

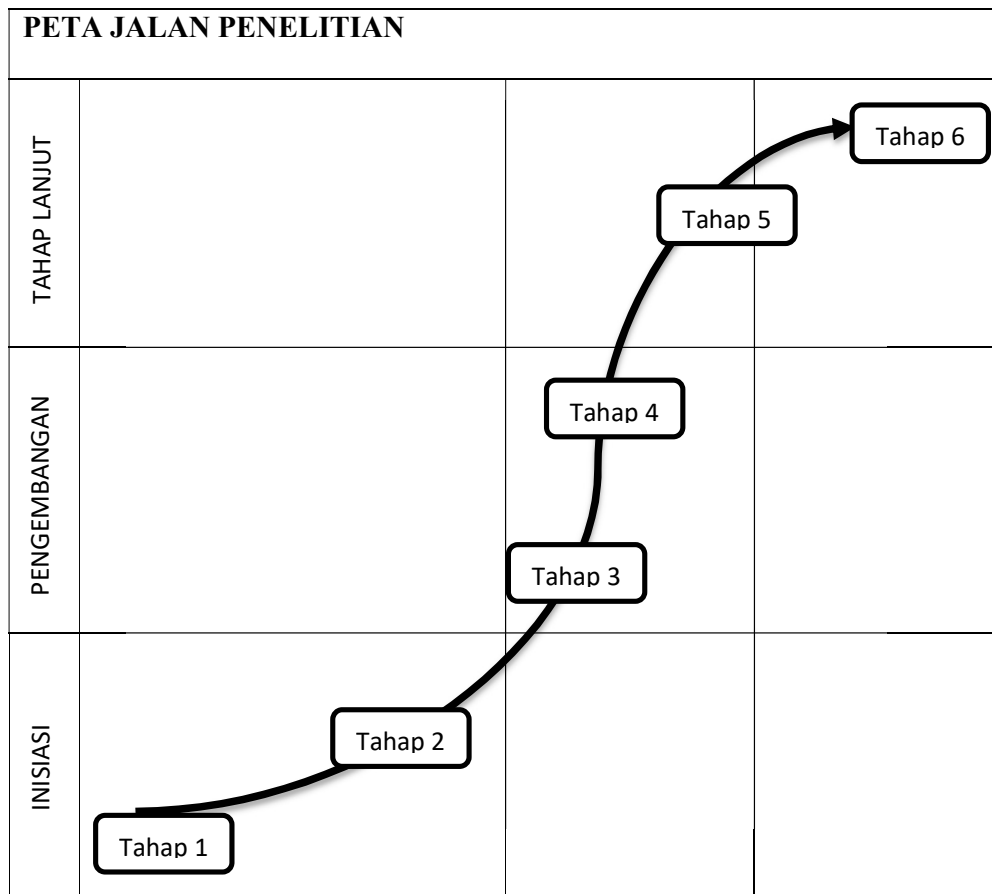
BAB III
METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

3.1.1 Peta Jalan (*Roadmap*) Penelitian

Roadmap merupakan panduan strategis yang menggambarkan arah dan tujuan dari penelitian ini. Dalam *roadmap* penelitian ini, terdapat tiga tahapan utama: tahap inisiasi sebagai langkah awal, diikuti oleh tahap pengembangan untuk memperluas dan menyempurnakan penelitian, serta tahap lanjutan sebagai tahap akhir yang mengarah pada penerapan atau penyempurnaan hasil penelitian.

Tabel 3. 1 Peta Jalan Penelitian



Keterangan :

Tahap 1 : Studi Literatur

Tahap 2 : Pengumpulan dan Analisis Data

Tahap 3 : Pengolahan Data (*preprocessing data*) dan implementasi Teknik *SMOTE*

Tahap 4 : Pemodelan Algoritma *K-Nearest Neighbor* dan *Adaptive boosting*

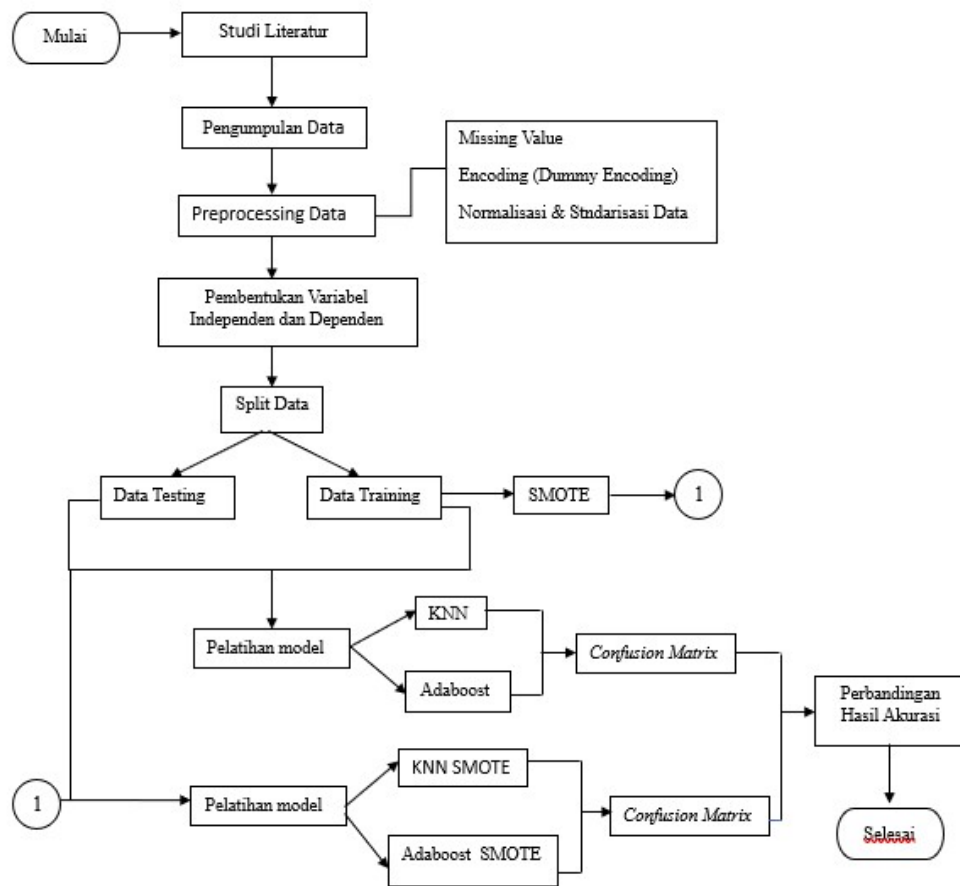
Tahap 5 : Evaluasi Model Algoritma *K-Nearest Neighbor* dan *Adaptive boosting*

Tahap 6 : Penarikan Kesimpulan

3.1.2 Tahapan Penelitian

Proses ini divisualisasikan dengan bentuk diagram alur untuk mempermudah pemahaman terhadap tahapan-tahapan yang dilalui. Penelitian ini menggunakan pendekatan metode kuantitatif. Penyusunan diagram alur bertujuan untuk menyampaikan informasi secara sistematis mengenai tahapan penelitian. Inovasi utama yang diusung pada penelitian adalah perbandingan dua metode klasifikasi, yakni *K-Nearest Neighbor* dan *Adaptive Boosting*, dalam memprediksi penyakit stroke dengan menerapkan teknik *SMOTE* guna mengatasi permasalahan ketidakseimbangan data. Keseluruhan proses akan direalisasikan melalui eksperimen dan observasi untuk mengidentifikasi algoritma yang memberikan hasil paling optimal.

Gambar 1 menyajikan tahapan yang akan dilakukan pada penelitian ini, mulai dari studi literatur, dilanjutkan dengan pengumpulan dan analisis data, analisis terhadap metode *adaptive boosting* dan algoritma *K-nearest neighbor*, pengujian metode yang digunakan, hingga tahap akhir berupa penarikan kesimpulan.



Gambar 3. 1 Tahapan Penelitian

3.1.3 Studi Literatur

Tahap ini yaitu mengumpulkan data serta sumber yang berhubungan dengan algoritma *adaboost* dan algoritma *k-nearest neighbor* serta penyakit stroke. Data dan referensi yang digunakan pada penelitian ini diperoleh dari jurnal ilmiah dan e-proceeding yang relevan dengan algoritma yang digunakan, serta didukung oleh literatur dari buku dan sumber daring (internet) yang berkaitan dengan objek penelitian.

3.1.4 Pengumpulan dan Analisis Data

Dari penelitian ini dibutuhkan data yang akan digunakan untuk menguji algoritma, Pada penelitian ini membutuhkan *dataset stroke disease*, *dataset* ini bersumber dari *kaggle* yang merupakan *dataset public*. Penelitian ini menggunakan data sekunder karena sumber *dataset* adalah dari *kaggle*. *Dataset stroke disease* memiliki 11 atribut yang terdiri dari 3 atribut numerik dan 7 atribut nominal atau *categorical* yang berfungsi sebagai prediktor serta memiliki 1 atribut tujuan yang bertujuan untuk memberikan output untuk membedakan hasil klasifikasi yang bernilai (1) untuk klasifikasi “*stroke*” serta nilai (0) untuk klasifikasi “*no stroke*”.

Dataset stroke disease masih terdapat *missing value* didalamnya, maka harus dilakukan pengolahan data terlebih dahulu agar data menjadi *valid* dan berkualitas maka harus dilakukan proses *preprocessing* data.

3.1.5 Pembagian Data dan Penerapan Teknik *SMOTE*

Tahapan ini melakukan pembagian data latih dan data uji yaitu 80%:20% dan menerapkan teknik *SMOTE* untuk mengatasi ketidakseimbangan pada data *stroke*, dimana *dataset* ini memiliki kelas mayor dan kelas minor. untuk kelas minor ini yaitu *stroke* dan kelas mayor yaitu *no stroke*, dengan diterapkannya teknik ini akan membuat data sintesis untuk kelas minor agar menjadi seimbang dengan kelas mayor, sehingga akurasi yang dihasilkan akan lebih akurat.

3.1.6 Pelatihan Model Algoritma *K-Nearest Neighbor* dan *Adaptive boosting*

Pada tahap ini, data yang telah dibagi menjadi data latih dan data uji, termasuk data uji yang telah melalui proses penerapan teknik *SMOTE*, digunakan untuk melatih model. Tujuannya adalah untuk membandingkan performa dua

algoritma klasifikasi, yaitu *KNN* dan *Adaptive Boosting*, guna menentukan algoritma yang mendapatkan nilai akurasi terbaik.

3.1.7 Pengujian Model

Untuk mengukur kinerja algoritma KNN (*K-Nearest Neighbor*) dan *adaptive boosting* akan dilakukan dengan menggunakan hasil perhitungan dari *confusion matrix*.

3.1.8 Penarikan Kesimpulan

Penarikan kesimpulan merupakan tahap akhir dalam penelitian ini. Hasil yang diperoleh berupa analisis perbandingan antara dua algoritma klasifikasi, yaitu knn (*K-Nearest Neighbor*) dan *Adaptive Boosting*, dalam memprediksi penyakit stroke. Selain itu, dilakukan pula evaluasi terhadap perbedaan performa model berdasarkan nilai akurasi, baik dengan penerapan teknik SMOTE maupun tanpa penerapan teknik SMOTE, untuk mengatasi ketidakseimbangan data.