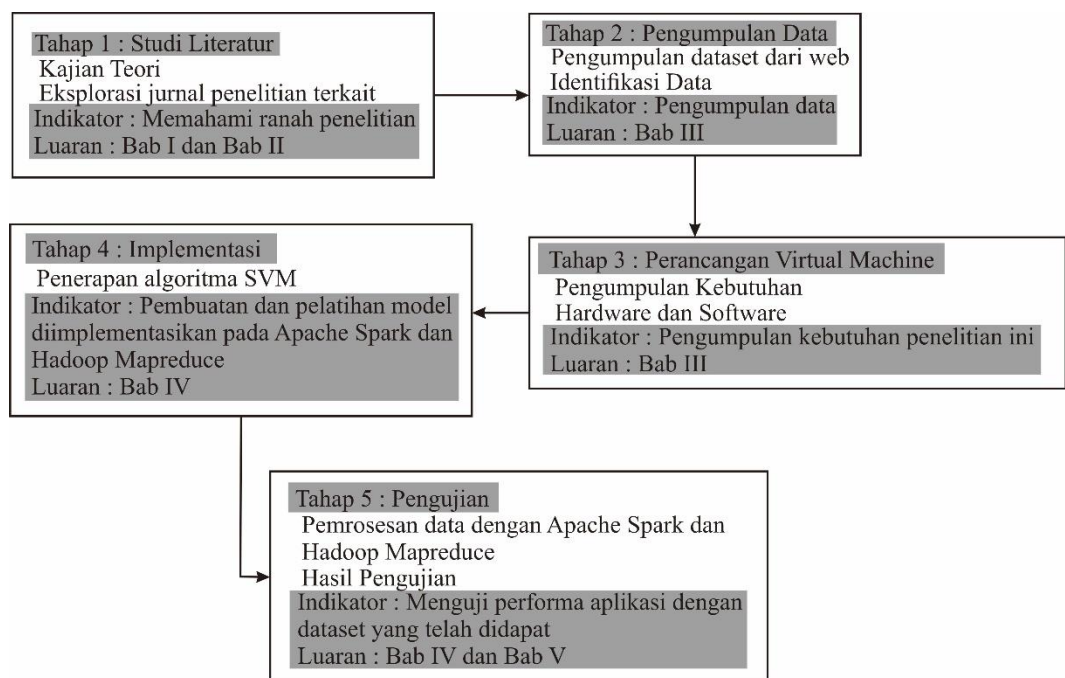


## BAB III

### METODOLOGI

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



Gambar 3.1 Diagram Alur 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 Analysis*, 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 *Mobile Price Classification* yang tersedia pada repositori *Kaggle* (Akses dataset: <https://www.kaggle.com/datasets/iabhishekoofficial/mobile-price-classification>).

#### **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 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	16192 MB RAM DDR4
Grafis	GeForce® GTX 1050 Ti ,GDDR5 4GB.

2. Perangkat lunak pengolah data yang dipakai dalam proses analisis perbandingan performa *Apache Spark* dan *Hadoop Mapreduce*. 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>
VMWare	Digunakan untuk membuat virtual <i>Operating System</i>
Linux Ubuntu	Sebagai <i>Operating System</i> yang digunakan
Apache Spark	Digunakan untuk penyimpanan dataset yang selanjutnya akan diproses menggunakan <i>Machine Learning</i> dengan algoritma <i>Support Vector Machine</i> dan nantinya menjadi bahan perbandingan dengan <i>Apache Hadoop</i> .

Lanjutan Tabel 3.2

Apache Hadoop	Digunakan untuk penyimpanan dataset yang selanjutnya akan diproses menggunakan <i>Machine Learning</i> dengan algoritma <i>Support Vector Machine</i> dan nantinya menjadi bahan perbandingan dengan <i>Apache Spark</i> .
Jupyter Notebook	Digunakan untuk menulis <i>code python</i> dalam proses analisis data.

### 3.3. Implementasi

Implementasi dibuat untuk rancangan program yang telah disusun. Pembuatan dan pelatihan model diimplementasikan pada *Apache Spark* dan *Hadoop Mapreduce* seperti proses analisis data dengan menggunakan *Support Vector Machine (SVM)* sebagai algoritma yang digunakan untuk proses *Big Data Analysis* pada penelitian ini..

#### 3.3.1. Proses Inisialisasi Data

Proses inisialisasi data dimulai dengan memasukan data pada *Apache Spark* dan *Hadoop Mapreduce* yang telah diperoleh dari website.

### **3.3.2. Visualisasi dan Analisis Data**

Visualisasi dilakukan setelah proses memasukan data pada *Apache Spark* dan *Hadoop Mapreduce*, visualisasi ini berupa grafik dari suatu keberagaman data. Visualisasi ini bermanfaat agar memudahkan dalam membaca data.

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

Proses penerapan algoritma *Support Vector Machine (SVM)* dilakukan setelah semua proses sebelumnya berhasil. Proses ini dilakukan untuk *Big Data Analysis* dengan mengukur tingkat akurasi dari proses komputasi yang berlangsung.

## **3.4. Pengujian**

Pengujian dilakukan untuk mengetahui kemampuan dari *Apache Spark* dan *Hadoop Mapreduce* dengan menggunakan algoritma *Support Vector Machine (SVM)* pada proses *Big Data Analysis* pada data berskala besar. Penggunaan algoritma *Support Vector Machine (SVM)* pada *Apache Spark* dan *Hadoop Mapreduce* yaitu untuk mengukur tingkat akurasi pada proses *Big Data Analysis*.