# BAB I
# PENDAHULUAN

## 1.1 Latar Belakang

Pada era persaingan pasar bebas *ASEAN* (*MEA*), perguruan tinggi, khususnya pendidikan vokasi yang berorientasi pada penerapan keahlian tertentu, dituntut untuk mampu menghasilkan lulusan yang memiliki daya saing tinggi dengan lulusan dari berbagai negara. Kondisi persaingan global yang semakin ketat ini mendorong para profesional untuk terus berinovasi, kreatif, serta meningkatkan tingkat keterampilan, kompetensi, dan profesionalitas yang dimilikinya secara berkelanjutan. Dalam dunia kerja modern, penguasaan hard skill dan soft skill memiliki pengaruh yang sangat signifikan terhadap kinerja dan prestasi kerja karyawan. Kedua jenis keterampilan tersebut sama-sama memiliki peran penting dan saling melengkapi satu sama lain dalam menunjang keberhasilan individu di lingkungan kerja (Vidyasari & Listiawati, 2022).

Klasifikasi yang digunakan pada portal lowongan kerja umumnya masih terbatas pada pengelompokan berdasarkan bidang pekerjaan, tanpa adanya penjabaran rinci mengenai keterampilan yang dibutuhkan. Oleh karena itu, untuk memperoleh pemetaan kebutuhan talenta digital yang lebih akurat, data lowongan kerja dari portal tersebut perlu diklasifikasikan terlebih dahulu guna mengidentifikasi lowongan yang secara spesifik memerlukan talenta digital. Inkonsistensi gaya penulisan serta format deskripsi pekerjaan yang bervariasi antar perusahaan menimbulkan hambatan dalam pemetaan kebutuhan industri secara cepat, baik bagi pencari kerja maupun institusi pendidikan (Geraldy et al., 2025).

Meskipun volume data lowongan pekerjaan tersedia secara masif (*big data*), informasi kualifikasi di dalamnya umumnya bersifat tidak terstruktur (*unstructured text*). Data iklan lowongan kerja dalam penelitian ini dihimpun melalui portal Jobstreet, yang dipilih karena reputasinya sebagai salah satu platform rekrutmen terbesar dengan volume iklan talenta digital yang signifikan di Indonesia. Penggunaan Jobstreet memungkinkan perolehan data teks kualifikasi yang

representatif, mencakup berbagai level jabatan dari berbagai industri, sehingga menyediakan basis data yang kaya untuk diolah dan dipetakan ke dalam unit kompetensi SKKNI bidang Informasi dan Komunikasi. Proses ekstraksi data yang dilakukan melalui pendekatan konvensional terhadap ribuan iklan kerja tidak hanya memerlukan durasi waktu yang signifikan, tetapi juga memiliki tingkat kerentanan yang tinggi terhadap faktor kesalahan manusia (*human error*) serta bias subjektivitas.

Kompleksitas permasalahan meningkat seiring adanya inkonsistensi terminologi dan penggunaan bahasa yang beragam (*bilingual*). Satu entitas kompetensi yang sama sering kali direpresentasikan melalui variasi diksi yang berbeda. Adanya ambiguitas terminologi dalam kualifikasi pekerjaan ini memerlukan standar kompetensi nasional sebagai basis pelabelan yang objektif (Geraldy et al., 2025). Oleh karena itu, penelitian ini menggunakan Standar Kompetensi Kerja Nasional Indonesia (SKKNI) Bidang Informasi dan Komunikasi sebagai acuan baku. Penggunaan standar ini krusial untuk menjamin validitas pemisahan antara kualifikasi teknis (*hard skill*) dan kemampuan perilaku (*soft skill*) agar selaras dengan parameter industri yang berlaku di Indonesia.

Sebagai solusi atas permasalahan tersebut, diperlukan sebuah kerangka kerja yang mampu mengombinasikan kepastian aturan dengan fleksibilitas komputasi bahasa. Penelitian ini mengusulkan implementasi metode berbasis *Rule-Based* untuk menjamin ketepatan pemetaan pada istilah teknis yang sudah baku dalam SKKNI, yang kemudian diintegrasikan dengan teknik *Similarity Matching*. Pemilihan metode *Similarity Matching* dengan menggunakan algoritma *Cosine Similarity* terbukti efektif dalam memberikan penilaian objektif dengan menghitung derajat kemiripan antara istilah hasil ekstraksi dengan dataset referensi yang bersumber dari Unit-Unit Kompetensi pada SKKNI. Hal ini memungkinkan sistem untuk memetakan istilah yang secara tekstual berbeda ke dalam kelas kompetensi yang sama meskipun istilah tersebut tidak tercantum secara harfiah dalam dokumen SKKNI.

Meskipun terdapat beberapa studi yang telah mengeksplorasi klasifikasi teks melalui pendekatan *Machine Learning* konvensional, riset

yang menggabungkan keunggulan pengetahuan pakar pada sistem berbasis aturan (*rule-based*) dengan fleksibilitas pencocokan kemiripan (*similarity matching*) untuk identifikasi talenta digital masih sangat terbatas. Studi ini diproyeksikan mampu memberikan sumbangsih ilmiah di bidang *Text Mining* sekaligus menghadirkan solusi teknologi dalam mengotomatisasi pemetaan kualifikasi pelamar guna mewujudkan proses seleksi yang lebih transparan, objektif, dan efisien (Malik & Malang, 2025).

Keunggulan dari penelitian ini terletak pada penggunaan standar SKKNI sebagai parameter klasifikasi yang objektif. Berbeda dengan pendekatan pada studi terdahulu yang cenderung melakukan pelabelan kualifikasi secara umum atau subjektif, penelitian ini menawarkan diferensiasi yang lebih presisi melalui pemisahan antara *hard skill* yang merujuk pada Unit Kompetensi Inti, serta *soft skill* yang didasarkan pada Unit Kompetensi Umum dan Sikap Kerja. Dengan demikian, hasil pemetaan yang dihasilkan memiliki validitas tinggi dan relevansi yang diakui oleh standar industri nasional.

Berdasarkan Uraian tersebut, Penulis tertarik melakukan penelitian dengan judul: **"Klasifikasi *Hard Skill* Dan *Soft Skill* IT Dari Iklan Kerja Menggunakan Metode Berbasis *Rule-Based* Dan *Similarity Matching*"**

## 1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana efektivitas metode berbasis *Rule-Based* dalam mengekstraksi informasi kualifikasi IT dari data iklan kerja Jobstreet yang memiliki format teks tidak terstruktur agar sesuai dengan terminologi baku SKKNI?
2. Bagaimana performa algoritma *Cosine Similarity* dalam mengklasifikasikan variansi istilah *hard skill* dan *soft skill* IT yang menggunakan kosakata multibahasa (Indonesia-Inggris) dengan mengacu pada Unit Kompetensi dalam SKKNI?
3. Sejauh mana tingkat akurasi dan performa sistem yang mengintegrasikan *metode Rule-Based* dan *Similarity Matching* dalam memetakan kualifikasi IT ke dalam kategori *Hard Skill* dan *Soft Skill* berdasarkan metrik Confusion Matrix?

## 1.3 Ruang Lingkup Masalah

Adapun ruang lingkup masalah yang akan di bahas adalah sebagai berikut:

1. Data penelitian bersumber dari iklan lowongan kerja bidang Teknologi Informasi (IT) pada platform Jobstreet Indonesia yang memiliki karakteristik teks tidak terstruktur.
2. Klasifikasi *hard skill* dan *soft skill* IT dilakukan dengan merujuk pada unit kompetensi dalam Standar Kompetensi Kerja Nasional Indonesia (SKKNI) sebagai dataset referensi (parameter objektif).
3. Implementasi metode terbatas pada penggunaan *Rule-Based* untuk ekstraksi kualifikasi dan algoritma *Cosine Similarity* untuk menghitung kemiripan istilah dalam bahasa Indonesia maupun Inggris (multibahasa).
4. Pengujian akurasi sistem dilakukan dengan menggunakan Confusion Matrix untuk mengukur nilai *Accuracy, Precision, Recall,* dan *F1-Score*.

## 1.4 Tujuan Penelitian

Adapun Tujuan penelitian sebagai berikut:

1. Mengimplementasikan metode *Rule-Based* untuk mengekstraksi data kualifikasi IT dari iklan kerja yang tidak terstruktur agar dapat dipetakan secara akurat ke dalam terminologi baku.
2. Menerapkan algoritma *Cosine Similarity* dalam mengukur derajat kemiripan antara variasi istilah kemampuan (*skill*) di industri dengan standar unit kompetensi yang ditetapkan dalam SKKNI.
3. Mengevaluasi tingkat akurasi sistem dalam mengklasifikasikan *hard skill* dan *soft skill* IT menggunakan *Confusion Matrix* untuk memastikan hasil pemetaan yang objektif dan valid.

## 1.5 Manfaat Penelitian

Adapun Manfaat penelitian ini adalah sebagai berikut:

1. Bagi Institusi

   Penelitian ini dapat menjadi bahan referensi dalam literatur akademik, khususnya mengenai implementasi teknik *Similarity Matching* melalui algoritma *Cosine Similarity* untuk klasifikasi teks. Kontribusi ini diharapkan dapat memperkuat kajian Program Studi Sistem Informasi dalam domain *Text Mining*, terutama yang berorientasi pada analisis kebutuhan tenaga kerja berdasarkan data rill dari portal lowongan pekerjan di sektor industri teknologi.
2. Bagi Peneliti

   Penelitian ini memberikan pengalaman praktis dalam mengimplementasikan alur kerja ekstraksi data dan klasifikasi teks, mulai dari tahap penarikan informasi menggunakan metode *Rule-Based* hingga tahap pengategorian otomatis. Selain itu, hasil penelitian ini dapat menjadi referensi bagi peneliti selanjutnya dalam mengevaluasi efektivitas algoritma kemiripan teks (*similarity*) pada dataset yang menggunakan terminologi multibahasa (Indonesia-Inggris).
3. Bagi Masyarakat dan Pencari Kerja

   Hasil penelitian ini berpotensi memberikan pemahaman yang lebih terukur mengenai kriteria *hard skill* dan *soft skill* yang dominan dibutuhkan di industri IT. Bagi pihak manajemen rekrutmen,

penelitian ini menawarkan solusi teknologi untuk mengotomatisasi pemetaan kualifikasi kandidat, sehingga proses seleksi dapat berjalan lebih objektif, cepat, dan efisien.