

# BAB I PENDAHULUAN

## 1.1. Latar Belakang Penelitian

Sepak bola adalah salah satu cabang olahraga bola besar yang terdiri dari 11 orang pemain dan dimainkan secara berkelompok. Faktanya, sepak bola merupakan olahraga paling populer di dunia yang dimainkan oleh sekitar 250 juta orang di lebih dari 200 negara [1]. Satu hal yang membuat sepak bola menjadi olahraga terpopuler di dunia yaitu dikarenakan banyaknya penggemar yang diperkirakan mencapai 3,3 sampai 3,5 miliar jiwa di berbagai penjuru dunia [2], di mana jumlah tersebut setara dengan 44% populasi manusia.

Menurut media sepak bola yang bernama Football Daily, dalam salah satu videonya yang diunggah di kanal Youtube mereka, mengatakan bahwa terdapat 10 alasan mengapa sepak bola disebut sebagai olahraga terpopuler dan juga terbaik di dunia. Alasan-alasan tersebut di antaranya yaitu: *The Global Game* yang bermakna sepak bola dapat dinikmati di seluruh penjuru dunia, *Awareness Generation* yang menjelaskan pengaruh sepak bola terhadap isu-isu sosial, *The Skill* yang merujuk pada aksi-aksi dari para pemain sepak bola di dalam lapangan, *The Passion* yang menggambarkan gairah serta perasaan para pemain, pelatih, dan pendukung sepak bola, baik itu di dalam atau bahkan di luar lapangan, *The United Community* yang menunjukkan bahwa komunitas sepak bola bisa ditemukan di segala penjuru dunia, *Bragging Rights* yang memiliki arti bahwasanya sepak bola dapat menjadi bahan candaan di beberapa kalangan, *The Unpredictability* yang menggambarkan bahwa sepak bola bukanlah sebuah olahraga yang mudah untuk ditebak, *The Peace* yang menunjukkan bahwa sepak bola dapat melambangkan kedamaian serta solidaritas, *The Opportunities* yang merujuk pada peluang sepak bola dalam menjadikannya sebagai mata pencaharian, dan terakhir adalah *The Accessibility* yang menunjukkan bahwa sepak bola merupakan olahraga yang dapat dimainkan oleh siapa pun, di mana pun, dan kapan pun [3]. Selain dicap sebagai olahraga yang paling populer di dunia, sepak bola sering dijuluki sebagai *The Beautiful Game* yang mengandung makna bahwasannya sepak bola itu sangat indah jika dimainkan dengan permainan yang ciamik dan penuh sportivitas.

Menilik alasan-alasan yang telah dipaparkan di atas, sangatlah masuk akal jika sepak bola menjadi sedemikian besar, kuat, dan berpengaruh dikarenakan dapat menimbulkan perasaan serta kepuasan tertentu terhadap diri manusia. Perasaan dan kepuasan tersebut hanya mampu diperoleh jika menjadi bagian dari sepak bola itu sendiri, sehingga mudah disimpulkan bahwa sepak bola merupakan jenis olahraga yang memiliki kekuatan magis dalam membangkitkan gairah, mengunggah gaya, mendobrak selera, serta memunculkan rasa bangga yang sebelumnya tersimpan di dalam diri manusia dikarenakan dapat membuat mereka merasa menjadi bagian dari sesuatu yang lebih besar di luar dirinya [4], itulah mengapa sepak bola mempunyai unsur *passion* di dalamnya.

*Passion* atau gairah ini sangatlah kental di dalam sepak bola, karena bukan hanya dimiliki oleh para pemain dan pelatih, akan tetapi dimiliki juga oleh sebagian penikmat sepak bola yang ada di seluruh dunia [5]. Khusus untuk penikmat sepak bola yang mendukung sebuah klub, *passion* di sini bersifat personal serta dapat memicu rasa memiliki hingga fanatisme terhadap sebuah klub [4]. Seseorang yang memahami sejarah perjuangan Manchester United dalam meraih gelar liga Inggris yang ke-20 kalinya dalam di tahun 2013 dapat memicu dirinya untuk mencintai hal-hal yang berhubungan dengan Manchester United itu sendiri, dengan kata lain *passion* di sini memiliki makna sebagai sebuah hal yang merepresentasikan gairah serta perasaan dari para pendukung sepak bola dalam menyalurkan *emotion* mereka terhadap klub yang dicintainya.

Berbicara mengenai *emotion* atau emosi, dalam ilmu psikologi didefinisikan sebagai kumpulan perasaan secara kompleks yang dapat mengakibatkan perubahan fisik serta psikologis dalam mempengaruhi pikiran dan perilaku manusia [6]. Emosi sering dikaitkan dengan berbagai macam fenomena psikologis yang ada, dari mulai kesenangan, kesabaran, keberanian, kesedihan, kemarahan, kegelisahan, dan masih banyak lagi fenomena psikologis lain yang bersifat positif hingga negatif, sehingga dapat dikatakan bahwa sentimen sangat berkaitan erat dengan emosi yang ada pada diri manusia. Emosi ini lah yang menjadikan para pendukung sepak bola memiliki sentimen yang berbeda-beda terhadap sebuah klub sepak bola, baik itu klub yang mereka dukung ataupun tidak.

Berkaitan dengan perbedaan sentimen yang ada di kalangan para pendukung sepak bola, sering kali mereka mengutarakan komentar berupa kritik ataupun opini terhadap sebuah klub, baik itu komentar yang bersifat positif, negatif, atau bahkan netral [7]. Hal tersebut dilakukan atas dasar mengekspresikan perasaan yang dialami sebagai pendukung sepak bola, seperti halnya mengomentari para pemain dan staf kepelatihan klub, mengkritik gaya permainan dan jalannya sebuah pertandingan, memprediksi masa depan sebuah klub sepak bola yang ada di bawah kepemimpinan manajemen klub tersebut, hingga memperdebatkan klub sepak bola mana yang akan keluar sebagai juara pada suatu kompetisi [4], semua opini itu sering mereka lontarkan di media sosial.

Fenomena tersebut sering terjadi karena adanya emosi yang didasari dengan unsur fanatisme. Namun terkadang komentar-komentar yang dituangkan di media sosial sering berlawanan dengan apa yang terjadi di dunia nyata, di mana efeknya membuat pandangan setiap pendukung sepak bola menjadi berbeda-beda dan sering beradu komentar hingga berujung pada perdebatan [7]. Hal tersebut menjadi salah satu permasalahan yang ada di kalangan pendukung sepak bola. Pasalnya, tidak sedikit pula dari mereka yang mengomentari sebuah klub tanpa melihat realita yang ada, sehingga untuk mengetahui kenyataannya, diperlukan sebuah pembuktian berupa data agar hal tersebut bisa menjadi nilai yang dapat diperdebatkan. Sebagai solusi permasalahan tersebut, diperlukannya sebuah sistem yang dapat memproses suatu kalimat untuk mengetahui sentimen dari komentar-komentar di media sosial secara otomatis, yang mana sistem tersebut biasa dikenal dengan nama *sentiment analysis* atau analisis sentimen.

Analisis sentimen merupakan suatu cara atau teknik yang dapat menentukan nilai sentimen dari suatu kalimat atau teks dengan menghitung polaritasnya sebagai tujuan dalam meninjau subjektivitas dari kalimat atau teks tersebut [8]. Seperti halnya pada penelitian ini, analisis sentimen digunakan untuk mengetahui sentimen dari opini serta komentar para pendukung sepak bola yang dilontarkan di media sosial, entah hasilnya bersifat positif, negatif ataupun netral. Fakta menariknya, sistem analisis sentimen dapat disebut sebagai salah satu teknologi *Artificial Intelligence* (AI) yang merupakan kombinasi antara *natural language processing* dan *machine learning*, dikarenakan dalam prosesnya dapat menetapkan sentimen

untuk suatu kalimat atau teks selayaknya otak manusia [9]. Hal itu sangatlah masuk akal lantaran proses analisis teks tersebut merupakan suatu masalah kognitif yang pada umumnya sangat berkaitan dengan kecerdasan otak manusia, contohnya seperti pembelajaran, pemecahan masalah, atau bahkan pengenalan pola [10].

Melihat era teknologi sekarang yang sebegitunya berkembang pesat dengan melimpahnya banyak data, rasanya sistem berbasis *Artificial Intelligence* sangatlah dibutuhkan dalam memecahkan berbagai macam permasalahan yang ada. Khusus untuk penelitian tugas akhir ini, ruang lingkup *natural language processing* serta tentunya *machine learning* merupakan subbidang yang tepat untuk menyelesaikan permasalahan tersebut, terlebih bersamaan dengan dipadukannya salah satu metode *deep learning* pada sistem analisis sentimen [11], secara tidak langsung terbukti dapat meningkatnya akurasi penilaian pada pemrosesannya [12].

*Deep learning* adalah sebuah metode pengembangan jaringan saraf tiruan atau *artificial neural network* yang merupakan subbidang dari *machine learning* dengan metode yang terinspirasi dari struktur otak pada manusia dan memiliki lebih banyak *hidden layer*. Dengan banyaknya *hidden layer* yang ada, *deep learning* dapat melakukan proses pengenalan secara kompleks [13], baik itu pengenalan citra digital, suara, atau bahkan teks, seperti halnya yang dilakukan dalam penelitian ini. Dibanding dengan metode generasi sebelumnya, salah satu kelebihan dari *deep learning* yaitu ada pada pemrosesan data. Semakin banyak data yang digunakan, maka kinerja dari metode tersebut akan lebih baik dibandingkan saat menggunakan sedikit data [14]. Terlepas dari keefektifannya, metode *deep learning* ini memiliki banyak sekali jenis yang salah satunya yaitu *Recurrent Neural Network*, di mana algoritma tersebut dirasa tepat untuk menyelesaikan permasalahan yang ada [15].

Sama halnya dengan *deep learning*, struktur yang terdapat pada algoritma *Recurrent Neural Network* pun memiliki banyak *hidden layer* yang berfungsi untuk menangkap informasi dari setiap *sequence*. Biasanya data yang digunakan untuk mengetahui kinerja dari algoritma *Recurrent Neural Network* ialah data sekuensial atau data yang berurutan serta berjenis data *time series*, seperti halnya *tweet* pada media sosial Twitter. Meski begitu, dalam pemrosesan *Recurrent Neural Network* ini sering mengalami permasalahan yang disebut sebagai *exploding* atau *vanishing gradient* [16].

Maksud dari *vanishing gradient* ini yaitu ketika jarak antara *input* data terlalu besar, *Recurrent Neural Network* tidak dapat menghubungkan informasi yang relevan dengan data tersebut [17]. Selain itu, masalah *vanishing gradient* di sini membatasi RNN untuk menerapkan konteks “*long-term dependencies*” [18]. Sehingga untuk menangani konteks “*long-term dependencies*” yang dibatasi oleh permasalahan *vanishing gradient* pada algoritma *Recurrent Neural Network*, Hochreiter dan Schmidhuber [15] mengajukan sebuah arsitektur *Recurrent Neural Network* modern yang bernama *Long Short-Term Memory*. Mengimplementasikan arsitektur *Long Short-Term Memory* merupakan suatu cara paling populer dalam mengatasi permasalahan tersebut, meskipun demikian [19] terdapat cara lainnya yaitu dengan menggunakan *Rectified Linear* atau ReLu saat menentukan *activation function* pada *Recurrent Neural Network*.

Berdasarkan hal-hal yang telah dipaparkan di atas, langkah yang dilakukan dalam menganalisis sentimen terkait opini para pendukung sepak bola yaitu dengan mengimplementasikan arsitektur *Long Short-Term Memory* yang merupakan varian dari algoritma *Recurrent Neural Network* dalam memproses banyaknya jenis data sekuensial serta data *time series*, sehingga judul yang tepat pada penelitian ini ialah **ANALISIS SENTIMEN TERHADAP OPINI PENDUKUNG SEPAK BOLA MENGGUNAKAN ALGORITMA *RECURRENT NEURAL NETWORK***.

### **1.2. Rumusan Masalah Penelitian**

Berdasarkan uraian latar belakang penelitian di atas, maka dapat ditentukan rumusan masalahnya sebagai berikut:

1. Bagaimana mengimplementasi serta mengetahui kinerja algoritma *Recurrent Neural Network* untuk menganalisis sentimen terhadap opini pendukung sepak bola?
2. Bagaimana menganalisis hasil sentimen pendukung sepak bola terhadap opini yang dilontarkan di media sosial?

### **1.3. Tujuan Penelitian**

Tujuan yang hendak dicapai dalam penelitian ini yaitu berdasarkan rumusan masalah penelitian yang telah ditentukan, yang mana di antaranya adalah:

1. Mengimplementasi serta mengetahui kinerja dari algoritma *Recurrent Neural Network* untuk menganalisis sentimen terhadap opini pendukung sepak bola.
2. Menganalisis hasil sentimen para pendukung sepak bola terhadap opini yang dilontarkan di media sosial.

#### 1.4. Manfaat Penelitian

Selain tujuan, adapun beberapa manfaat yang dapat diambil dari penelitian ini, yang mana manfaat tersebut dibagi menjadi dua golongan:

1. Bagi penulis, penelitian ini jelas dapat menambah wawasan serta pengetahuan dalam memanfaatkan teknologi *Artificial Intelligence* untuk membangun suatu model *machine learning* yang mampu menentukan sentimen secara otomatis dengan memanfaatkan algoritma *Recurrent Neural Network* sebagai metode penyelesaian masalahnya. Selain itu, manfaat lainnya ialah dapat memberikan informasi terkait sentimen terhadap sebuah klub sepak bola yang dilontarkan oleh para pendukung sepak bola di media sosial Twitter.
2. Bagi pembaca, penelitian ini dapat dijadikan referensi untuk mengembangkan model *machine learning* sebagai sistem yang mampu mempermudah pekerjaan manusia.

#### 1.5. Batasan Masalah Penelitian

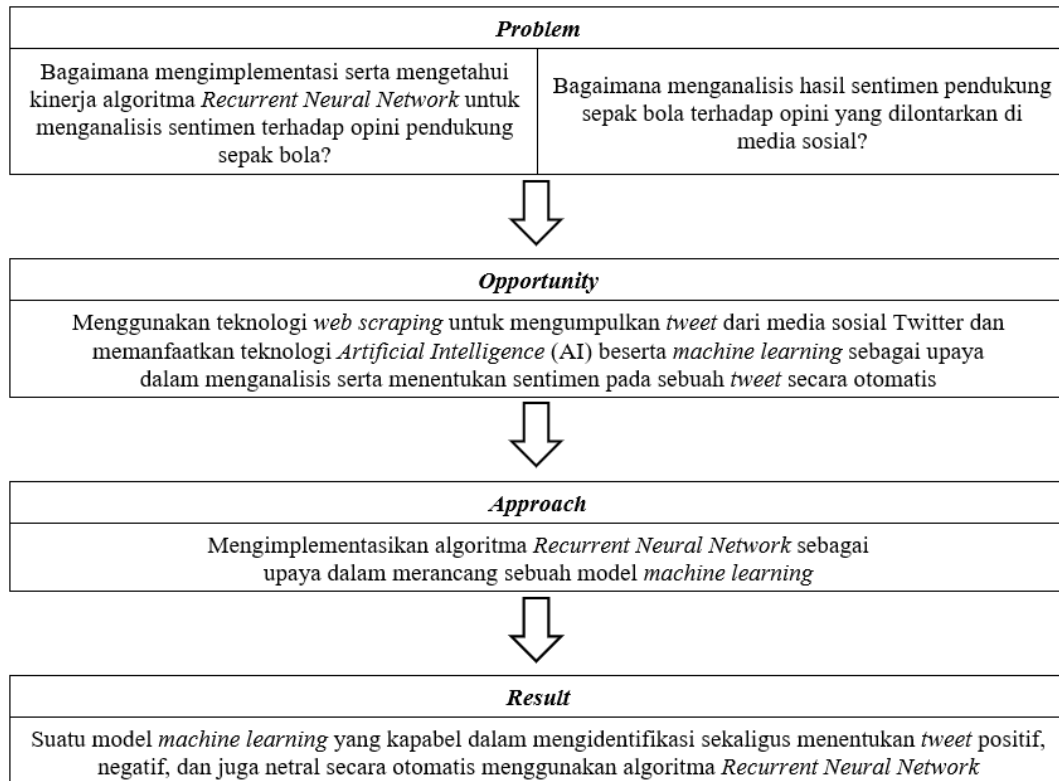
Ruang lingkup pada penelitian ini sangat luas, sehingga diperlukan sejumlah batasan masalah agar tujuan penelitian tidak menyimpang dan tetap sesuai dengan yang direncanakan. Batasan masalah tersebut di antaranya yaitu:

1. Sistem ini dirancang menggunakan bahasa pemrograman Python yang dibuat dalam sebuah *notebook* berbasis *cloud* bernama Google Colaboratory (Google Colab).
2. Model pengembangan perangkat lunak yang digunakan dalam penelitian yaitu CRISP-DM.
3. Algoritma yang dimanfaatkan untuk menentukan sentimen sebuah *tweet* yaitu *Long Short-Term Memory*, yang mana algoritma tersebut merupakan salah satu turunan dari algoritma *Recurrent Neural Network*.

4. Data yang digunakan penelitian ini merupakan sebuah *dataset* berisi kumpulan *tweet* berbahasa Inggris terkait tiga klub sepak bola tersukses sepanjang masa liga Inggris yaitu Arsenal, Liverpool, dan Manchester United [20] [21], yang diambil dari media sosial Twitter dengan menggunakan kata kunci:
  - “@Arsenal” untuk *dataset* klub Arsenal.
  - “@LFC” untuk *dataset* klub Liverpool.
  - “@ManUtd” untuk *dataset* klub Manchester United.
5. Pengumpulan *dataset* dari media sosial Twitter diambil dengan memanfaatkan salah satu *library* bahasa pemrograman Python bernama Snsrape.
6. Rentang waktu yang diformulasikan pada saat pengambilan *dataset* yaitu sejak dibukanya jendela transfer musim panas 2021/22 di liga Inggris, atau tepatnya pada tanggal 24 Mei 2021 hingga tanggal pengambilan *dataset* pada masing-masing klub.
7. Tanggal dan waktu pada saat pengambilan data yaitu:
  - *Dataset* klub Arsenal: Sabtu, 9 April 2022, 15:23:41 PM.
  - *Dataset* klub Liverpool: Sabtu, 9 April 2022, 16:07:18 PM.
  - *Dataset* klub Manchester United: Sabtu, 9 April 2022, 17:08:27 PM.
8. Jumlah keseluruhan data yang digunakan pada penelitian ini yaitu sebanyak 225.000 *tweet*, yang mana jumlah tersebut terdiri dari 75.000 *tweet* per *dataset*-nya.
9. *Output* dari penelitian tugas akhir ini akan menentukan sentimen pada *dataset* yang belum dilabeli dengan menggunakan model *machine learning* yang telah dirancang sekaligus mengevaluasi hasil dari proses *testing* pada model tersebut menggunakan teknik evaluasi *Confusion Matrix* dan *Classification Report*.

#### 1.6. Kerangka Pemikiran Penelitian

Berikut adalah struktur dari kerangka pemikiran yang menjelaskan secara singkat penelitian ini:



Gambar 1.1 Kerangka pemikiran penelitian

Berdasarkan struktur kerangka pemikiran yang tertera pada gambar di atas, dapat disimpulkan bahwasanya terdapat dua permasalahan pada penelitian ini yang mana di antaranya yaitu “Bagaimana mengimplementasi serta mengetahui kinerja algoritma *Recurrent Neural Network* untuk menganalisis sentimen terhadap opini pendukung sepak bola?” serta “Bagaimana menganalisis hasil sentimen pendukung sepak bola terhadap opini yang dilontarkan di media sosial?”.

Dari permasalahan yang telah dipaparkan di atas, terdapat sebuah peluang dalam menyelesaikan permasalahan tersebut dengan menggunakan teknologi *web scraping* untuk mengumpulkan *tweet* dari media sosial Twitter dan memanfaatkan teknologi *Artificial Intelligence (AI)* beserta *machine learning* sebagai upaya dalam menganalisis serta menentukan sentimen pada sebuah *tweet* secara otomatis.

Sebagai pendekatan terhadap peluang tersebut, penelitian ini akan didukung dengan mengimplementasikan algoritma *Recurrent Neural Network* sebagai upaya dalam merancang sebuah model *machine learning* yang pada akhirnya *result* dalam penelitian ini yaitu terbentuknya suatu model *machine learning* yang kapabel dalam



mengidentifikasi sekaligus menentukan *tweet* positif, negatif, dan juga netral secara otomatis menggunakan algoritma *Recurrent Neural Network*.

## 1.7. Model Penelitian

### 1.7.1. Model Pengerjaan Tugas Akhir

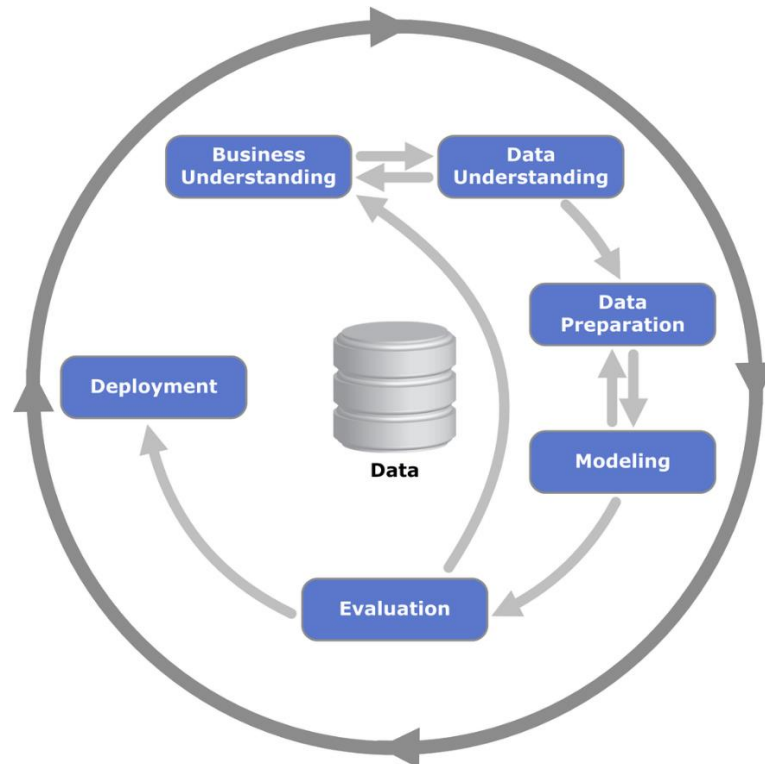
Demi kelancaran dalam pengerjaan tugas akhir ini, terdapat tiga tahapan yang harus dijalankan agar hasil yang diperoleh sesuai dengan yang diharapkan. Tahapan tersebut di antaranya sebagai berikut:

- PERENCANAAN : Menetapkan tujuan serta langkah-langkah yang harus dijalankan dan mengidentifikasi berbagai macam teknologi serta sumber daya yang dapat dimanfaatkan dalam merancang sistem.
- STUDI LITERATUR : Mengumpulkan, mempelajari, memahami, serta mencatat berbagai jenis bahan penelitian seperti informasi, teori, dan prosedur dalam merancang sebuah model *machine learning* yang kapabel dalam menentukan sentimen suatu *tweet* dengan memanfaatkan salah satu turunan dari algoritma *Recurrent Neural Network* yang dikenal sebagai *Long Short-Term Memory* atau LSTM, di mana hal tersebut merupakan inti pembahasan dalam penelitian ini.
- PENGUMPULAN DATA : Mengaplikasikan teknologi *web scraping* dalam mengumpulkan *tweet* dari media sosial Twitter dengan memanfaatkan salah satu *library* bahasa pemrograman Python bernama Snsrape.

### 1.7.2. Model Pengembangan Perangkat Lunak

Metode atau model pengembangan perangkat lunak yang dimanfaatkan dalam membuat sistem pada penelitian ini yaitu model *Cross-Industry Standard Process for Data Mining* atau CRISP-DM. Model tersebut dikembangkan oleh

raturan organisasi dan perusahaan di Eropa untuk dijadikan standar metodologi tanpa pemilik bagi *data mining* [22].



Gambar 1.2 Model CRISP-DM [23]

Terdapat enam tahapan yang dilakukan dalam pelaksanaan CRISP-DM, yang mana urutannya terdiri dari *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan terakhir yaitu *Deployment*, seperti yang ada pada gambar di atas. Dalam gambar tersebut, tanda panah yang berada pada lingkaran menjelaskan bahwa alur CRISP-DM tidak akan berhenti hingga proses *Deployment* saja, melainkan proses tersebut akan terus-menerus mengalir hingga adanya data baru yang digunakan. Berikut adalah penjelasan untuk setiap tahapan pada model CRISP-DM:

*Business Understanding* : Tahap *Business Understanding* adalah sebuah tahapan yang berfokus untuk melakukan *brainstorming* sebagai upaya dalam memahami permasalahan pada penelitian, misalnya seperti menentukan objek beserta peluangnya, menentukan tujuan dan persyaratan yang harus dicapai serta diperlukan selama proses penelitian berlangsung,

- menentukan data yang akan digunakan pada penelitian, dan yang paling penting ialah menentukan rencana serta strategi untuk mencapai tujuan dalam pembuatan model *machine learning*.
- Data Understanding* : Tahap *Data Understanding* diawali dengan melakukan prosedur pengumpulan data yang kemudian dilanjutkan dengan mendeskripsikan dan mengeksplorasi data yang telah diperoleh sebelum akhirnya memverifikasi *weight* atau bobot pada data tersebut.
- Data Preparation* : Tahap *Data Preparation* bertujuan dalam memperbaiki dan menyiapkan data sebagai upaya dalam membangun *dataset* akhir yang nantinya akan digunakan pada tahap *Modeling*.
- Modeling* : Tahap *Modeling* atau pemodelan mengacu pada proses pembuatan model menggunakan algoritma yang sesuai dengan permasalahan pada penelitian.
- Evaluation* : Tahap *Evaluation* atau tahap evaluasi dilakukan dengan tujuan untuk mengevaluasi hasil yang diperoleh setelah melakukan proses *modeling*.
- Deployment* : Tahap *Deployment* meliputi *delivery*, *support*, dan juga *feedback* yang berfokus untuk merepresentasikan hasil serta informasi yang didapat ke dalam sebuah perangkat lunak untuk digunakan oleh *user*.

## 1.8. Sistematika Penulisan

Adapun sistematika penulisan penelitian yang disusun menjadi lima bagian demi memberikan gambaran secara singkat mengenai penelitian yang dilaksanakan dengan urutannya sebagai berikut:

BAB I : PENDAHULUAN

Berisi pemaparan terkait latar belakang penelitian, rumusan masalah penelitian, tujuan serta manfaat penelitian, batasan masalah penelitian, kerangka pemikiran penelitian, metode

atau model penelitian, serta yang terakhir ialah sistematika penulisan pada penelitian.

**BAB II : KAJIAN LITERATUR**

Berisi uraian secara singkat mengenai penelitian terdahulu yang memiliki relevansi terhadap bidang yang diteliti serta terdapat beberapa landasan teori guna menjelaskan berbagai macam konsep dalam penelitian ini.

**BAB III : METODOLOGI PENELITIAN**

Berisi pemaparan mengenai prosedur atau langkah-langkah dalam menjalankan proses penelitian yang dijelaskan secara kronologis serta sistematis. Umumnya, uraian dalam bab ini mengacu pada model pengembangan perangkat lunak yang sesuai dengan kebutuhan dan karakteristik penelitian.

**BAB IV : HASIL DAN PEMBAHASAN**

Berisi uraian mengenai temuan atau hasil penelitian berupa gambar, diagram, atau tabel yang diperoleh melalui tahapan penelitian. Pada akhir bab, terdapat pembahasan mengenai temuan tersebut yang disesuaikan dengan rumusan masalah dalam penelitian ini.

**BAB V : SIMPULAN DAN SARAN**

Berisi uraian mengenai simpulan yang merupakan jawaban untuk rumusan masalah penelitian serta terdapat saran yang dipusatkan pada harapan peneliti dalam menyempurnakan penelitian.