

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Gambaran Umum Objek Penelitian

Penelitian dilakukan untuk menjawab berbagai permasalahan yang dihadapi oleh perusahaan Digital Hub Sinar Mas Land dengan memanfaatkan penggunaan teknologi AI & ML untuk meningkatkan operasional serta kualitas layanan pelanggan melalui rancang bangun dan optimasi model *chatbot* Digital Hub. Digital Hub Sinar Mas Land membutuhkan *chatbot* yang mampu merespons pertanyaan dengan cepat, akurat, memahami konteks percakapan, serta memberikan jawaban yang relevan untuk meningkatkan efisiensi komunikasi antara perusahaan dan pelanggan. Objek penelitian ini dipilih karena Digital Hub Sinar Mas Land sebagai pusat inovasi digital memiliki kebutuhan yang mendesak untuk memberikan layanan pelanggan yang cepat, efektif, dan responsif terhadap berbagai pertanyaan dan permasalahan yang muncul.

##### 3.1.1 Sejarah Perusahaan

Sinar Mas Land, sebagai salah satu pengembang properti terbesar di Indonesia, telah berkontribusi secara signifikan dalam sektor properti nasional selama lebih dari 40 tahun. Didirikan oleh Eka Tjipta Widjaja, perusahaan ini memiliki beragam proyek, yang mencakup pengembangan perumahan hingga kawasan perkotaan terintegrasi, seperti BSD City.[18]. [18].



Gambar 3.1 Logo Sinar Mas Land

Sinar Mas Land, dengan visi menjadi pemimpin dalam industri properti, merancang proyek Digital Hub sebagai ekosistem bisnis yang inovatif dan strategis di Indonesia. Proyek ini didasarkan pada enam pilar utama yang terintegrasi, menjadikannya pusat bagi perusahaan teknologi dan *startup*.

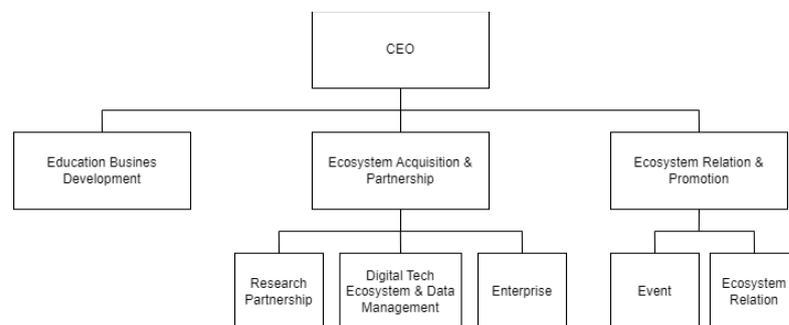
Digital Hub menggabungkan elemen real estat komersial dengan fasilitas pendidikan dan rekreasi, seperti Biomedical Campus dan Urban Forest.[61].



Gambar 3.2 Logo Digital Hub

### 3.1.2 Struktur Organisasi Perusahaan

Struktur organisasi Sinar Mas Land dirancang untuk mendukung pencapaian tujuan bisnis, dengan pembagian tugas yang jelas dan hierarki efektif yang memungkinkan koordinasi optimal antar unit kerja. Struktur ini mendukung efisiensi dan pencapaian target, terutama dalam proyek Digital Hub.



Gambar 3.3 Struktur Perusahaan Digital Hub by Sinar Mas Land

Struktur organisasi Sinar Mas Land mencerminkan komitmennya untuk menjadi pemimpin dalam industri properti di Indonesia. Pembagian tugas yang terstruktur dalam berbagai divisi, seperti pengembangan pendidikan, akuisisi ekosistem, dan pengelolaan data, mengindikasikan upaya perusahaan dalam menciptakan nilai tambah yang berkelanjutan. Fokus yang diberikan pada ekosistem mengungkapkan pandangan perusahaan bahwa properti tidak hanya

sekadar sebuah entitas, tetapi merupakan bagian dari suatu sistem yang lebih komprehensif, yang melibatkan integrasi infrastruktur, fasilitas, dan pengembangan komunitas.

### 3.2 Metode Penelitian

#### 3.2.1 Metode Machine Learning

Penelitian ini menggunakan *framework* CRISP-ML sesuai dengan tujuan penelitian yang berorientasi pada optimasi *chatbot* yang dilakukan menggunakan model *machine learning* sehingga lebih akurat dan cocok menggunakan *framework* CRISP-ML dibanding dengan model lainnya. Metode CRISP-ML telah dijelaskan pada bab 2 tentang landasan teori. Dalam konteks ini, perlu dilakukan perbandingan antara kedua metode tersebut untuk memberikan alasan yang jelas mengenai pemilihan CRISP-ML sebagai metode yang digunakan dalam penelitian ini. Berikut adalah perbandingan dari kedua kerangka kerja tersebut, antara lain:

Tabel 3.1 Perbandingan CRISP-DM dan CRISP-ML

Aspek	CRISP-DM	CRISP-ML
<b>Kepanjangan</b>	<i>Cross-Industry Standard Process for Data Mining.</i>	<i>Cross-Industry Standard Process for Machine Learning.</i>
<b>Tujuan</b>	Digunakan sebagai standar proses dalam proyek data mining.	Dikembangkan untuk proyek machine learning dengan fokus pada jaminan kualitas.
<b>Tahapan</b>	<ol style="list-style-type: none"> <li>1. <i>Business Understanding</i></li> <li>2. <i>Data Understanding</i></li> <li>3. <i>Data Preparation</i></li> <li>4. <i>Modeling</i></li> <li>5. <i>Evaluation</i></li> <li>6. <i>Deployment</i></li> </ol>	<ol style="list-style-type: none"> <li>1. <i>Business and Data Understanding</i></li> <li>2. <i>Data Preparation</i></li> <li>3. <i>Modeling</i></li> <li>4. <i>Evaluation</i></li> <li>5. <i>Deployment</i></li> <li>6. <i>Monitoring and Maintenance</i></li> </ol>

<b>Aspek</b>	<b>CRISP-DM</b>	<b>CRISP-ML</b>
<b>Dokumentasi &amp; Transparansi</b>	Dokumentasi proses tidak menjadi fokus utama.	Menekankan dokumentasi lengkap pada setiap langkah untuk menjaga transparansi dan kualitas.
<b>Penggunaan Pada Industri</b>	Digunakan secara luas dalam proyek <i>data mining</i> .	Cocok untuk proyek ML di industri yang membutuhkan <i>Quality Assurance (QA)</i> .
<b>Kelebihan</b>	<ul style="list-style-type: none"> <li>a. Struktur yang jelas dan terdefinisi, memandu proses <i>data mining</i> dari awal hingga akhir.</li> <li>b. Fleksibel dan dapat diterapkan di berbagai industri.</li> <li>c. Menekankan pemahaman bisnis sebelum teknis.</li> </ul>	<ul style="list-style-type: none"> <li>a. Menambahkan aspek jaminan kualitas (QA) dalam proses <i>machine learning</i>.</li> <li>b. Memperhatikan siklus hidup model secara lengkap, termasuk pemantauan dan pemeliharaan.</li> <li>c. Cocok untuk proyek <i>machine learning</i>.</li> </ul>
<b>Kekurangan</b>	<ul style="list-style-type: none"> <li>a. Kurang spesifik untuk proyek <i>machine learning</i> modern.</li> <li>b. Tidak menekankan pemantauan dan pemeliharaan model setelah implementasi.</li> </ul>	<ul style="list-style-type: none"> <li>a. Relatif baru dan kurang dikenal dibandingkan CRISP-DM.</li> <li>b. Implementasi yang lebih kompleks karena penambahan tahapan jaminan kualitas.</li> <li>c. Membutuhkan sumber daya tambahan untuk pemantauan model.</li> </ul>

Berdasarkan perbandingan framework CRISP-DM dan CRISP-ML, pemilihan framework CRISP-ML dalam penelitian ini didasarkan pada beberapa pertimbangan yang relevan dengan kebutuhan *proyek machine learning* seperti yang telah dipaparkan diatas. Pertama, CRISP-ML menekankan pentingnya dokumentasi yang komprehensif pada setiap langkah proses untuk menjaga transparansi dan kualitas, yang menjadi faktor krusial dalam proyek *machine learning* yang memerlukan pemantauan dan pembaruan model secara berkelanjutan khususnya pada model chatbot yang terus diperbarui jika terdapat informasi baru. Selain itu, CRISP-ML dirancang untuk memenuhi kebutuhan industri yang memerlukan jaminan kualitas dalam implementasi *machine learning*, sehingga kerangka kerja ini lebih relevan bagi aplikasi industri yang mengutamakan keandalan dan konsistensi model. CRISP-ML juga mengintegrasikan siklus hidup model yang lebih lengkap, termasuk tahap pemantauan dan pemeliharaan, yang penting untuk memastikan performa model tetap optimal setelah diimplementasikan. Oleh karena itu, CRISP-ML dianggap lebih sesuai untuk penelitian ini dibandingkan dengan CRISP-DM, yang cenderung bersifat umum dan lebih relevan untuk proses *data mining*.

### 3.2.2 Metode Vektorisasi Teks

Dalam melakukan analisis teks, representasi data dalam bentuk numerik merupakan langkah krusial untuk memungkinkan penerapan berbagai algoritma pembelajaran mesin. Selain *Term Frequency-Inverse Document Frequency (TF-IDF)* terdapat berbagai metode vektorisasi teks lainnya yang telah dianalisis secara mendalam dalam literatur. Contohnya: metode *Bag of Words (BoW)* dan *Word2Vec*. Berikut adalah perbandingannya:

Tabel 3.2 Perbandingan Vektorisasi Teks

Keterangan	TF-IDF	BoW	Word2Vec
Definisi	Model yang merepresentasika	Teknik untuk menghitung bobot	Model pembelajaran representasi kata

<b>Keterangan</b>	<b>TF-IDF</b>	<b>BoW</b>	<b>Word2Vec</b>
	n teks sebagai vektor berdasarkan jumlah kemunculan kata dalam dokumen.	pentingnya kata dalam dokumen dengan mempertimbangkan frekuensi kata dan pentingnya kata dalam seluruh korpus.	berbasis pembelajaran mendalam, menghasilkan vektor kata yang mencerminkan hubungan semantik antar kata.
<b>Model Representasi</b>	Vektor frekuensi kata.	Vektor dengan bobot berdasarkan frekuensi kata dan IDF.	Vektor kata berdimensi rendah, dihasilkan melalui pembelajaran dari konteks kata dalam korpus.
<b>Konteks</b>	Tidak mempertimbangkan konteks kata.	Tidak mempertimbangkan konteks kata, hanya memperhitungkan frekuensi dalam dokumen dan korpus.	Mempertimbangkan konteks kata dalam kalimat atau dokumen dan hubungan semantiknya.
<b>Ukuran Vektor</b>	Ukuran vektor tergantung pada jumlah kata unik dalam korpus.	Ukuran vektor tergantung pada jumlah kata unik dalam korpus (seperti BoW).	Ukuran vektor tetap meskipun ada ribuan kata dalam korpus (dimensi vektor bisa 50, 100, atau 300, dll.).
<b>Ketergantungan pada Urutan Kata</b>	Tidak mempertimbangkan urutan kata.	Tidak mempertimbangkan urutan kata.	Menangkap konteks dua arah dalam kalimat, menangani urutan kata.
<b>Sifat Kata</b>	Kata dianggap independen, tanpa	Kata dianggap independen, tetapi	Kata diwakili sebagai vektor yang dapat

<b>Keterangan</b>	<b>TF-IDF</b>	<b>BoW</b>	<b>Word2Vec</b>
	memperhatikan hubungan antar kata.	ada penyesuaian bobot berdasarkan keberadaannya di seluruh korpus.	menangkap hubungan semantik dan sintaksis antar kata.
<b>Penggunaan</b>	Berguna untuk analisis teks dasar seperti klasifikasi atau clustering.	Digunakan dalam pencarian informasi, ekstraksi fitur, dan analisis teks yang membutuhkan bobot kata.	Digunakan untuk tugas-tugas yang membutuhkan pemahaman semantik lebih mendalam, seperti analisis sentimen, klasifikasi teks, dan penerjemahan otomatis.
<b>Kelebihan</b>	Sederhana, mudah diimplementasikan, dan efektif untuk tugas dasar.	Memberikan bobot kata berdasarkan pentingnya dalam korpus, meningkatkan relevansi kata dalam dokumen.	Menangkap hubungan semantik antar kata, sehingga dapat lebih baik dalam menangani sinonim dan perbedaan makna kata.
<b>Kekurangan</b>	Tidak menangkap konteks dan hubungan antar kata.	Tidak mempertimbangkan hubungan semantik kata, hanya menghitung frekuensi dan IDF.	Memerlukan korpus besar dan sumber daya komputasi yang tinggi untuk pelatihan dan penggunaan.

Dalam pembuatan *chatbot* khususnya informasi spesifik sangat efektif menggunakan TF-IDF karena dapat memfokuskan pada kata-kata yang relevan dan penting dalam konteks tertentu. TF-IDF memberikan bobot lebih pada kata-kata yang sering muncul dalam dokumen tertentu tetapi jarang ditemukan

di seluruh korpus, yang sering kali adalah kata kunci yang terkait langsung dengan produk atau layanan. Hal ini memungkinkan *chatbot* untuk memberikan jawaban yang lebih tepat dan relevan dengan mengenali kata-kata penting yang membedakan pertanyaan pelanggan. Selain itu, TF-IDF juga mengurangi pengaruh kata-kata umum atau *stopwords* yang tidak memberikan informasi berarti, sehingga *chatbot* dapat lebih fokus pada inti pertanyaan. Dengan cara ini, TF-IDF membantu *chatbot* dalam menyaring kata-kata tidak relevan dan memberikan respons yang lebih cepat dan tepat sasaran.

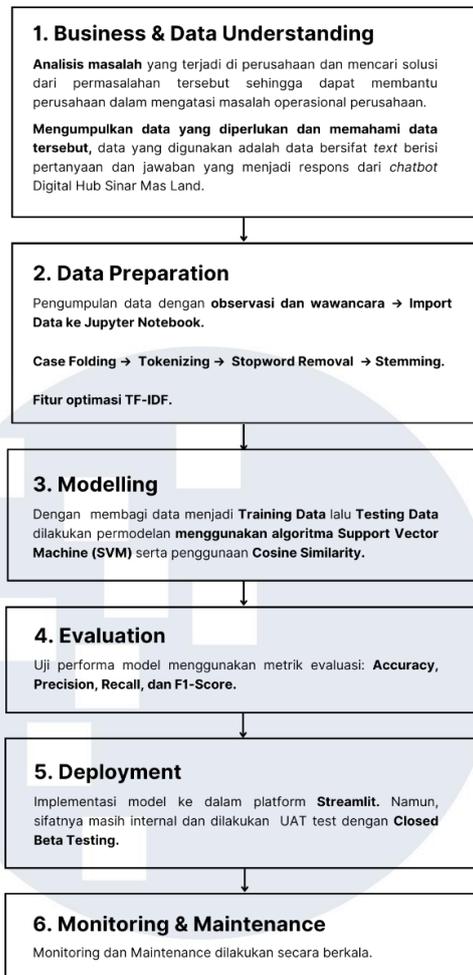
### 3.2.3 Alur Penelitian

Alur Penelitian dirancang untuk memberikan ilustrasi yang komprehensif mengenai tahapan proses penelitian yang akan dilaksanakan. Alur penelitian juga bertujuan agar penelitian dapat dilaksanakan dengan terstruktur menggunakan *framework* CRISP-ML. Berdasarkan pada gambar 3.1, penelitian diawali dengan *Business & Data Understanding* yaitu dengan menganalisis permasalahan perusahaan dan perkiraan solusi yang dapat dilakukan untuk mengatasi permasalahan perusahaan tersebut. Pada Digital Hub Sinar Mas Land, teridentifikasi banyaknya pertanyaan yang masuk tanpa didukung oleh sistem respons otomatis yang memadai. Pada tahap ini juga dilakukan pengumpulan data dengan cara observasi dan wawancara kepada petinggi perusahaan Digital Hub Sinar Mas Land. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan serta mengoptimalkan *chatbot* guna meningkatkan kecepatan, efisiensi dalam merespons pertanyaan tersebut, dan membantu operasional perusahaan. Tahap kedua adalah melakukan *Data Preparation* yaitu melakukan beberapa tahapan dalam NLP karena data yang digunakan untuk memaksimalkan model sehingga mendapat akurasi yang maksimal. Langkah-langkah tersebut meliputi proses *case folding*, *tokenizing*, *stopword removal*, and *stemming*. Selanjutnya, data-data yang telah dilakukan *preprocessing* akan di optimasi menggunakan fitur TF-IDF agar model menjadi lebih akurat lagi.

Tahap berikutnya adalah tahap *Modelling*, yang dimulai dengan membagi dataset menjadi dua bagian: data training dan data testing. Pembagian ini penting untuk memastikan bahwa model dapat belajar dari data training dan kemudian diuji pada data testing untuk mengukur kinerjanya. Dalam proses ini, algoritma Support Vector Machine (SVM) digunakan sebagai model utama.. Tahapan *Evaluation* ini mencakup berbagai metrik, seperti akurasi, *presisi*, *recall*, dan nilai *F1-score*, guna memastikan bahwa model memiliki kemampuan yang baik dalam menggeneralisasi pola dari data yang ada.

Setelah model *chatbot* selesai dibangun, tahap berikutnya adalah *Deployment*, yaitu proses mengimplementasikan model ke dalam platform Streamlit. Dengan menggunakan Streamlit sebagai antarmuka, pihak-pihak yang berhubungan langsung, seperti pengguna internal dari perusahaan, dapat melakukan uji coba atau eksplorasi interaktif terhadap kinerja *chatbot* secara langsung. Proses ini memungkinkan tim untuk melihat bagaimana *chatbot* merespons berbagai pertanyaan dan skenario, serta mengidentifikasi area yang mungkin memerlukan penyempurnaan. Selama tahap ini, juga dilakukan *Closed Beta Testing*, di mana staff Digital Hub secara terbatas diberikan akses untuk mencoba *chatbot*. Uji coba ini bertujuan untuk mengevaluasi apakah model telah sesuai dengan kebutuhan dan ekspektasi perusahaan Digital Hub, serta mengidentifikasi potensi perbaikan berdasarkan masukan dari pengguna.

Tahapan akhir dalam pengembangan ini adalah *Monitoring & Maintenance*. Setelah implementasi selesai, model akan secara berkala dimonitor dan diperbarui untuk memastikan kinerjanya tetap optimal dan relevan. Staff Digital Hub Sinar Mas Land akan melakukan pengecekan rutin terhadap performa *chatbot* dan menambahkan modul atau dataset baru sesuai dengan perkembangan dan kebutuhan perusahaan. Hal ini bertujuan agar *chatbot* selalu *up-to-date* dan mampu merespons informasi terbaru, sehingga tetap memberikan pengalaman yang informatif dan akurat bagi pengguna.



Gambar 3.4 Alur Penelitian

### 3.3 Teknik Pengumpulan Data

#### 3.3.1 Metode Pengumpulan Data

Penelitian ini akan menggunakan metode kualitatif yang menggunakan data primer dan data sekunder. Data primer berupa data teks yang berasal dari hasil observasi dan wawancara dengan petinggi Digital Hub Sinar Mas Land. Wawancara ini dilakukan untuk memahami kebutuhan spesifik perusahaan terkait layanan *chatbot*, serta untuk memastikan bahwa jawaban yang diberikan oleh *chatbot* nantinya akan akurat dan berintegritas sesuai dengan visi dan misi perusahaan. Observasi yang dilakukan juga bertujuan untuk menganalisis permasalahan yang sering dihadapi oleh pengguna, sehingga *chatbot* yang

dikembangkan mampu memberikan solusi yang relevan dan responsif terhadap kebutuhan pengguna. Pemilihan dan pengambilan sejumlah data tersebut juga didasari pada ketentuan perusahaan yang membatasi penggunaan data. Selanjutnya, terdapat juga data sekunder dari beberapa studi literatur dan penelitian terdahulu yang relevan dengan penelitian untuk memperkuat analisis data. Berikut adalah contoh dari dataset untuk penelitian ini.

	Pertanyaan	Jawaban	Label
1	Apa itu Digital Hub Sinar Mas Land?	Digital Hub adalah kawasan bisnis teknologi di BSD City.	Informasi Umum
2	Di mana lokasi Digital Hub?	Digital Hub terletak di BSD City, Tangerang Selatan.	Informasi Umum
3	Apa saja fasilitas yang tersedia di Digital Hub?	Fasilitas di Digital Hub meliputi coworking space, ruang kantor, dan area rekreasi.	Informasi Umum
4	Bagaimana cara menyewa ruang kantor di Digital Hub?	Hubungi tim marketing atau kunjungi situs resmi untuk menyewa ruang.	Informasi Kontak dan Publikasi
5	Apa keunggulan Digital Hub untuk startup?	Digital Hub menawarkan lingkungan kolaboratif bagi startup dan perusahaan teknologi.	Informasi Umum
6	Apakah ada coworking space di Digital Hub?	Ya, terdapat coworking space dengan berbagai fasilitas.	Informasi Umum
7	Bagaimana keamanan di Digital Hub?	Kesamanan di Digital Hub dijaga 24 jam dengan CCTV dan akses terbatas.	Informasi Umum
8	Apakah Digital Hub terhubung dengan transportasi umum?	Digital Hub terhubung dengan MRT dan bus, memudahkan akses.	Informasi Umum
9	Apa saja perusahaan besar yang ada di Digital Hub?	Perusahaan seperti Apple dan Microsoft memiliki kantor di sini.	Startup, Event, dan Komunitas
10	Bagaimana cara menghubungi tim marketing Digital Hub?	Anda dapat menghubungi tim marketing melalui website Digital Hub.	Informasi Kontak dan Publikasi
11	Apakah ada layanan internet cepat di Digital Hub?	Ya, tersedia layanan internet fiber optik yang cepat.	Informasi Umum
12	Bagaimana konsep arsitektur Digital Hub?	Arsitektur Digital Hub didesain modern dengan fokus pada kolaborasi.	Informasi Umum
13	Apakah Digital Hub memiliki ruang konferensi?	Tersedia ruang konferensi untuk acara bisnis dan pertemuan.	Informasi Umum
14	Apakah Digital Hub memiliki ruang konferensi?	Tersedia ruang konferensi untuk acara bisnis dan pertemuan.	Informasi Umum
15	Apa peran Digital Hub dalam ekosistem startup?	Digital Hub membantu membangun jaringan bagi startup teknologi.	Informasi Umum

Gambar 3.5 Dataset chatbot Digital Hub

### 3.3.2 Narasumber

Narasumber utama dalam penelitian ini adalah Pak Aristianto, yang menjabat sebagai *Assistant Manager of Digital Tech & Ecosystem Data Management* di perusahaan Digital Hub Sinar Mas Land. Beliau memiliki tanggung jawab utama dalam menjalin komunikasi dengan para *stakeholder* terkait, khususnya dengan seluruh informasi dan pengembangan yang ada pada Digital Hub Sinar Mas Land.

### 3.3.3 Periode Pengambilan Data

Pengambilan data melalui wawancara dan observasi dilakukan selama dengan rentan waktu kurang lebih 11 hari yaitu pada tanggal 10 – 21 September 2024. Pertanyaan disusun berdasarkan intensitas pertanyaan dan jawaban yang sering ditanyakan oleh pihak diluar kepada pihak Digital Hub karena perusahaan lain ingin lebih memahami dan mengetahui lebih detail tentang

Digital Hub Sinar Mas Land. Peneliti juga melakukan pengambilan data melalui observasi secara langsung pada kantor, berkas, dan data perusahaan selama proses pengerjaan penelitian dari bulan Juni hingga Oktober 2024.



UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA