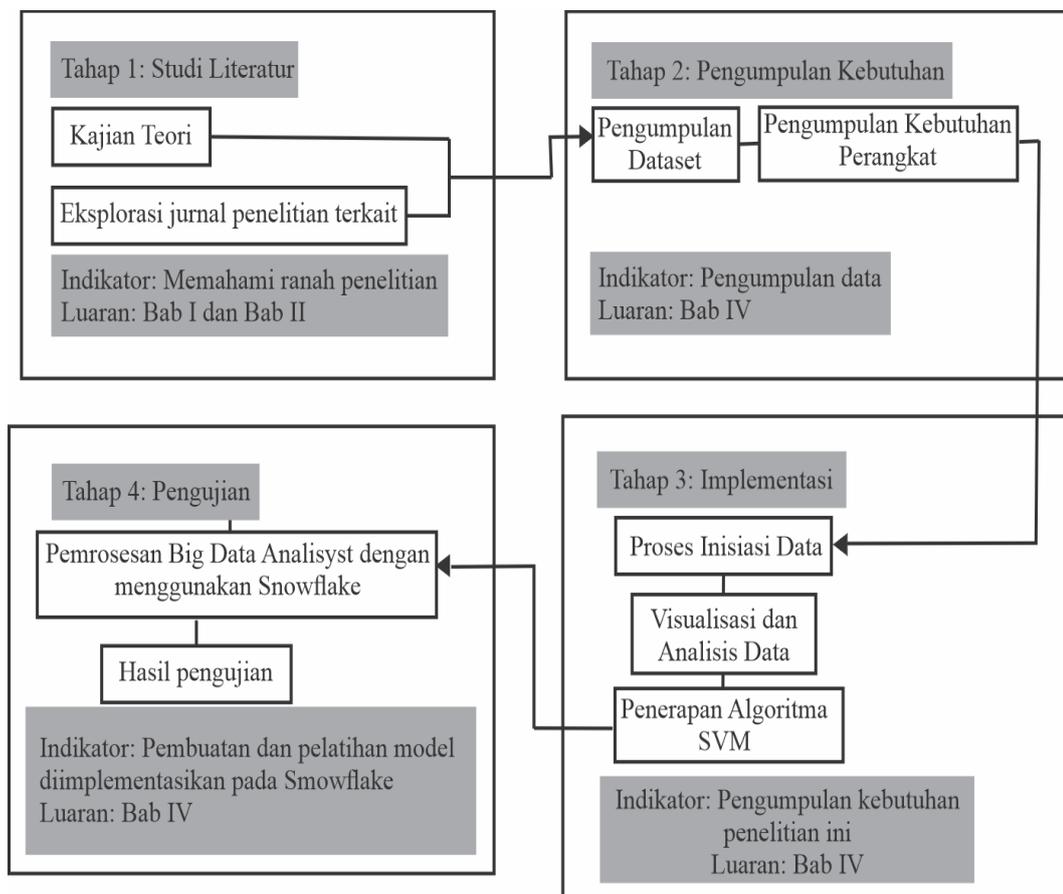


### BAB III

#### METODOLOGI

Alur metode penelitian yang digunakan, dimulai dari hasil studi literatur, pengumpulan data, perancangan program, dan implementasi program, dan Kesimpulan. Diagram alur penelitian dapat dilihat pada gambar 3.1.



Gambar 3. 1 Alur diagram penelitian

### **3.1 Studi Literatur**

Studi pustaka dan literatur yang dilakukan, dipelajari, dan dikaji dalam menyelesaikan penelitian ini diambil dari buku, jurnal, internet, serta dokumen-dokumen yang berkaitan dengan penelitian. Jurnal-jurnal yang berkaitan dengan penelitian *Big Data Analyst*, dikaji secara mendalam dan membandingkan antar literatur hasil penelitian sebelumnya, dicari kelemahannya, serta metode-metode yang digunakan dalam penelitian pada jurnal tersebut.

### **3.2 Pengumpulan Kebutuhan**

#### **3.2.1 Pengumpulan Dataset**

Data yang digunakan dalam penelitian ini adalah dataset mengenai *Human Activity Recognition* yang tersedia pada repositori *UCI Machine Learning*(Aksesdataset:<https://archive.ics.uci.edu/ml/machine-learning-databases/00240/>).

#### **3.2.2 Pengumpulan Kebutuhan Perangkat**

Kebutuhan berupa perangkat keras dan perangkat lunak yang digunakan dalam penelitian ini dijabarkan dibawah ini:

1. Perangkat keras pengolah data dengan spesifikasi optimal. Tabel perangkat keras pengolah data yang digunakan pada penelitian ini dapat dilihat pada tabel 3.1.

Tabel 3. 1 Perangkat keras pengolah data

<i>Hardware</i>	<b>Spesifikasi</b>
Prosesor	Intel® Core™ i5-9300H Processor (8M Cache, up to 4.10 GHz)
RAM	16192MB RAM DDR4
Grafis	GeForce® GTX 1050 Ti ,GDDR5 4GB.

2. Perangkat lunak pengolah data yang dipakai dalam proses analisis data dan implementasi snowflake *cloud data warehousing*. Tabel perangkat lunak pengolah data yang digunakan pada penelitian ini dapat dilihat pada tabel 3.2.

Tabel 3. 2 Perangkat lunak pengolah data

<i>Software</i>	<b>Fungsi</b>
Google Colab	Digunakan untuk menulis <i>code</i> python dalam proses analisis data. Pada google colab sudah tersedia berbagai macam <i>library</i> yang akan digunakan dalam penelitian ini.
Google Drive	Digunakan untuk menyimpan dataset yang akan digunakan dalam penelitian ini, serta menyimpan data hasil pengolahan data yang dilakukan.

Lanjutan tabel 3.2

Snowflake	<p>Digunakan untuk menyimpan semua data berbasis <i>cloud warehousing</i> yang sudah diolah terlebih dahulu menjadi sebuah data sempurna yang selanjutnya akan dianalisis menggunakan <i>Machine Learning</i> dengan algoritma <i>Support Vector Machine</i>.</p>
-----------	---

### 3.3 Implementasi

Implementasi dilakukan setelah proses pengumpulan semua kebutuhan yang telah disusun. Pembuatan dan pelatihan model diimplementasikan pada google colab dengan bantuan berbagai *Library* dan menghubungkannya ke Snowflake *Cloud*. Berbagai proses dilakukan dalam penelitian ini, seperti proses analisis data dengan menggunakan SVM sebagai algoritma yang digunakan untuk proses *Big Data Analyst* pada penelitian ini.

#### 3.3.1 Proses Inisialisasi Data

Proses inisialisasi data dimulai dengan memasukan data pada google drive dan *mounting* menggunakan google colab sehingga bisa diolah dan dianalisis.

#### 3.3.2 Visualisasi dan Analisis Data

Visualisasi dilakukan setelah proses memasukan data pada google colab, visualisasi ini berupa grafik dari suatu keberagaman data. Visualisasi ini bermanfaat agar memudahkan dalam membaca data, mengecek kekosongan

dan duplikasi data, dan menyatukan dataset menjadi satu data yang akan diuji pada Snowflake.

### **3.3.3 Penerapan Algoritma Support Vector Machine (SVM)**

Proses penerapan algoritma SVM dilakukan setelah semua proses sebelumnya berhasil. Proses ini dilakukan untuk *Big Data Analysisist* dengan mengukur tingkat akurasi dari proses komputasi yang berlangsung.

## **3.4 Pengujian**

Pengujian dilakukan untuk mengetahui kemampuan dari Snowflake dengan menggunakan algoritma SVM pada proses *Big Data Analysisist* pada data data berskala besar. Selain itu pengujian akhir penggunaan Snowflake yaitu membangun *data warehouse* sederhana secara *cloud*. Penggunaan algoritma SVM pada Snowflake yaitu untuk mengukur tingkat akurasi pada proses *Big Data Analysisist*.