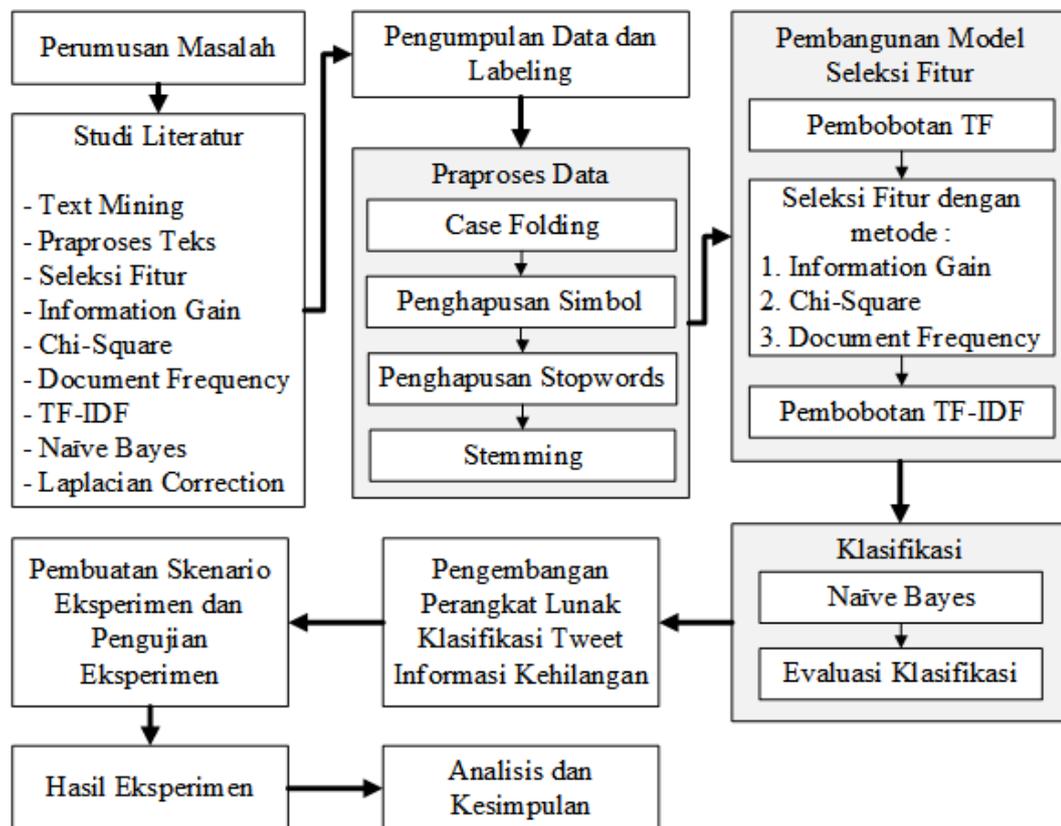


## BAB III METODOLOGI PENELITIAN

### 3.1 Desain Penelitian

Desain penelitian merupakan kerangka kerja yang digunakan dalam penelitian. Pada bagian ini akan dipaparkan kerangka kerja dari mulai penelitian sampai selesai. Desain penelitian yang digunakan dapat dilihat pada gambar 3.1.



Gambar 3.1 Desain Penelitian

Pada Gambar 3.1 di atas menggambarkan secara garis besar kerangka kerja yang akan digunakan oleh penulis pada penelitian ini. Penjelasan lebih lengkap akan dijabarkan sebagai berikut.

#### 3.1.1 Perumusan Masalah

Pada tahap ini merupakan awal alur penelitian dengan merumuskan masalah mengenai kategorisasi informasi kehilangan pada media sosial Twitter yang bisa

terbagi dalam beberapa kategori. Selain itu merumuskan juga permasalahan terkait reduksi fitur dalam data teks yang memiliki banyak sekali fitur. Setelah perumusan selesai, maka selanjutnya penelitian dapat dimulai dengan mempelajari beberapa literatur yang terkait dengan permasalahan dalam penelitian dan mulai mengumpulkan data.

### 3.1.2 Studi Literatur

Pada tahap ini, penulis melakukan pengumpulan dan pemahaman berbagai teori dan penghitungan serta pembahasan yang terkait dengan penelitian. Sumber literatur berasal dari berbagai sumber seperti jurnal, buku, video dan artikel penelitian untuk dijadikan referensi. Adapun subjek teori yang didapat untuk dipahami adalah mengenai *preprocessing text*, *feature selection* menggunakan metode Information Gain, Chi-Square, dan Document Frequency, pembobotan TF-IDF, klasifikasi dengan metode Naïve Bayes beserta teknik *Laplacian Correction*, dan penelitian terkait yang telah dilakukan sebelumnya.

### 3.1.3 Pengumpulan Data dan *Labeling*

Pada tahap ini dilakukan pengumpulan data yang akan digunakan sebagai dataset. Data yang digunakan berasal dari media sosial Twitter dengan kata kunci “kehilangan”, “hilang”, “ditemukan”, “menemukan”, “pencurian”, dan “dicuri”. Proses pengumpulan data dilakukan dengan bantuan program Scweet milik Altimis yang dapat diakses pada link <https://github.com/Altimis/Scweet>. Setelah data terkumpul, selanjutnya dilakukan pelabelan kelas atau kategori secara manual.

### 3.1.4 Praproses Data

Data yang sudah terkumpul adalah data berbentuk teks, di mana data tersebut merupakan data tidak terstruktur. Agar bisa diolah secara komputasi, maka diperlukan praproses data terlebih dahulu supaya menjadi data yang terstruktur. Tahapan praproses data meliputi *case folding*, penghapusan simbol, penghapusan *stopwords*, dan *stemming*. Tahapan ini juga berfungsi untuk membersihkan data dari simbol, karakter, dan kata yang tidak diperlukan.

### 3.1.5 Pembangunan Model Seleksi Fitur

Setelah data dipraproses, selanjutnya dilakukan tahapan pembangunan model untuk seleksi fitur. Tahap ini terdiri dari tiga proses, yaitu pembobotan TF, pembangunan seleksi fitur, dan pembobotan TFIDF. Proses pembobotan TF akan mengubah bentuk data menjadi berbentuk matriks TF berdasarkan frekuensi kemunculan kata. Kemudian untuk proses pembangunan seleksi fitur digunakan tiga metode yaitu Information Gain, Chi-Square, dan Document Frequency. Metode Information Gain dan Chi-Square akan menggunakan rumus penghitungan tertentu, sedangkan Document Frequency hanya menggunakan jumlah dokumen yang mengandung fitur yang dicari. Dengan pembangunan model seleksi fitur ini maka akan mereduksi jumlah fitur dengan harapan bisa meningkatkan performa klasifikasi dan kinerja komputasi. Setelah model seleksi fitur dibuat, selanjutnya dilakukan pembobotan untuk setiap fitur menggunakan metode TFIDF.

### 3.1.6 Klasifikasi

Pada tahap ini dilakukan pembangunan model klasifikasi dengan menggunakan metode Naïve Bayes. Data yang digunakan dalam klasifikasi ini adalah data yang sudah dilakukan seleksi fitur dan yang tidak dilakukan seleksi fitur. Hal ini dilakukan sebagai eksperimen dan dianalisis hasilnya kemudian seperti apa. Kemudian setelah model klasifikasi dibuat, selanjutnya dilakukan evaluasi, apakah proses dalam klasifikasinya sudah sesuai dan apakah hasilnya sudah optimal.

### 3.1.7 Pengembangan Perangkat Lunak Klasifikasi Informasi Kehilangan

Pengembangan perangkat lunak pada penelitian didasarkan pada tahapan-tahapan sebelumnya, yaitu praproses data, seleksi fitur, dan klasifikasi. Pengembangan pada tahap ini menggunakan model *Waterfall* (Sommerville, 2010). Proses pengembangan menggunakan Bahasa pemrograman Python 3 dan Jupyter Notebook.

### 3.1.8 Pembuatan Skenario Eksperimen dan Pengujian Eksperimen

Pada tahap ini perangkat lunak sudah selesai dikembangkan dan akan dibuat skenario eksperimen. Tujuannya yaitu untuk menguji model yang telah dibuat serta untuk mendapatkan hasil yang beragam. Skenario yang akan dilakukan antara lain sebagai berikut.

Tabel 3.1 Skenario Eksperimen Penelitian

Eksperimen	Feature Selection			Pembobotan	
	IG	X <sup>2</sup>	DF	TF	TF-IDF
1				✓	
2					✓
3	✓			✓	
4	✓				✓
5		✓		✓	
6		✓			✓
7			✓	✓	
8			✓		✓

Tabel 3.1 merupakan gambaran dari skenario eksperimen yang akan dilakukan pada penelitian ini. Secara garis besar, eksperimen terbagi menjadi 2 tahap, yaitu pemilihan fitur yang akan digunakan dan pembobotan dari fitur tersebut. Metode *feature selection* yang digunakan adalah Information Gain, Chi-Square, dan Document Frequency. Batasan pemilihan fitur yang digunakan adalah sebesar 20%, 40%, 60%, dan 80% dari total keseluruhan fitur. Kemudian untuk pembobotan fitur dilakukan dengan menghitung *term frequency* (TF) dan TF-IDF. Pengujian akan dilakukan untuk klasifikasi 2 dan 4 kelas. Klasifikasi 2 kelas terdiri dari kelas ‘Informasi Kehilangan’, dan ‘Bukan Informasi Kehilangan’. Sedangkan klasifikasi 4 kelas terbagi atas kelas ‘Kehilangan Barang dan Dokumen’,

‘Kehilangan Kendaraan’, ‘Kehilangan Orang dan Hewan Peliharaan’, dan ‘Tidak Diketahui – Tidak Relevan’.

### **3.1.9 Hasil Eksperimen**

Hasil eksperimen pada skenario yang telah dilakukan akan dicatat dan dikumpulkan. Hasil ini akan dilakukan analisis pada tahap berikutnya.

### **3.1.10 Analisis dan Kesimpulan**

Pada tahap ini hasil dari eksperimen dilakukan analisis. Mengenai hasil dari skenario dibahas pada tahap ini. Hasil evaluasi ditampilkan ke dalam sebuah tabel dan grafik untuk mempermudah analisis. Dari beberapa analisis yang dilakukan, akan dibuat menjadi kesimpulan.

## **3.2 Metode Penelitian**

Metode penelitian yang dilakukan dalam melakukan penelitian ini terbagi ke dalam dua bagian, meliputi metode yang digunakan dalam proses pengumpulan data serta metode yang digunakan dalam proses pengembangan perangkat lunak. Penjelasan mengenai kedua tahapan metode tersebut akan dijelaskan sebagai berikut.

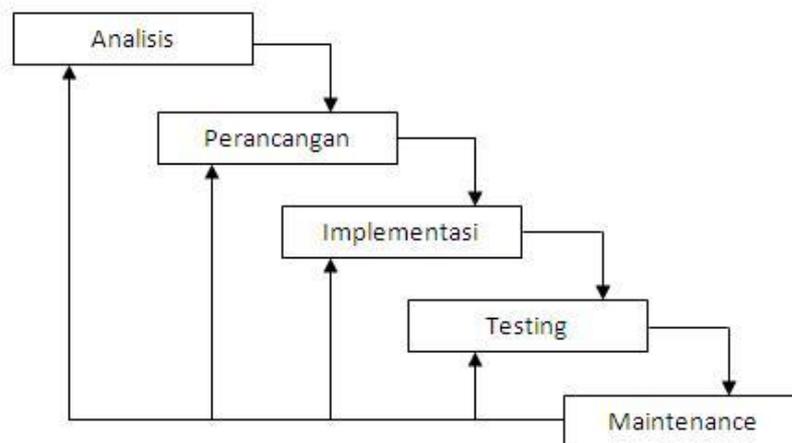
### **3.2.1 Metode Pengumpulan Data**

Metode pengumpulan data pada penelitian ini dilakukan dengan cara *scraping* data Twitter dengan kata kunci “kehilangan”, “hilang”, “ditemukan”, “menemukan”, “pencurian”, dan “dicuri”. Proses pengumpulan data dilakukan dengan bantuan program Scweet milik Altimis yang dapat diakses pada link <https://github.com/Altimis/Scweet>. Data yang didapat berupa data teks yang tidak terstruktur. Kemudian dilakukan pelabelan secara manual.

### **3.2.2 Metode Pengembangan Perangkat Lunak**

Pengembangan perangkat lunak dalam penelitian ini dibangun menggunakan metode *Waterfall*. Model ini membangun perangkat lunak secara

bertahap atau sekuensial. Menurut Sommerville (2010) pada gambar 3.2, model ini terdiri dari tahapan analisis, perancangan, implementasi, dan pengujian.



Gambar 3.2 Model *Waterfall* (Sommerville, 2010)

Pada model *waterfall*, pembangunan dilakukan secara sekuensial sehingga dapat terjadi kemungkinan untuk kembali ke tahap sebelumnya jika saat berada pada aktivitas tertentu membutuhkan perbaikan. Berikut ini adalah penjelasan mengenai model *waterfall*.

### 1. Analisis

Analisis merupakan tahapan awal dalam pengembangan perangkat lunak. Pada tahap ini dilakukan pengumpulan data informasi mengenai latar belakang mengapa perangkat lunak perlu dibangun. Kemudian dilakukan juga analisis terhadap data yang akan digunakan. Pada tahap ini pula ditentukan spesifikasi dari perangkat lunak untuk mencapai tujuan dari dibuatnya perangkat lunak.

### 2. Perancangan

Pada tahap ini perangkat lunak dilakukan perancangan agar dapat berjalan sesuai dengan hasil analisis yang telah dilakukan sebelumnya. Tahapan yang dilakukan diantaranya adalah membuat desain alur dari perangkat lunak, mulai dari pengumpulan data, *training*, dan *testing*.

### 3. Implementasi

Pada tahap ini adalah implementasi hasil perancangan ke dalam perangkat lunak dengan diterjemahkan ke bahasa yang dapat dibaca oleh mesin komputer. Implementasi pembuatan perangkat lunak dilakukan bertahap sesuai dengan alur pada tahap perancangan sebelumnya.

### 4. Testing

Pada tahap ini dilakukan proses pengujian terhadap perangkat lunak yang telah selesai dibuat. Pengujian dilakukan untuk memastikan bahwa perangkat lunak yang dibangun dapat menjalankan perintah dengan baik dan sesuai dengan skenario.

### 5. Maintenance

Tahap Ini merupakan tahap akhir pada pengembangan perangkat lunak Di mana perangkat lunak mengalami perubahan dengan tujuan pemeliharaan. Perubahan terjadi karena perangkat lunak mengalami kesalahan atau untuk menyesuaikan kembali dengan kebutuhan.

## 3.3 Perangkat dan Data Penelitian

Dalam melakukan penelitian, penulis menggunakan alat dan bahan yang meliputi perangkat keras, perangkat lunak, bahan kajian beserta sejumlah data yang akan dijelaskan sebagai berikut.

### 3.3.1 Perangkat Penelitian

Perangkat yang digunakan dalam penelitian ini meliputi perangkat keras dan perangkat lunak. Berikut adalah daftar alat yang digunakan dalam penelitian.

Perangkat keras yaitu komputer dengan spesifikasi:

- Processor Intel® Core™ i5-3360M CPU @ 2.80GHz (4 CPUs), ~2.8GHz
- Memori 12288MB RAM
- Intel® HD Graphics 4000
- SSD 240GB
- HDD 500GB

Perangkat lunak yang digunakan sebagai berikut:

- Sistem Operasi Windows 10 Pro 64-bit
- Browser Google Chrome
- Python 3.6.8
- Jupyter Notebook
- Anaconda

### **3.3.2 Data Penelitian**

Data penelitian yang digunakan dalam penelitian ini terdiri atas tweet informasi kehilangan yang diambil melalui platform media sosial Twitter. Data diambil dalam jangka waktu 3 tahun mulai dari Januari 2019 sampai Desember 2021.