

BAB III

METODE PENELITIAN

3.1 Metode Penelitian

Penggunaan mixed method research dengan desain sequential explanatory didasarkan pada dua tahapan yang dilakukan, yaitu tahap kuantitatif awal dengan mengumpulkan data terstruktur untuk mengidentifikasi pola atau hubungan, kemudian tahap kualitatif menggunakan analisis tematik untuk mengeksplorasi pola secara mendalam (Caesar, 2024; Wipulanusat et al., 2020). Metode penelitian campuran memberikan pemahaman yang lebih komprehensif dibandingkan dengan pendekatan Tunggal, karena dapat mengatasi keterbatasan masing-masing metode dengan memberikan validasi silang terhadap temuan penelitian (Subedi, 2023).

Dalam konteks penelitian ini, pendekatan kuantitatif digunakan untuk menganalisis pola dan frekuensi kompetensi yang dibutuhkan industri melalui text mining dan analisis statistic, sementara pendekatan kualitatif digunakan untuk menginterpretasi dan memahami makna mendalam dari kebutuhan kompetensi tersebut melalui content analysis. Selanjutnya penelitian ini mengadopsi paradigma computational linguistics dan natural language processing yang bertujuan untuk merepresentasikan aturan linguistic secara komputasional untuk memfasilitasi tugas-tugas pemrosesan bahasa alami (Sadiku et al., 2024; Yomantas, 2022). Penggunaan teknik TF-IDF (Term Frequency-Inverse Document Frequency) dan cosine similarity karena umum digunakan dalam sistem pencarian untuk meningkatkan akurasi dan relevansi hasil pencarian (Rasyid & Ningsih, 2024), dimanfaatkan juga untuk mengidentifikasi kemiripan antar dokumen (Halim & Lasut, 2024)

3.2 Jenis penelitian

Berdasarkan tujuannya, penelitian ini termasuk dalam kategori penelitian deskriptif-eksploratif dengan pendekatan analisis konten. Penelitian deskriptif dapat berfokus pada perangkuman berbagai kondisi atau fenomena di masyarakat, sering kali menggunakan data naratif untuk menggali karakteristik dan pola (Bancin & Corry, 2023; Furidha, 2024). Selain itu juga dapat melibatkan analisis statistic untuk menggambarkan fenomena secara numerik, meskipun tidak bertujuan untuk menetapkan hubungan sebab-akibat (Mao & Huo, 2023).

Aspek eksploratif dari penelitian ini terletak pada upaya mengidentifikasi dan menganalisis pola-pola kompetensi yang belum terdokumentasi secara sistematis, serta mengeksplorasi kesenjangan (gap) antara kebutuhan industri dengan standar kompetensi nasional. Penelitian eksploratif cukup penting untuk studio tahap awal penelitian karena dapat mengidentifikasi pola dan mengusulkan kerangka teori baru jika diperlukan (Swedberg, 2020). Dari segi pendekatan waktu, penelitian ini merupakan penelitian cross-sectional yang mengambil data pada satu titik waktu tertentu untuk memberikan gambaran kondisi saat ini (Ziauddin et al., 2023) dengan hanya mengamati dan mencatat informasi sebagaimana adanya (Legiran, 2022), dalam hal ini mengenai kebutuhan kompetensi BIM di industri.

Penelitian ini juga menggunakan pendekatan analisis sekunder terhadap data lowongan pekerjaan yang tersedia di platform digital, yang kemudian dianalisis menggunakan kerangka teoritis O*NET (Occupational Information Network) untuk klasifikasi kompetensi.

3.3 Teknik Pengumpulan Data

Data yang digunakan pada penelitian ini berupa data sekunder yang merupakan sekumpulan informasi atau iklan informasi pekerjaan yang didapatkan dari beberapa platform pencarian kerja (LinkedIn, Jobstreet, Glints

dan KitaLulus) untuk dapat dilakukan mapping kemampuan yang dibutuhkan pada jabatan kerja BIM. Penggunaan data sekunder dipilih karena dapat memberikan akses terhadap volume data yang besar dan representatif dari kebutuhan industri secara real time (Yeole, 2023). Pemilihan platform LinkedIn, Jobstreet, Glints dan KitaLulus tersebut didasarkan pada kriteria memiliki volume lowongan kerja yang tinggi, dan menyediakan deskripsi pekerjaan yang detail. Proses pengumpulan data dilakukan secara manual scrapping yaitu dengan mengumpulkan iklan informasi kerja dengan kata kunci pada saat pencarian yaitu, BIM, Building Information Modelling, BIM Hiring Indonesia, BIM Drafter, BIM Modeler, BIM Coordinator, BIM Engineer, BIM Indonesia, BIM Specialist, BIM Technician, Drafter, Revit dan Tekla.

Data lainnya yang digunakan dalam penelitian ini adalah Standar Kompetensi Kerja Nasional Indonesia (SKKNI) sektor konstruksi, khususnya yang berkaitan dengan BIM, Kerangka Kualifikasi Nasional Indonesia (KKNI) untuk pemetaan level kompetensi dan framework O*NET untuk klasifikasi kompetensi dan aktivitas kerja.

3.4 Populasi dan Sampel

Penelitian ini menggunakan informasi yang terbuka secara luas dari berbagai *platform* pencarian kerja, dengan mengumpulkan data dari LinkedIn, Jobstreet, Glints dan KitaLulus. Data dikumpulkan dalam rentang waktu tanggal 27 oktober 2023 – 27 November 2023. Kemudian rumus sampel yang dipilih adalah total sampling, sehingga jumlah keseluruhan yang ada pada kedua populasi merupakan sampel pada penelitian ini. Pada tabel 3.1 merupakan contoh iklan informasi kerja yang didapatkan dari *platform* pencarian kerja secara online.

Tabel 3. 1 Contoh iklan informasi kerja hasil manual scrapping

						Title
16/7/2024	LinkedIn	BIM	GeoBIM Indonesia	BIM Modeler MEP PT Geo BIM Indonesia adalah salah satu anak perusahaan dari PT Madhava Persada Group yang menawarkan jasa BIM untuk konstruksi. Saat ini GeoBIM sedang membuka lowongan pekerjaan freelance BIM Modeler MEP untuk pengerjaan Project 3D Modelling MRT Jakarta. Kualifikasi 1. Berpengalaman atau terbiasa melakukan modeling menggunakan Autodesk Revit 2. Diutamakan yang pernah atau berpengalaman menggunakan Dynamo pada Revit 3. Diutamakan yang pernah atau berpengalaman mengerjakan MEP 4. Lulusan jurusan teknik sipil arsitektur atau sejenisnya 5. Bersedia bekerja dengan sistem WFO (10jakarta) yang akan dikontrak sesuai durasi pelaksanaan proyek (bisa diperpanjang apabila personel memiliki kinerja yang bagus) 6. Tidak sedang terikat kontrak kerja dengan instansi manapun 7. Diutamakan tidak sedang berkuliah atau memiliki rencana lanjut kuliah dalam waktu dekat Jobdesk Membuat model As Build Drawing MEP MRT Jakarta menggunakan Autodesk Revit sesuai KAK dan BEP	https://www.linkedin.com/jobs/view/3973901120	BIM Modeler MEP

3.5 Instrumen Penelitian

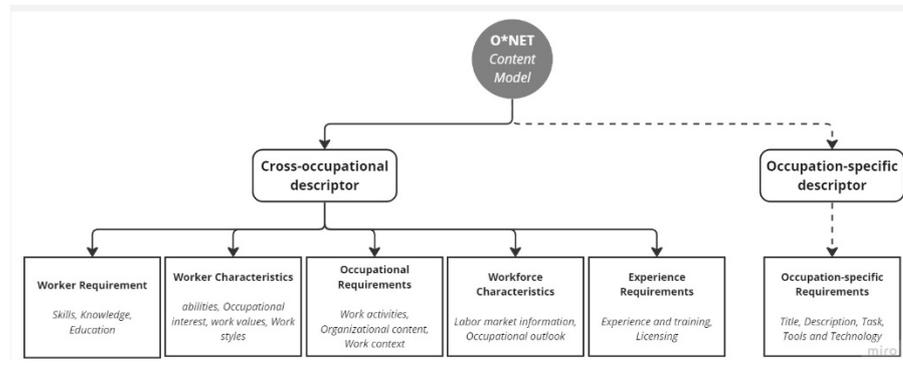
Instrumen penelitian dalam studi ini dirancang untuk memfasilitasi memfasilitasi pengumpulan dan analisis data secara sistematis menggunakan pendekatan *computational linguistic* dan *text mining*.

3.5.1 Instrumen Pengumpulan Data

Pengumpulan data dilakukan secara manual untuk memastikan akurasi data yang lebih tinggi, serta memungkinkan seleksi data yang lebih tepat sesuai dengan kriteria penelitian. Pengumpulan data manual melibatkan penyalinan dan penempelan data secara manual dari situs web, metode ini sering digunakan ketika menangani dataset kecil atau ketika otomatisasi tidak memungkinkan (Fikri et al., 2022).

Template pengumpulan data dirancang menggunakan spreadsheet berdasarkan klasifikasi O*NET content model yang merupakan sistem informasi pekerjaan dan kompetensi, klasifikasi pekerjaan dan kompetensi yang digunakan oleh pemerintah Amerika Serikat dan di dasarkan pada kerangka kerja knowledge (pengetahuan), skills (keterampilan) dan abilities (kemampuan) (KSA) (Uhm et al., 2017). Sejak tahun 1995, O*NET telah mengkategorikan 974 jabatan dari 1110 pekerjaan dengan menggunakan 277 elemen kompetensi dalam 6 domain kompetensi. O*NET content model ini terdiri atas 2 kategori yang di dalamnya terdapat 6 sub kategori,

kategori pertama disebut dengan kelompok “karakteristik” yang berisi kumpulan unsur kompetensi dan kategori kedua disebut sebagai persyaratan yang berisi kumpulan unsur kompetensi khusus yang berfokus pada pekerjaan.



Gambar 3. 1 ONET Content model

Kriteria seleksi data ditetapkan berdasarkan relevansi yang mensyaratkan bahwa iklan informasi kerja dapat secara eksplisit menyebutkan BIM (*Building Information Modeling*) atau software BIM secara spesifik (Autodesk Revit, Tekla Structures atau ArchiCAD). Kelengkapan data dalam konteks dataset pasar kerja, sangat penting untuk menganalisis tren dan membuat keputusan yang tepat (Yamashita et al., 2024). Ambang batas kelengkapan data minimum 80% (Margaritopoulos et al., 2012), di mana 8 dari 10 *core fields* (nama posisi, nama Perusahaan, *job description*, *requirement*, *skills*, level pengalaman, level pendidikan dan Lokasi) harus terisi dengan informasi substantif untuk memenuhi keberagaman dalam himpunan data penelitian. Preferensi bahasa dapat dalam bahasa Indonesia atau bahasa Inggris, asalkan lokasi penempatan kerjanya di Indonesia.

Alat pendukung untuk membantu proses *manual scraping* digunakan *Google Spreadsheet*, karena dapat memfasilitasi kolaborasi secara langsung, manajemen data dan pemantauan proyek sehingga sesuai digunakan sebagai alat dalam pengumpulan dan pemrosesan data.

3.6 Teknik Analisis Data

Analisis data dalam penelitian ini menggunakan pendekatan sequential explanatory mixed method yang mengintegrasikan Teknik analisis kuantitatif dan kualitatif secara bertahap (Creswell, J. W., & Plano Clark, 2017). Proses analisis data dibagi menjadi tiga tahap utama yaitu, preprocessing data, analisis kuantitatif menggunakan text mining, dan analisis kualitatif menggunakan content analysis.

3.6.1 Pra Pemrosesan Data

Prapemrosesan penting untuk mengonversi data teks tidak terstruktur menjadi format terstruktur yang dapat dianalisis oleh model komputasional. Hal ini melibatkan pembersihan “noise” seperti stop words, tanda baca, dan istilah yang tidak relevan, yang tidak memberikan kontribusi signifikan terhadap konteks teks (Kalra & Aggarwal, 2018; Kathuria et al., 2021). Teknik prapemrosesan yang umum digunakan meliputi tokenization, penghapusan stop words, stemming dan lemmatization. Metode-metode ini membantu mengurangi dimensi data dan meningkatkan efisiensi algoritma text mining (da Silva et al., 2023; Nayak & Kanive, 2016).

Pembersihan dan normalisasi data deskripsi pekerjaan dilakukan dengan menghilangkan elemen-elemen yang tidak relevan seperti karakter khusus dan noise lainnya. Proses normalisasi dilakukan untuk menyeragamkan format penulisan, termasuk konversi huruf kapital menjadi huruf kecil. Selanjutnya tokenisasi dilakukan untuk memecah teks menjadi unit-unit kata atau frasa yang dapat dianalisis secara individual. Penghapusan stop words atau kata-kata umum yang tidak memberikan makna signifikan terhadap teks dan sering dihapus untuk meningkatkan efisiensi komputasi serta memfokuskan analisis pada kata-kata yang lebih informatif.

Stemming dan lemmatization diterapkan untuk mereduksi kata-kata ke bentuk dasarnya. Stemming mengubah kata menjadi bentuk akar dengan

menghapus imbuhan, proses ini biasanya spesifik terhadap bahasa dan memerlukan seperangkat aturan atau algoritma untuk menangani berbagai struktur morfologis (Amri & Zenkour, 2018; Trishala & Mamatha, 2021). Lemmatization melibatkan penghapusan akhiran infleksi untuk mengembalikan kata ke bentuk dasar atau bentuk kamusnya (lemma), proses ini memerlukan pengetahuan semantic dan sering mengandalkan kamus digital atau informasi morfologi bahasa (Abidin et al., 2024; Amri & Zenkour, 2018; Trishala & Mamatha, 2021).

Prapemrosesan data dalam penelitian ini menggunakan bantuan alat yaitu berupa extension dari google chrome yaitu Apps Script, digunakan untuk mengintegrasikan Google Spreadsheet dengan layanan google lainnya, meningkatkan fungsionalitas, dan mengotomisasi proses (S & Dirgahayu, 2024). Google apps script mengotomatiskan tugas-tugas manajemen data di dalam Google spreadsheet, seperti entri data, pembaruan dan penghitungan, sehingga mengurangi kesalahan manual dan menghemat waktu.

3.6.2 Analisis Kuantitatif

TF-IDF atau *Term Frequency–Inverse Document Frequency* digunakan untuk mengukur kemiripan teks dengan mengevaluasi pentingnya suatu istilah dalam dokumen relative terhadap kumpulan dokumen (Lan, 2022). Metode ini menghitung bobot istilah berdasarkan frekuensinya dalam teks tertentu dan kelangkaannya di seluruh kumpulan dokumen. *Term frequency* mengacu pada semakin sering sebuah kata muncul dalam sebuah teks, semakin relevan kata tersebut dengan topik teks, *Inverse Document Frequency* mengacu pada semakin sering sebuah istilah muncul dalam beberapa teks dalam koleksi teks, semakin buruk istilah tersebut (Zhang & Ge, 2019). Dengan menggunakan metode tersebut berikut adalah rumus yang dapat digunakan dalam perhitungannya :

$$TF - IDF (w_i) = tf(w_i) \times idf(w_i) = tf(w_i) \times \log\left(\frac{N}{df(w_i)}\right),$$

$$tf(w_i) = \frac{n_{ij}}{\sum_{k=1}^m n_{kj}}$$

$$df(w_i) = |\{j: w_i \in d_j\}|$$

Data yang akan melewati proses TF-IDF adalah jabatan kerja beserta elemen kompetensi yang terdapat pada SKKNI serta iklan informasi kerja dan deskripsi kerja yang sudah dikumpulkan. Data-data tersebut kemudian diukur tingkat kemiripannya menggunakan sebuah algoritma yang disebut dengan *cosine similarity*. Pada prinsipnya *cosine similarity* mengukur kemiripan dalam ruang vektor (*vector space similarity measure*), menggunakan kata kunci dalam sebuah dokumen untuk menghitung kemiripan antara dokumen-dokumen tersebut, dan dinyatakan dalam bentuk vektor (Aohana & Fitri Bimantoro, 2023). Berikut adalah rumus yang dapat digunakan untuk menghitung *cosine similarity* :

$$\text{Cos } a = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Untuk mengolah data menggunakan TF-IDF dan *cosine similarity* menggunakan bantuan alat yang sama yaitu google apps script.

Automated gap analysis sebagai tambahan dari analisis TF-IDF, gap analysis dilakukan menggunakan *google apps script* yang terintegrasi dengan proses *text mining* sebelumnya. *Gap analysis* membandingkan kompetensi yang teridentifikasi dari analisis iklan informasi kerja dengan standar kompetensi yang tercantum dalam SKKNI terkait BIM dan KKNi (Gupta et al., 2007). Proses *gap analysis* menggunakan pendekatan *systematic comparison* melalui *matrix mapping* untuk mengidentifikasi tiga kategori utama : (1) kompetensi yang sudah tercakup dalam standar nasional; (2) kompetensi yang kurang memadai atau memerlukan updating; dan (3) kompetensi baru yang belum terakomodasi dalam standar nasional

3.6.3 Analisis Kualitatif

Untuk analisis kualitatif dalam penelitian ini menggunakan content analysis dengan pendekatan deduktif dengan kerangka teoritis O*NET content model sebagai skema koding awal. Pendekatan deduktif dipilih

karena penelitian ini bertujuan untuk memetakan kompetensi yang teridentifikasi dari job market ke dalam framework klasifikasi yang sudah ada. Pendekatan ini memfasilitasi pengembangan kerangka kompetensi standar yang data diterapkan di berbagai sektor. Standarisasi ini membantu menciptakan bahasa bersama untuk kompetensi, sehingga memudahkan organisasi menilai dan mengembangkan keterampilan karyawan (Bahri & Salsiati, 2024; Freedden & Meeks, 2022).

Proses content analysis meliputi tiga tahapan utama, yaitu preparation phase dengan penentuan unit analisis berupa kalimat atau frase yang mengandung deksripsi kompetensi. Selanjutnya tahap organizing phase dengan kategorisasi kompetensi berdasarkan 6 domain O*NET (Worker Requirements, Worker Characteristics, Experience Requirements, Occupational Requirements, Workforce Characteristics, dan Occupation-Specific Information) dan reporting phase dengan interpretasi dan validasi hasil kategorisasi (DeVellis, 2017).

Proses pengkodean dilakukan secara sistematis untuk mengelompokkan dan menmentranformasikan raw data menjadi format terstruktur yang dapat dianalisis menggunakan framework O*NET. Prosedur pengkodean meliputi empat Langkah utama :

1. Identifikasi kategori dan kode. Pengkategorian disesuaikan dengan elemen yang tersedia pada O*NET content model, dengan menggunakan 277 elemen kompetensi dalam 6 domain kompetensi sebagai coding scheme.
2. Atribusi data kompetensi. Data iklan informasi rekrutmen kerja yang terdiri atas nama posisi dan deskripsi pekerjaan dianalisis satu persatu untuk mengidentifikasi requirement yang dapat dikategorisasikan ke dalam elemen kategori kompetensi O*NET
3. Organisasi data dalam format terstruktur. Data yang telah di kodekan di organisasikan dalam google spreadsheet dengan format yang memfasilitasi analisis lebih lanjut dan memungkinkan tracking dari raw data ke kategorisasi akhir

Penggunaan established framework ini didasarkan pada prinsip theoretical validation Dimana instrument atau kerangka kerja yang telah terbukti reliable dan valid dalam konteks serupa dapat diadopsi untuk penelitian baru tanpa perlu re-validation dari awal, selama konteks penelitian masih dalam domain yang sama (DeVellis, 2017). Hal ini didukung oleh fakta bahwa O*NET telah mengkategorikan lebih dari 974 posisi dengan menggunakan 277 elemen kompetensi yang dikembangkan melalui analisis pekerjaan yang luas dan validasi ahli. Kategorisasi ini membantu mengidentifikasi soft skills dan kemampuan penting yang krusial untuk keberhasilan di bidang tertentu (Fantozzi et al., 2024)

