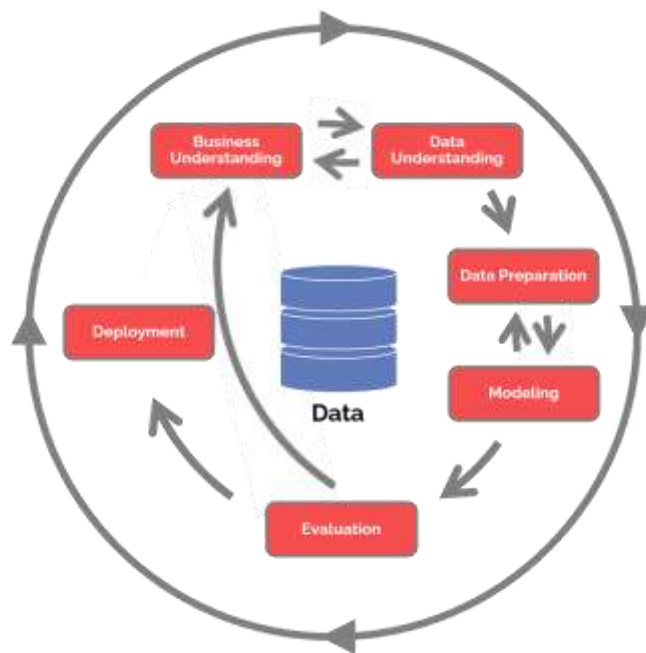


BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Tahapan penelitian adalah serangkaian tahapan yang sistematis terkait aktivitas ilmiah guna mencapai tujuan penelitian. Penelitian ini menggunakan CRISP-DM sebagai acuan tahapan penelitian yang dilakukan. CRISP-DM (*Cross-Industry Standard Process for Data Mining*) merupakan kerangka proses *data mining* yang ditujukan untuk memecahkan masalah bisnis dan penelitian. (Adhi Putra, 2021). Gambar 3.1 mendeskripsikan mengenai tahapan dari metode CRISP-DM.



Gambar 3.1 Tahapan CRISP-DM

(Sumber : Kotz, 2023)

Menurut (Adhi Putra, 2021) menuturkan bahwa pada CRISP-DM memiliki beberapa tahapan, yaitu:

3.1.1 *Business Understanding* (Pemahaman Bisnis)

Business Understanding dilakukan untuk mengidentifikasi permasalahan yang akan diangkat yaitu membandingkan performa algoritma *Support Vector Machine* (SVM) dengan menerapkan metode pelabelan *Textblob* dan *Vader* dalam melakukan analisis sentimen ulasan DANA dari Google Play Store.

3.1.2 *Data Understanding* (Pemahaman Data)

Data Understanding dilakukan untuk mengumpulkan data, mengidentifikasi data, dan memahami data. Data yang digunakan dalam penelitian adalah data ulasan aplikasi DANA pada Google Play Store. Data tersebut diambil menggunakan metode *scraping*.

3.1.3 *Data Preparation* (Persiapan Data)

Data Preparation dilakukan untuk mentransformasikan dan mempersiapkan data terlebih dahulu sebelum dilakukan pelabelan menggunakan *Textblob* dan *Vader*. Setelah data memiliki label, data akan diubah dalam bentuk vektor melalui perhitungan bobot kata menggunakan TF-IDF.

3.1.4 *Modelling* (Permodelan)

Modelling dilakukan terhadap data ulasan yang telah melewati *data preparation*. Data ulasan dibagi jadi data *training* dan data *testing* dengan rasio perbandingan 60:40, 70:30, dan 80:20. Selanjutnya data diklasifikasikan menggunakan algoritma *Support Vector Machine* (SVM) guna menciptakan nilai sentimen positif, negatif atau netral.

3.1.5 *Evaluation* (Evaluasi)

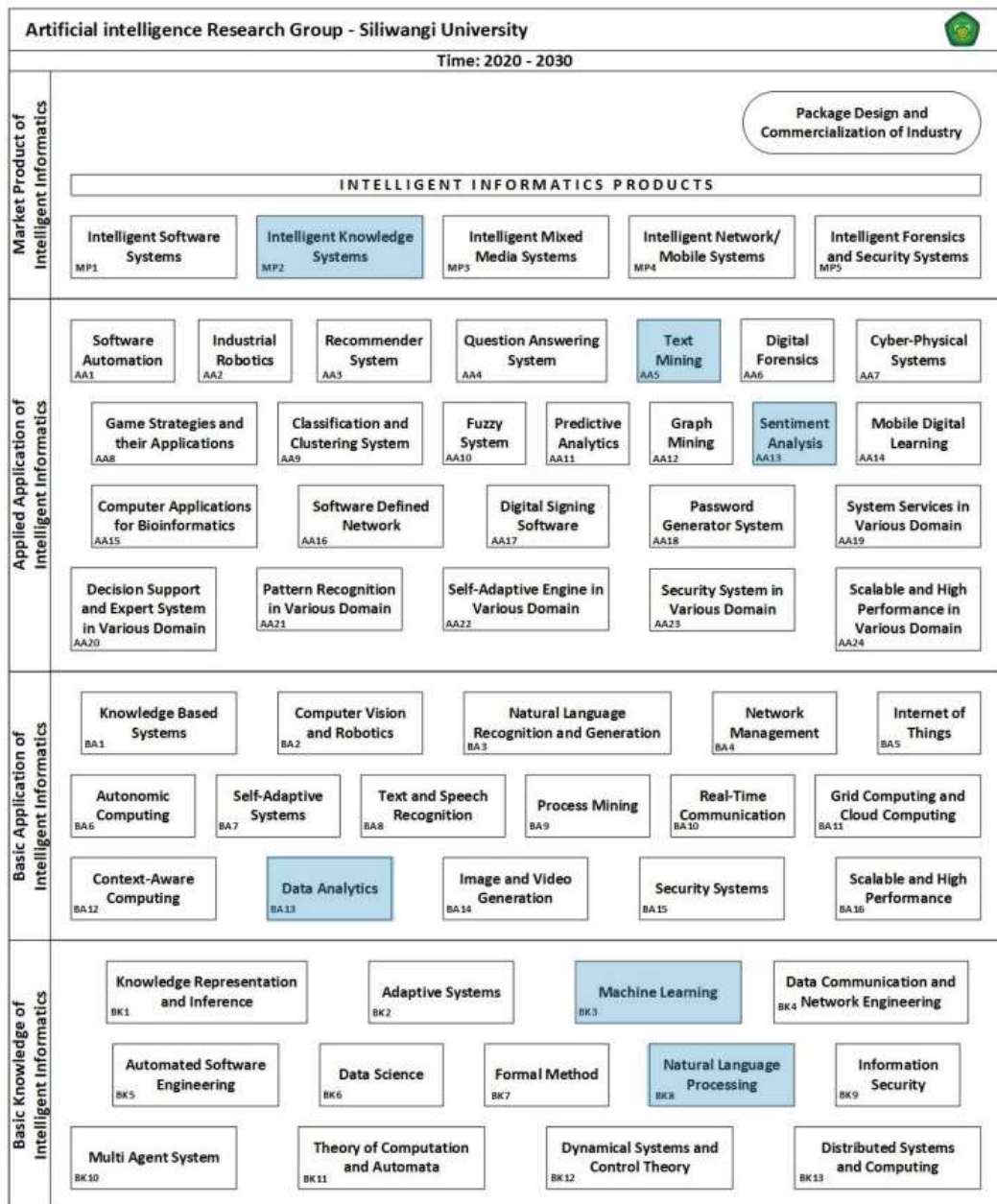
Evaluation dilakukan untuk mengukur performa metode *Textblob* dan *Vader* menggunakan *confusion matrix* dengan klasifikasi SVM.

3.1.6 *Deployment* (Penyampaian)

Deployment merupakan tahapan untuk menciptakan penyebaran informasi terkait beberapa visualisasi atau *reporting* dari analisis yang dilakukan sebelumnya.

3.2 **Peta Jalan (*Road Map*) Penelitian**

Peta jalan atau yang biasa disebut dengan *roadmap* merupakan suatu konsep arah penelitian yang dimaksudkan untuk menjelaskan ke arah mana penelitian akan dituju. Penelitian ini mengacu pada *roadmap* yang dikembangkan oleh *Artificial Intelligence Research Group* – Universitas Siliwangi tahun 2020 – 2030 yang ditunjukkan pada Gambar 3.2.



Gambar 3.2 Roadmap AI Research Group Universitas Siliwangi

(Sumber : AIS, 2020)

Berdasarkan Gambar 3.2 terdapat beberapa kajian penelitian pada penelitian ini ditandai dengan kolom berwarna biru. Kajian penelitian tersebut diantaranya, *Basic Knowledge* yang digunakan yaitu *Natural Language Processing* dan *Machine Learning*. *Basic Application* dalam penelitian ini yaitu *Data Analytics*, dan *Applied*

Application dalam penelitian ini yaitu *Text Mining* serta *Sentiment Analysis*. Berdasarkan beberapa kajian tersebut ditujukan untuk mencapai target yaitu, *Intelligent Knowledge System* yang berkaitan dengan komparasi metode pelabelan otomatis pada analisis sentimen aplikasi DANA.

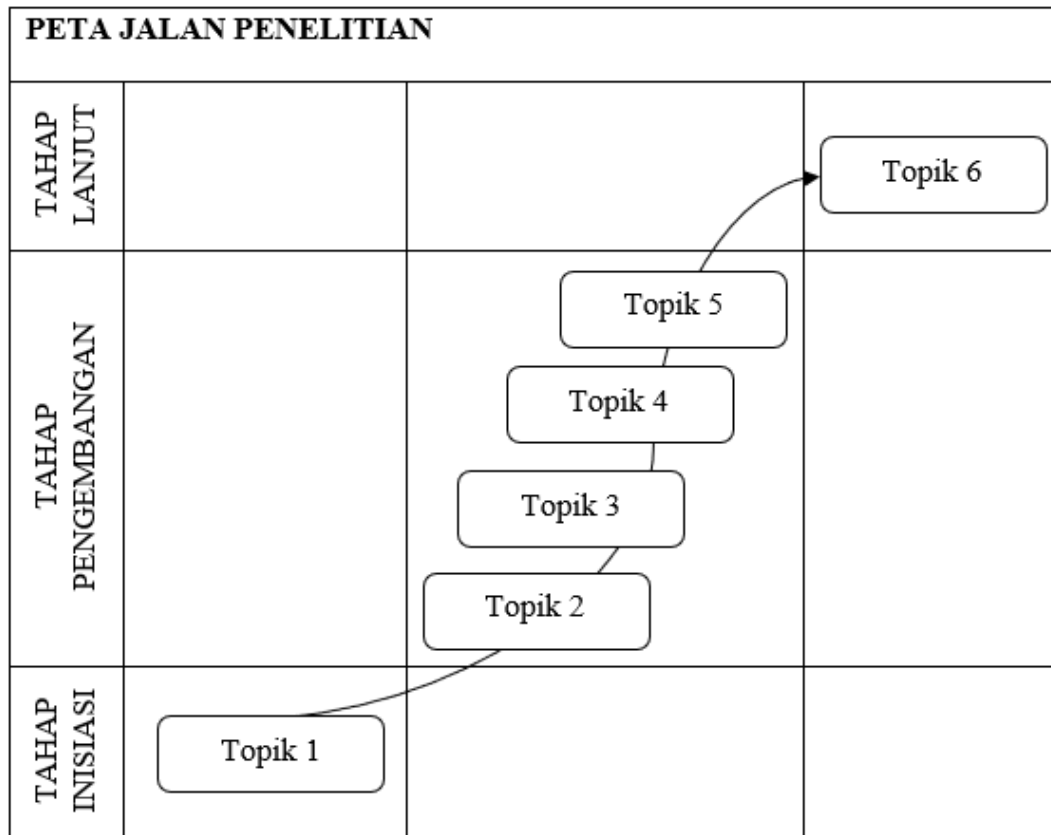
Berdasarkan *roadmap* pada Gambar 3.2, terdapat penelitian yang berkaitan dengan kajian-kajian penelitian yang digunakan. Berdasarkan penelitian terkait yang berjudul “Analisis Sentimen Ciutan Twitter Terkait Penerapan Permendikbudristek Nomor 30 Tahun 2021 Menggunakan TextBlob dan Support Vector Machine”.

Penelitian ini dilakukan karena di era digital seluruh aspek implementasi kebijakan pemerintah menjadi topik yang banyak diperbincangkan di jejaring sosial, khususnya Twitter. Salah satu topiknya terkait implementasi Peraturan Menteri Pendidikan dan Kebudayaan Nomor 30 Tahun 2021. Berdasarkan hasil pengujian oleh *Textblob*, didukung dengan penggunaan algoritma *Support Vector Machine*, menggunakan 48 data pengujian dengan 34 data positif serta 14 data negatif dan mencapai akurasi 70,8%. Diharapkan penelitian ini dapat dilanjutkan dengan topik dan metode lain untuk membandingkan hasil akurasi sistem dari masing-masing metode (Afrilia et al., 2022).

Selain itu, terdapat penelitian terkait lain yang mengacu pada Gambar 3.2 terkait kajian-kajian penelitian yang digunakan. Berdasarkan penelitian terkait yang berjudul “Analisis Sentimen Terhadap Sekolah Saat Covid-19 Pada Twitter Menggunakan Metode Lexicon Based”. Alasan dari penelitian ini adalah pembelajaran daring dinilai kurang efektif dan usulan Menteri Pendidikan pada 30

Maret 2020 untuk membuka kembali sekolah tatap muka secara bertahap di masa pandemi COVID-19. Kebijakan tersebut menuai banyak komentar masyarakat, khususnya di media sosial Twitter. Dengan menggunakan metode *Vader* untuk menganalisis sentimen kelas tatap muka di Twitter selama pandemi, hasil penelitian menunjukkan 5,4% sangat positif, 6,04% positif, 0,54% sangat negatif, 3,9% negatif dan 88,23% netral. Jika dilihat dari persentase opini masyarakat pengguna Twitter mengenai pendidikan tatap muka di masa pandemi virus corona, lebih banyak orang yang memberikan komentar netral. Terdapat batasan penelitian yaitu belum adanya *confusion matrix* sehingga hasil yang didapatkan belum terbukti validitasnya (Mulyadi and Lestari, 2022).

Gambar 3.3 merupakan tahapan dalam *roadmap* penelitian ini, yaitu tahap inisiasi, tahap pengembangan, dan tahap lanjut.

Gambar 3.3 *Roadmap* Penelitiannya**Keterangan:**

Topik 1 : *Business Understanding*

Topik 2 : *Data Understanding*

Topik 3 : *Data Preparation*

Topik 4 : *Modelling*

Topik 5 : *Evaluation*

Topik 6 : *Deployment*

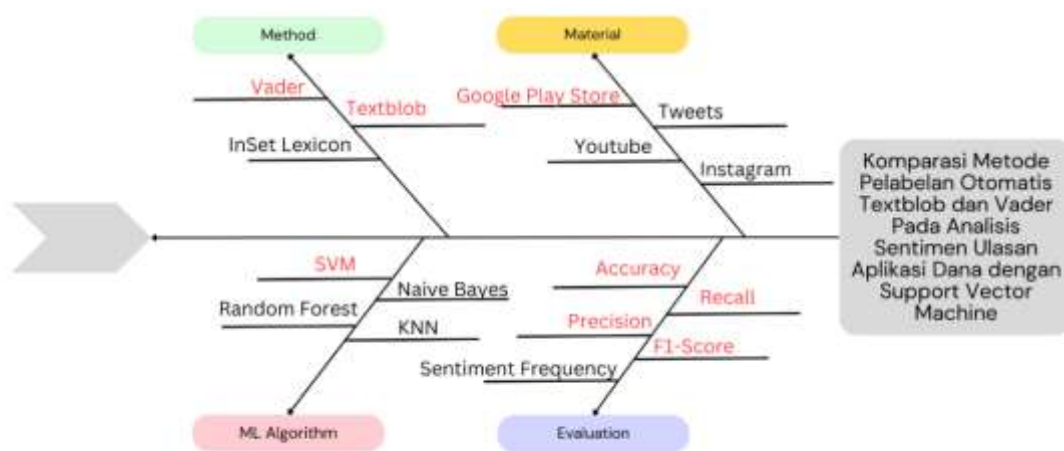
Tahap inisiasi merupakan tahapan pertama dalam penelitian ini yaitu *Business Understanding*. Tahapan ini merupakan tahapan mengidentifikasi pengetahuan yang berkaitan dengan penelitian.

Tahap pengembangan merupakan tahapan paling penting pada penelitian ini. Tahapan ini terdiri dari 1) *Data Understanding* untuk memahami data yang digunakan, 2) *Data Preparation* untuk membersihkan dan mempersiapkan data sebelum dilakukan permodelan, 3) *Modelling* untuk melatih model klasifikasi dan menguji permodelan tersebut, dan 4) *Evaluation* untuk mengukur hasil klasifikasi.

Tahap lanjut merupakan tahapan pengembangan secara lanjut terhadap model yang dibuat sebelumnya. Tahapan ini merupakan tahapan *Deployment* yang bertujuan untuk menyampaikan hasil analisis data yang dilakukan sebelumnya.

3.3 Fishbone Diagram

Prinsip dasar dari *fishbone diagram* adalah menempatkan akar permasalahan di bagian kepala atau sebelah kanan dari struktur rangkaian tulang ikan. Fungsi utama dari diagram ini adalah untuk menggambarkan faktor-faktor kunci yang memengaruhi kualitas dan menghasilkan konsekuensi terhadap masalah yang sedang diteliti (Tamardina, Yasin and Ispriyanti, 2022).



Gambar 3.4 Fishbone Diagram

Gambar 3.4 menunjukkan beberapa faktor kunci pada penelitian yang ditandai dengan teks berwarna merah diantaranya yaitu *Vader*, *Textblob*, Google Play Store, SVM, *Accurary*, *Recall*, *F1-Score*, dan *Precision*. Masing – masing bagian pada tulang memiliki keterangan sebagai berikut:

- a. *Material*, yaitu sumber data yang digunakan dalam penelitian. Data yang digunakan yaitu ulasan aplikasi DANA yang diperoleh dari Google Play Store.
- b. *Method*, yaitu metode yang digunakan untuk pelabelan otomatis yang mana pada penelitian ini menggunakan *Vader* dan *Textblob*.
- c. *ML Algorithm*, yaitu algoritma *Machine Learning* yang digunakan dalam penelitian. Algoritma yang digunakan yaitu SVM.
- d. *Evaluation*, yaitu hasil yang diperoleh dari penelitian. Adapun hasil dari penelitian yang dilakukan adalah mengetahui *Accurary*, *Recall*, *F1-Score*, dan *Precision*.

3.4 Parameter Penelitian

Parameter penelitian merupakan segala hal yang dipilih untuk dipelajari dengan cara apapun dalam mengumpulkan data untuk membuat kesimpulan (Moto, 2019). Tabel 3.1 merupakan parameter penelitian ini yang terdiri dari 3 bagian.

Tabel 3.1 Parameter Penelitian

No	Nama	Tipe Data	Deskripsi
1	<i>Username</i>	<i>object</i>	Nama pengguna yang memberikan ulasan
2	<i>At</i>	<i>object</i>	Tanggal dibuatnya ulasan
3	<i>Content</i>	<i>datetime</i>	Isi ulasan yang dibuat oleh pengguna

Berdasarkan Tabel 3.1 parameter dalam penelitian ini yaitu *userName* sebagai informasi nama *user* yang memberikan ulasan. Selain itu, terdapat parameter *at* sebagai informasi terkait waktu dibuatnya ulasan serta parameter *content* sebagai informasi isi ulasan yang disampaikan *user*. Berdasarkan ketiga parameter tersebut, parameter yang digunakan untuk proses analisis sentimen yaitu parameter *content*, sedangkan untuk parameter *userName* dan *at* digunakan sebagai parameter pendukung sebelum proses analisis sentimen.