dampak sosial-ekonomi yang sangat signifikan. Laporan World Economic Forum (2023) memperkirakan bahwa pergeseran menuju rekrutmen berbasis keterampilan (*skills-first approach*) dapat membebaskan lebih dari 100 juta pekerja global dari hambatan syarat gelar formal. Bagi perusahaan, pendekatan ini memperluas talent pool potensial hingga hampir sepuluh kali lipat [22]. Dengan demikian, pengembangan sistem rekomendasi pekerjaan menggunakan algoritma *Content-Based Filtering* pada penelitian ini tidak hanya berkontribusi secara akademis, tetapi juga memberikan dampak nyata dalam menciptakan proses pencocokan pasar kerja yang lebih adil, inklusif, dan tepat sasaran.

2.2 Tinjauan Teori

2.2.1 *Skill-Based Recruitment*

Skill-Based Recruitment atau rekrutmen berbasis keterampilan adalah metode pengadaan tenaga kerja yang menitikberatkan evaluasi pada kompetensi

teknis, kemampuan praktis, dan portofolio kandidat, dibandingkan menjadikan ijazah atau kredensial akademik sebagai *filter* utama. Menurut laporan dari International Labour Organization (ILO) pada tahun 2024, pendekatan *skills-first* menjadi strategi krusial secara global untuk mengatasi ketimpangan pasar kerja (*skills mismatch*) dan memberikan peluang yang setara bagi individu yang memperoleh keahlian melalui jalur nonformal [21].

Lebih lanjut, riset dari LinkedIn Economic Graph (2024) membuktikan bahwa ketika perusahaan beralih menggunakan pencarian berbasis keterampilan alih-alih gelar pendidikan, ukuran kelompok bakat (*talent pool*) yang dapat dijangkau meluas hingga hampir 10 kali lipat [23]. Hal ini mendasari perancangan sistem dalam penelitian ini, di mana ekstraksi teks difokuskan secara eksklusif pada parameter keahlian dan alat (*tools*) guna menciptakan sistem penyaringan kandidat yang lebih objektif dan inklusif.

2.2.2 Sistem Rekomendasi (*Content-Based Filtering*)

Sistem rekomendasi adalah perangkat lunak dan teknik yang memberikan saran item yang dinilai paling relevan bagi pengguna tertentu. Dalam domain rekrutmen tenaga kerja, pendekatan yang paling umum digunakan adalah *Content-Based Filtering* (CBF). Menurut Ricci dkk. (2015), metode CBF bekerja dengan cara merekomendasikan item (lowongan pekerjaan) yang memiliki kemiripan atribut dengan profil atau karakteristik kompetensi yang dimiliki oleh pengguna (*user profile*) [24]. Pendekatan CBF sepenuhnya independen dari interaksi atau data historis pengguna lain (*community behaviors*), melainkan bertumpu murni pada analisis kecocokan konten teks (*content matching*) yang melekat pada entitas dokumen terkait.

Secara arsitektural, menurut Aggarwal (2016) [8] serta Lops dkk. (2011) [25], efektivitas operasional sistem rekomendasi berbasis konten sangat bergantung pada tiga modul atau komponen utama sebagai berikut:

1. Penganalisis Konten (*Content Analyzer*): Komponen ini berfungsi untuk mengekstrak dan merepresentasikan informasi dari item (dalam penelitian ini berupa teks deskripsi lowongan pekerjaan atau *job description*) ke dalam bentuk representasi fitur yang terstruktur. Proses ini umumnya melibatkan teknik prapemrosesan teks (*text preprocessing*) dan ekstraksi bobot statistik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-

IDF) untuk mengonversi teks tidak terstruktur menjadi bentuk representasi vektor (*Vector Space Model*).

2. Pembelajar Profil (*Profile Learner*): Komponen ini bertugas untuk mengonstruksi entitas profil pengguna (kandidat pekerja) berdasarkan riwayat data, kemampuan, portofolio proyek, maupun kueri preferensi yang dimasukkan. Pembelajar profil akan merangkum karakteristik dokumen kandidat ke dalam representasi vektor yang sejenis dengan vektor lowongan pekerjaan.

3. Komponen Penyaringan (*Filtering Component*): Komponen ini merupakan modul penentu keputusan yang mengevaluasi tingkat kemiripan (*similarity matrix*) antara vektor profil kandidat pekerja dan vektor deskripsi lowongan pekerjaan. Algoritma komputasi seperti *Cosine Similarity* diaplikasikan pada tahap ini untuk mengukur kedekatan sudut antar-vektor guna menghasilkan urutan rekomendasi berdasarkan nilai pemeringkatan akhir (*ranking score*).

Integrasi ketiga komponen arsitektur di atas memposisikan sistem CBF sebagai instrumen pencarian informasi (*Information Retrieval*) yang transparan dan dapat dijelaskan (*explainable*). Karakteristik ini sangat krusial dalam mengatasi fenomena ketidaksesuaian hubungan kerja (*skill mismatch*), karena sistem mampu secara objektif mengukur tingkat relevansi kompetensi kandidat talenta berdasarkan bukti-bukti teks portofolio aktual mereka.

2.2.3 *Text Preprocessing*

Sebelum teks profil pengguna dan deskripsi pekerjaan diproses oleh algoritma komputasi, teks tersebut harus melalui tahapan pra-pemrosesan (*text preprocessing*). Tujuannya adalah untuk mengurangi noise dan menstandardisasi format kata sehingga mesin dapat membaca fitur teks dengan optimal. Menurut Adriani dkk. (2007) dalam penelitiannya mengenai pemrosesan teks, tahapan utama yang umumnya dilakukan meliputi [26]:

1. *Case Folding*: Mengubah seluruh huruf dalam dokumen teks menjadi huruf kecil (*lowercase*).

2. *Tokenizing*: Memecah kalimat menjadi kumpulan kata tunggal atau unit terkecil yang disebut *token*.

3. *Stopword Removal (Filtering)*: Menghapus kata-kata hubung atau kata umum yang tidak memiliki nilai informasi spesifik.

4. *Stemming*: Mengembalikan kata yang memiliki imbuhan menjadi kata dasar untuk menyamakan makna komputasional.

2.2.4 *Term Frequency - Inverse Document Frequency (TF-IDF)*

TF-IDF adalah metode pembobotan statistik yang digunakan dalam *Information Retrieval* untuk mengevaluasi seberapa penting sebuah kata dalam suatu dokumen di dalam sebuah kumpulan dokumen (korpus). Manning dkk. (2008) menjelaskan bahwa TF-IDF memastikan kata kunci spesifik mendapat bobot lebih tinggi daripada kata-kata yang terlalu sering muncul secara umum [27]. Rumus matematis untuk TF-IDF diuraikan sebagai berikut:

1. Term Frequency (TF): Rumus 2.1 Mengukur frekuensi kemunculan term t dalam dokumen d.

$$TF(t,d) = \frac{f_{t,d}}{\sum_{t \in d} f_{t,d}} \tag{2.1}$$

Keterangan:

- $TF(t,d)$ = frekuensi kemunculan term t dalam dokumen d
- $f_{t,d}$ = jumlah kemunculan term t dalam dokumen d
- $\sum_{t' \in d} f_{t',d}$ = total keseluruhan kemunculan seluruh term dalam dokumen d

2. Inverse Document Frequency (IDF): Rumus 2.2 Mengukur nilai kelangkaan suatu term t di seluruh korpus dokumen D. Semakin sedikit dokumen yang mengandung term t, semakin besar nilai IDF-nya.

$$IDF(t,D) = \log\left(\frac{N}{df_t}\right) \tag{2.2}$$

Keterangan:

- $IDF(t,D)$ = nilai kelangkaan (inverse document frequency) dari term t

- N = total jumlah dokumen di dalam korpus (kumpulan profil/lowongan)
- df_t = jumlah dokumen yang mengandung term t

Di mana N adalah total jumlah dokumen dalam korpus, dan df_t adalah jumlah dokumen yang mengandung kata t.

3. Bobot TF-IDF: Rumus 2.3 Bobot akhir teks didapatkan dengan mengalikan nilai TF dan IDF.

$$W_{t,d} = TF(t,d) \times IDF(t,D) \tag{2.3}$$

Keterangan:

- $W_{t,d}$ = bobot akhir TF-IDF untuk term t pada dokumen d
- $TF(t,d)$ = nilai term frequency
- $IDF(t,D)$ = nilai inverse document frequency

2.2.5 *Cosine Similarity*

Setelah dokumen teks diubah menjadi representasi vektor numerik melalui TF-IDF, *Cosine Similarity* digunakan untuk mengukur tingkat kemiripan antara dua vektor teks tersebut. Menurut Manning dkk. (2008), metode ini menghitung kosinus sudut di antara dua vektor [27]. Semakin kecil sudut antar vektor, semakin nilai kosinusnya mendekati angka 1, yang mengindikasikan tingkat kecocokan yang sangat tinggi.

Rumus 2.4 Rumus *Cosine Similarity* didefinisikan sebagai:

$$\text{Similarity}(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{2.4}$$

Keterangan:

- $Similarity(A,B)$ = nilai kemiripan kosinus antara vektor A dan B
- A = vektor representasi numerik dari profil pengguna
- B = vektor representasi numerik dari deskripsi lowongan pekerjaan

- A_i = bobot term ke-i pada vektor pengguna A
- B_i = bobot term ke-i pada vektor pekerjaan B
- n = jumlah total term unik (dimensi vektor) dalam korpus

Di mana A merepresentasikan vektor dari profil pengguna, B merepresentasikan vektor dari deskripsi lowongan pekerjaan, serta A_i dan B_i adalah bobot term ke-i pada masing-masing vektor.

2.2.6 Pendekatan *Hybrid Content-Based Filtering* dan *Rule-Based Weighting*

Dalam sistem rekomendasi berbasis konten (*Content-Based Filtering*), ketergantungan murni pada kemiripan teks (seperti TF-IDF) seringkali kurang optimal jika entitas memiliki atribut kategorikal yang bersifat mutlak, seperti tingkat keahlian atau kategori keahlian utama. Untuk mengatasi hal ini, diterapkan pendekatan Hybrid atau kombinasi pembobotan (*Rule-Based Weighting*). Menurut Aggarwal (2016) [8], sistem hibrida dapat menggabungkan berbagai skor kemiripan fitur menjadi satu rumusan linier untuk meningkatkan akurasi sistem. Dalam konteks penelitian ini, skor akhir (*Final Similarity*) dihitung dengan mengombinasikan skor *Cosine Similarity* dari ekstraksi teks, dengan skor pencocokan persis (*Exact Match*) pada fitur *Main Skill* dan *Skill Level* menggunakan persamaan pembobotan linier.

Menurut Burke (2002) [28] dalam penelitiannya yang berjudul "*Hybrid Recommender Systems: Survey and Experiments*", hibridasi dalam sistem rekomendasi tidak hanya terbatas pada penggabungan antara metode *Collaborative Filtering* dan *Content-Based Filtering*. Burke menegaskan bahwa sistem hibrida dapat dibangun dengan mengombinasikan berbagai komponen rekomendasi (*recommendation components*) atau teknik representasi fitur yang berbeda untuk saling melengkapi kelemahan masing-masing paradigma tunggal.

Berdasarkan taksonomi hibridasi yang dikemukakan oleh Burke (2002), metode yang diterapkan dalam penelitian ini diklasifikasikan secara spesifik sebagai *Weighted Hybrid Recommender System*. Mekanisme Weighted Hybrid bekerja dengan cara menggabungkan lembar skor (*score sheets*) dari beberapa komponen analisis fitur yang berbeda secara linier untuk menghasilkan satu nilai bobot pemeringkatan tunggal (*single combined recommendation ranking*).

Dalam konteks penelitian sistem rekomendasi pekerjaan ini, sifat hibrida (*hybrid*) diwujudkan melalui penggabungan dua komponen evaluasi berbasis

konten, yaitu:

1. Komponen Tidak Terstruktur (*Unstructured Content Component*): Menggunakan algoritma *Term Frequency-Inverse Document Frequency (TF-IDF)* dan *Cosine Similarity* untuk mengekstrak dan menghitung kemiripan tekstual pada narasi portofolio proyek dan deskripsi pekerjaan.
2. Komponen Terstruktur (*Knowledge-Based/Metadata Component*): Menggunakan pendekatan aturan logis (*Rule-Based Matching*) untuk memvalidasi atribut kategorikal mutlak rekrutmen berupa *main skill*, *skill level*, dan kalkulasi *stability score*.

Penggabungan kedua komponen ini melalui kombinasi linier terbobot menghasilkan arsitektur hibrida yang tangguh karena mampu menilai kedalaman kompetensi kandidat (melalui teks portofolio) sekaligus menjaga kepatuhan terhadap batasan administratif formal rekrutmen (melalui aturan metadata).

Rumus 3.1 *Hybrid Similarity* (Skor Akhir) didefinisikan sebagai:

$$\text{Final\_Score} = (0.45 \times S_{tfidf}) + (0.35 \times S_{skill}) + (0.10 \times S_{level}) + (0.10 \times S_{stability}) \tag{2.5}$$

Keterangan:

- Final_Score = nilai kemiripan akhir (probabilitas kecocokan) antara profil pengguna dan deskripsi lowongan pekerjaan
- S_{tfidf} = skor *Cosine Similarity* yang didapatkan dari ekstraksi teks portofolio menggunakan algoritma TF-IDF (memiliki bobot 0.45 atau 45%)
- S_{skill} = skor *Exact Match* untuk fitur *Main Skill*. Bernilai 1 jika keahlian utama pengguna sama persis dengan yang dicari rekruter, dan bernilai 0 jika berbeda (memiliki bobot 0.35 atau 35%)
- S_{level} = skor *Exact Match* untuk fitur tingkat keahlian (*Skill Level*). Bernilai 1 jika tingkat keahlian cocok (misal: *Junior* = *Junior*), dan bernilai 0 jika tidak cocok (memiliki bobot 0.10 atau 10%)
- $S_{stability}$ = skor stabilitas (*Stability Score*) kandidat pekerja yang merepresentasikan metrik loyalitas atau potensi retensi kerja berdasarkan rekam jejaknya (memiliki bobot 0.10 atau 10%)

2.2.7 Metrik Evaluasi Sistem Rekomendasi

Sistem rekomendasi teks umumnya dievaluasi menggunakan metrik berbasis pemeringkatan (*ranking-based metrics*) yang diadaptasi dari disiplin *ilmu Information Retrieval*. Karena pengguna atau rekruter umumnya hanya memperhatikan beberapa item teratas (Top-K) yang disajikan oleh sistem [29], penelitian ini menggunakan metrik Precision@K, Recall@K, dan F1-Score@K dengan nilai batas (K) 3 dan 5. Berdasarkan rumusan dari Manning dkk. (2009) [27], penjabaran evaluasi tersebut adalah sebagai berikut:

1. Precision@K: Rumus 2.6 Precision pada Top-K mengukur proporsi item yang benar-benar relevan dari sejumlah K item teratas yang direkomendasikan sistem.

$$Precision@K = \frac{\text{Jumlah kandidat relevan pada peringkat } K}{K} \tag{2.6}$$

   Keterangan:

   - Precision@K = rasio ketepatan rekomendasi pada K urutan teratas
   - K = batas pemotongan jumlah item rekomendasi yang dievaluasi (Top-K)

2. Recall@K: Rumus 2.7 Recall pada Top-K mengukur seberapa banyak item yang relevan yang berhasil ditemukan oleh sistem di dalam batas K peringkat teratas, dibandingkan dengan keseluruhan *ground truth* (total data yang seharusnya relevan).

$$Recall@K = \frac{\text{Jumlah kandidat relevan pada peringkat } K}{\text{Total kandidat yang relevan pada sistem}} \tag{2.7}$$

   Keterangan:

   - Recall@K = rasio perolehan kandidat relevan pada K urutan teratas
   - K = batas pemotongan jumlah item rekomendasi yang dievaluasi (Top-K)

3. F1-Score@K: Rumus 2.3 F1-Score merupakan nilai *harmonic mean* (rata-rata harmonik) yang menggabungkan dan menyeimbangkan metrik *Precision* dan *Recall*. Metrik ini krusial untuk memastikan bahwa sistem tidak hanya akurat menebak pada urutan teratas, tetapi juga tidak mengabaikan kandidat yang seharusnya relevan.

$$\text{F1-Score@}K = 2 \times \frac{\text{Precision@}K \times \text{Recall@}K}{\text{Precision@}K + \text{Recall@}K} \tag{2.8}$$

Keterangan:

- F1-Score@K = rata-rata harmonik dari *Precision* dan *Recall* pada Top-K
- Precision@K = nilai Precision pada batas K
- Recall@K = nilai *Recall* pada batas K

4. *Mean Reciprocal Rank (MRR):* Rumus 2.9 *Mean Reciprocal Rank* (MRR) digunakan untuk mengevaluasi efektivitas posisi penempatan peringkat rekomendasi item. MRR mengukur seberapa cepat sistem dapat menempatkan kandidat relevan pertama pada posisi teratas dari keseluruhan kueri pencarian.

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i} \tag{2.9}$$

Keterangan:

- MRR = nilai rata-rata kebalikan peringkat (*reciprocal rank*) dari sistem
- $|Q|$ = total jumlah kueri pencarian rekruter yang dievaluasi dalam pengujian
- rank_i = posisi peringkat pertama di mana kandidat relevan ditemukan pada kueri ke-i

5. Hit Rate@K: Rumus 2.10 *Hit Rate* pada *Top-K* mengukur probabilitas keberhasilan sistem dalam menyajikan setidaknya satu buah rekomendasi kandidat yang tepat sasaran di dalam batas daftar K teratas yang ditampilkan kepada rekruter.

$$\text{Hit Rate@}K = \frac{\sum_{q \in Q} \text{is\_hit}(q, K)}{|Q|} \tag{2.10}$$

Keterangan:

- Hit Rate@K = nilai persentase keberhasilan tembakan rekomendasi kandidat pada batas K
- is_hit(q, K) = fungsi kondisi bernilai 1 jika terdapat minimal satu kandidat relevan dalam daftar *Top-K* pada kueri q, dan bernilai 0 jika tidak ditemukan kecocokan sama sekali
- $|Q|$ = total jumlah seluruh kueri pencarian yang diuji dalam sistem

2.3 Tools

Dalam tahap perancangan, implementasi, dan pengujian model sistem rekomendasi, penelitian ini memanfaatkan beberapa perangkat lunak dan bahasa pemrograman sebagai berikut:

2.3.1 Python

Python adalah bahasa pemrograman tingkat tinggi yang berorientasi objek dan memiliki sintaks yang dirancang untuk mudah dibaca serta dipahami. Menurut McKinney (2022), Python telah menjadi bahasa standar de facto dalam analisis data, *machine learning*, dan pemrosesan bahasa alami (*Natural Language Processing*) karena dukungan ekosistem pustaka (*library*) yang sangat ekstensif [30]. Dalam penelitian ini, Python digunakan sebagai bahasa utama untuk mengembangkan seluruh alur kerja komputasi, mulai dari generasi dataset sintetis dengan pustaka pandas dan numpy, hingga ekstraksi fitur teks (TF-IDF) dan perhitungan kemiripan kosinus (*Cosine Similarity*) menggunakan pustaka scikit-learn.

2.3.2 Visual Studio Code

Visual Studio Code (VS Code) adalah *Integrated Development Environment (IDE)* berbasis *open-source* yang dikembangkan oleh Microsoft. Perangkat lunak ini menyediakan lingkungan penulisan kode yang komprehensif, dilengkapi dengan fitur *syntax highlighting, intelligent code completion (IntelliSense), debugging*

terintegrasi, serta manajemen kontrol versi (*version control*) seperti Git secara bawaan [31]. VS Code digunakan dalam penelitian ini sebagai teks editor utama untuk menulis, mengelola struktur repositori, dan merapikan kode program secara lokal sebelum dieksekusi atau diunggah ke lingkungan berbasis cloud.

2.3.3 Google Colab

Google Colaboratory, atau yang lebih dikenal dengan sebutan Google Colab, adalah lingkungan eksekusi Jupyter Notebook berbasis cloud yang disediakan secara gratis oleh Google. Google Colab memungkinkan pengembang untuk menulis dan mengeksekusi kode Python langsung di dalam peramban (*browser*) tanpa memerlukan konfigurasi lingkungan lokal yang rumit [32]. Keunggulan utama dari Google Colab adalah kemampuannya dalam menyediakan akses ke sumber daya komputasi tinggi secara gratis, sehingga mempermudah proses evaluasi model *machine learning* (seperti pengujian metrik Precision dan Recall pada dataset) dengan cepat dan efisien.