

BAB III

METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

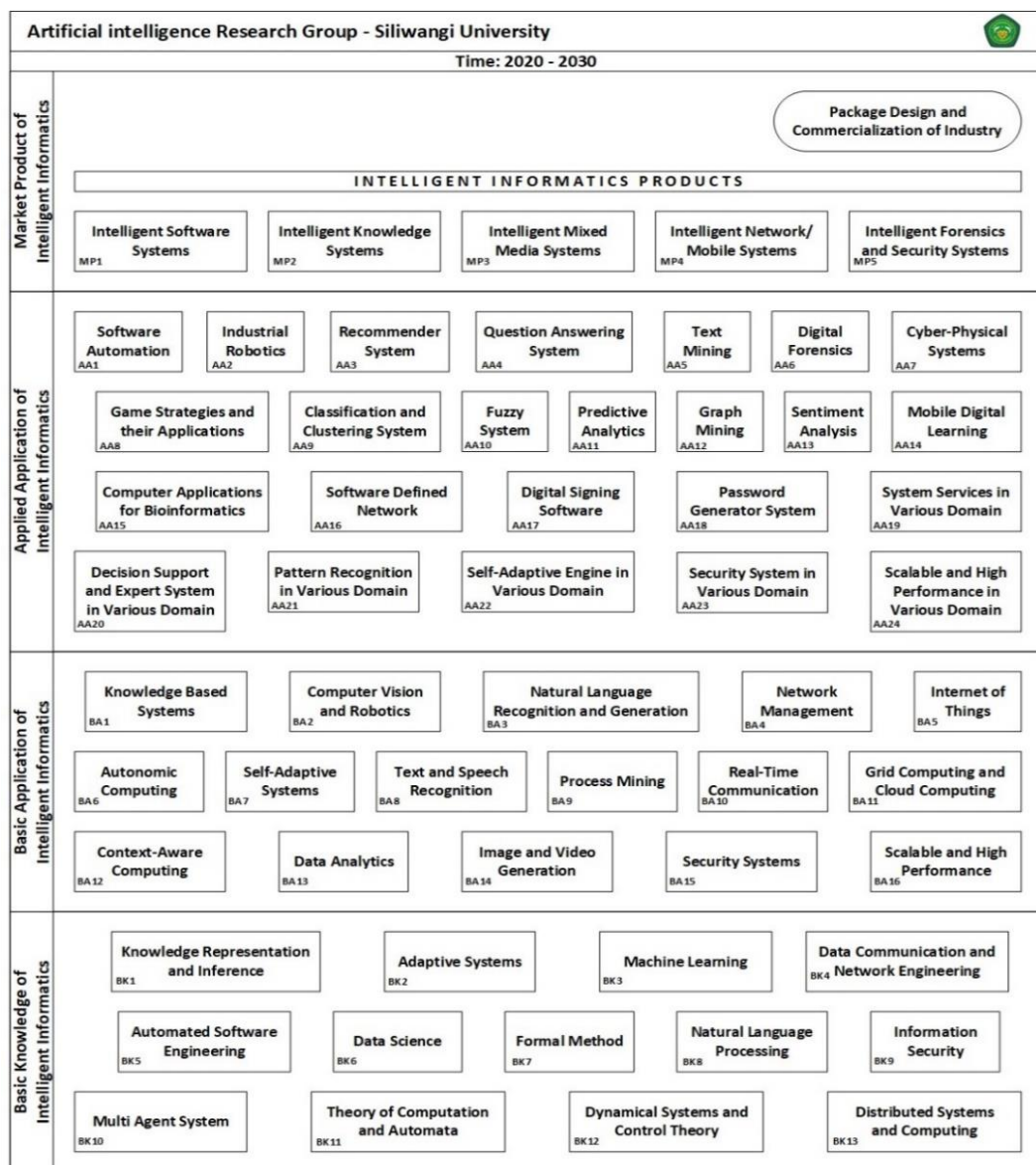
Penelitian ini menggunakan metodologi kuantitatif yang merupakan jenis penelitian sistematis, terencana dan terstruktur secara jelas dari awal sampai pembuatan desain penelitiannya. Berlandaskan pada filsafat *positivism* untuk meneliti populasi atau sampel tertentu. Dilanjutkan dengan analisis data bersifat kuantitatif atau statistik untuk menguji hipotesis yang telah ditetapkan.

Data yang diambil merupakan review dari website Google Play mengenai aplikasi investasi online yaitu Ajaib Sekuritas. Pengambilan sampel dilakukan secara acak dengan waktu tertentu untuk memperoleh data yang akan dianalisis menggunakan algoritma *machine learning* untuk didapatkan klasifikasi antara positif, negatif atau netral setiap sentimen.

3.2 Peta Jalan (Road Map) Penelitian

Suatu konsep arah penelitian yang dimaksudkan untuk menjelaskan ke arah mana penelitian akan dituju disebut peta jalan (Roadmap). Roadmap pada penelitian ini mengacu pada Roadmap Artificial Intelligence Research Group \rightarrow Universitas Siliwangi tahun 2020 – 2030. Roadmap Artificial Intelligence Research Group \rightarrow Universitas Siliwangi tahun 2020 – 2030 ini merupakan kolaborasi antara Kelompok Keahlian (KK) Informatika dan Sistem Inteligen (ISI) dengan Kelompok Keahlian (KK) Jaringan, Keamanan dan Digital Forensik (JKF) jurusan

Informatika, Fakultas Teknik, Universitas Siliwangi. Roadmap Artificial Intelligence Research Group \cap Universitas Siliwangi tahun 2020 – 2030 ditunjukkan pada Gambar 3.1



Gambar 3.1 Roadmap AI Research Group Universitas Siliwangi 2020 – 2030 (Sumber AIS, 2019)

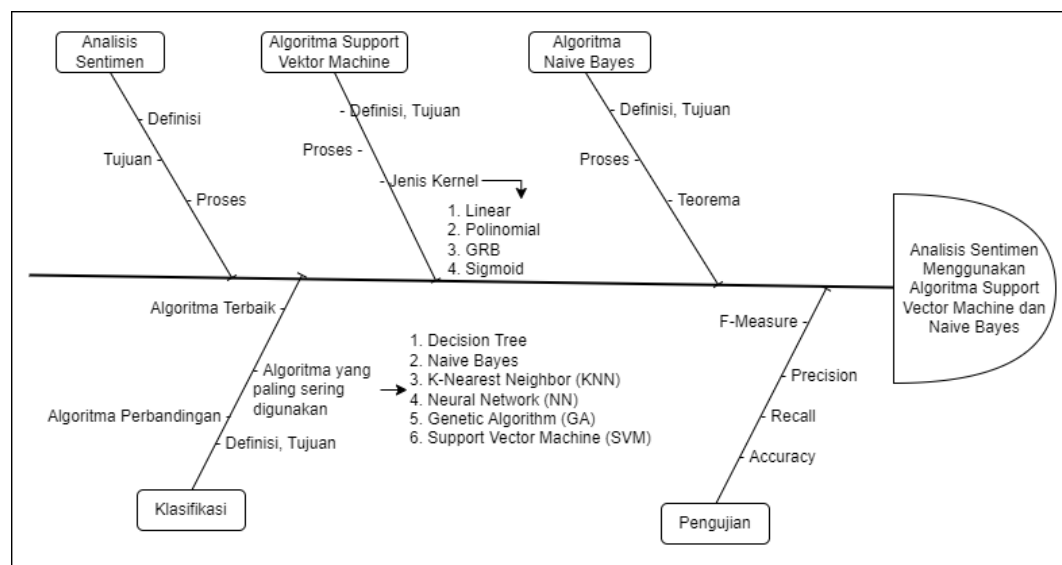
Berdasarkan Roadmap Artificial Intelligence Research Group – Universitas Siliwangi tahun 2020 – 2030. Basic Knowledge yang digunakan pada penelitian ini

adalah *Natural Language Processing* (NLP) dan *Machine Learning* (ML), *Basic Application* pada penelitian ini adalah *Data Analytics* dan *Process Mining*, dan *Applied Application* yang digunakan pada penelitian ini adalah *Classification and Clustering System*, *Text Mining* dan *Sentiment Analysis*.

3.3 Tahapan Penelitian

Penelitian ini menggunakan algoritma klasifikasi yaitu *Naïve Bayes* dan *Support Vector Machine (SVM)*, pemilihan kedua algoritma tersebut tidak serta merta asal pilih. Penggunaan diagram *fishbone* membantu untuk menentukan algoritma mana yang terbaik dalam mengklasifikasin data untuk analisis sentimen.

Gambar 3.2 merupakan diagram *fishbone* penelitian ini:



Gambar 3.2 Diagram Fishbone

Analisis Sentimen, terdiri dari defenisi, tujuan dan proses yang dilakukan dengan membaginya menjadi beberapa tingkatan.

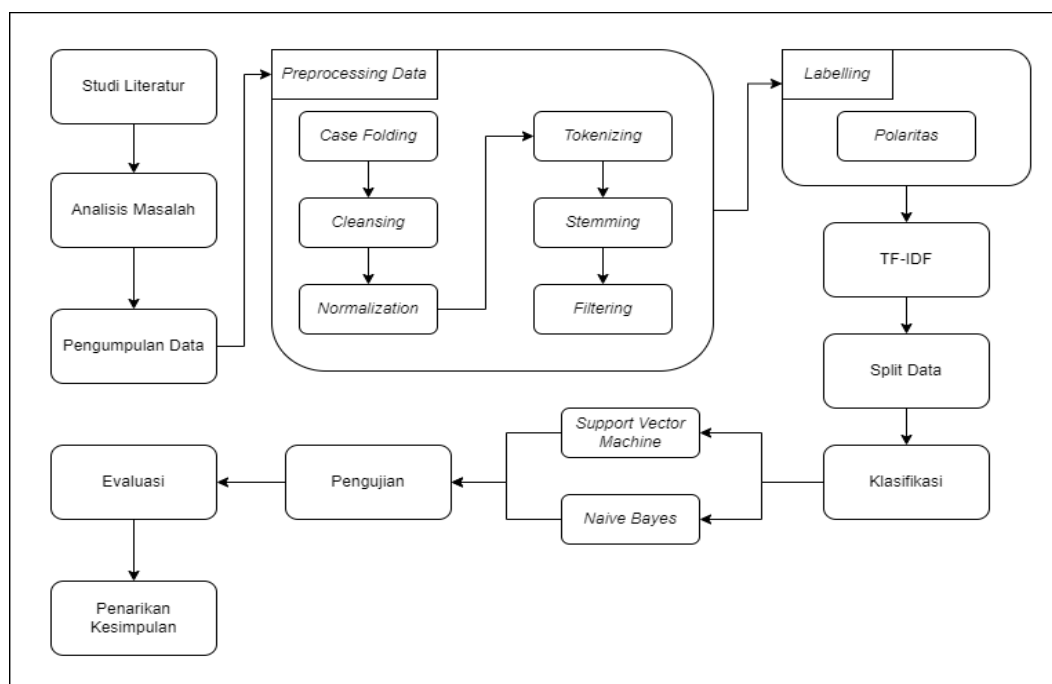
Klasifikasi, yang terdiri dari defenisi, tujuan, algoritma serupa yang paling sering digunakan, algoritma pembanding dan algoritma terbaik dari metode serupa.

Algoritma *Naïve Bayes*, pembahasan algoritma *Naïve Bayes* meliputi defenisi, tujuan, teorema dan proses/tahapan algoritma.

Algoritma *Support Vector Machine*, pembahasan algoritma *Support Vector Machine* meliputi defenisi, tujuan, teorema dan proses/tahapan algoritma.

Pengujian, dilakukan dengan mengukur nilai akurasi, presisi, *recall* dan *F-Measure*.

Pada penelitian ini dilakukan tahapan dengan alur yang sesuai dengan tahapan penelitian lainnya, yaitu dengan mencari data yang berhubungan dengan penelitian, menganalisis masalah pada penelitian dan mengumpulkan data untuk kebutuhan penelitian. Untuk lebih lengkapnya digambarkan pada Gambar 3.3.



Gambar 3.3 Skema Penelitian

3.3.1 Studi Literatur

Studi literatur dilaksanakan dengan cara mencari segala bentuk informasi dan penelitian yang berkaitan dengan Analisis Sentimen, Klasifikasi dengan algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)*, *Text Mining*, *Machine Learning*, *Web Scraping* dan segala hal yang berkaitan dengan model pemrogramannya.

3.3.2 Analisis Masalah

Pada penelitiannya Mardiana, Syahreva dan Truslaela di tahun 2019 yang berjudul “Komparasi Metode Klasifikasi Pada Analisis Sentimen Usaha Waralaba Berdasarkan Data Twitter” menghasilkan bahwa algoritma *Support Vector Machine* merupakan salah satu algoritma terbaik dalam mengklasifikasikan *review*. Pemilihan algoritma *Naïve Bayes* dan *Support Vector Machine* dikarenakan kedua algoritma tersebut sudah banyak digunakan, memiliki nilai akurasi yang tinggi dan tidak memakan waktu pemrosesan. Ditambah situs *review online* terus bertambah populer karena semakin banyak orang mencari saran dari sesama pengguna layanan dan produk. Selain itu juga tidak jarang ditemukan beberapa isi *review* yang tidak relevan dengan *rating* yang diberikan, contohnya sebuah *review* yang berisi komentar positif namun diberi *rating* 2 dari 5, dan sebaliknya. Maka dari itu, dibutuhkan sebuah algoritma yang dapat secara cepat dan tepat menganalisis *review* pengguna dan mengekstrak informasi yang paling relevan.

3.3.3 Pengumpulan Data

Pengumpulan data dilakukan dengan teknik *web scraping*. Menurut Vargiu (2012) *web scraping* merupakan suatu gabungan teknik yang digunakan untuk mendapatkan informasi dari sebuah *website* secara otomatis tanpa harus menyalinnya secara manual. *Web scraping* dilakukan di *website Google Play* pada *review* aplikasi Ajaib Investasi. Data yang telah didapatkan kemudian diambil data dengan rentang waktu dari bulan Maret 2021 sampai dengan Agustus 2021. Pada proses pengumpulan data ini menghasilkan data sebanyak 7943 data. Dari data yang telah dikumpulkan, didapatkan data yang memiliki 3 atribut, *date* (tanggal pembuatan ulasan), *rating* (tingkat kepuasan pengguna) dan *review* (isi ulasan).

3.3.4 Preprocessing Data

Tahap *Preprocessing* Data merupakan proses yang berfungsi untuk membersihkan teks sebelum dilakukan pengolahan lebih lanjut. Data teks mentah yang diperoleh biasanya tidak terstruktur serta terdapat banyak *noise* seperti tanda baca, imbuhan, angka, karakter-karakter khusus dan lainnya. Beberapa tahapan *text processing* yang digunakan pada penelitian ini yaitu:

a. *Case Folding*

Tahap ini dilakukan untuk mengkonversi keseluruhan teks menjadi suatu bentuk standar atau mengubah keseluruhan teks menjadi *lowercase* (format penulisan huruf kecil)

b. *Tokenizing*

Tahap ini merupakan tahapan yang dilakukan untuk memotong *string* input berdasarkan kata penyusunnya. Proses *tokenizing* akan memecah teks sekumpulan karakter yang terdapat dalam teks ke dalam satuan kata, sehingga dapat memudahkan untuk membedakan karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan.

c. *Filtering*

Pada tahap ini dilakukan pengambilan kata-kata penting hasil dari proses *tokenizing*. Dalam tahap ini juga dapat digunakan algoritma *stoplist* (membuang kata kurang penting) dan *wordlist* (menyimpan kata penting).

3.3.5 Pembobotan Data

Pada tahap ini, data yang dihasilkan pada tahap *preprocessing* dilakukan pembobotan untuk setiap kata yang terdapat pada data tersebut. Pembobotan ini dilakukan menggunakan metode *TF-IDF*. Selain untuk menghasilkan bobot kata, proses ini juga akan mengubah data teks menjadi data numerik.

3.3.6 Klasifikasi

Klasifikasi adalah proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han, 2006). Algoritma yang digunakan pada penelitian ini adalah:

a. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi (Santosa, 2007). Tujuan dari metode *Support Vector Machine* adalah untuk menemukan pemisah linier diruang pencarian yang dapat memisahkan berbagai kelas (Jayalekshmi & Mathew, 2017). Pada ruang berdimensi tinggi, akan dicari *hyperplane* yang dapat memaksimalkan jarak (margin) antara kelas data. Menurut (Santosa, 2007). Disini *SVM* tidak ditujukan untuk memprediksi kelas diskrit ($y_i \in \{+1, -1\}$), melainkan untuk upaya untuk mempelajari hubungan input-output antara masukan pelatihan tupel X_i , dan dengan nilai output, $y_i \in R$.

b. *Naïve Bayes*

Naïve Bayes adalah sebuah metode klasifikasi dengan metode probabilitas dan statistic yang dikemukakan oleh ilmuwan Inggris yaitu Thomas Bayes, yang digunakan untuk memprediksi di masa depan berdasarkan pengalaman di masa sebelumnya (Bustami, 2013). *Naïve Bayes* didasarkan pada asumsi penyerdehanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Keuntungan menggunakan algoritma *Naïve Bayes* adalah bahwa algoritma ini hanya membutuhkan jumlah data pelatihan (*data training*) yang kecil untuk menemukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Dalam kebanyakan situasi dunia nyata yang kompleks *Naïve Bayes* cenderung bekerja lebih baik dari pada yang diharapkan (Pattekari & Parveen, 2012).

3.3.7 Evaluasi

Metode yang digunakan untuk melakukan uji performansi pada penelitian ini adalah *Confusion Matrix*. *Confusion matrix* merupakan metode yang digunakan untuk melakukan perhitungan akurasi pada *data mining* atau *supervised learning*.

3.3.8 Penarikan Kesimpulan

Penarikan kesimpulan dilakukan berdasarkan hasil pengujian dan evaluasi kerja pengujian yang dilakukan dengan metode *confusion matrix* mendapatkan nilai yang lebih tinggi daripada kegagalan dan data dapat diklasifikasikan dengan tepat. Dengan demikian penelitian dapat dikatakan berhasil dalam melakukan pemrosesan data dan membantu dalam mengklasifikasikan *review online*.