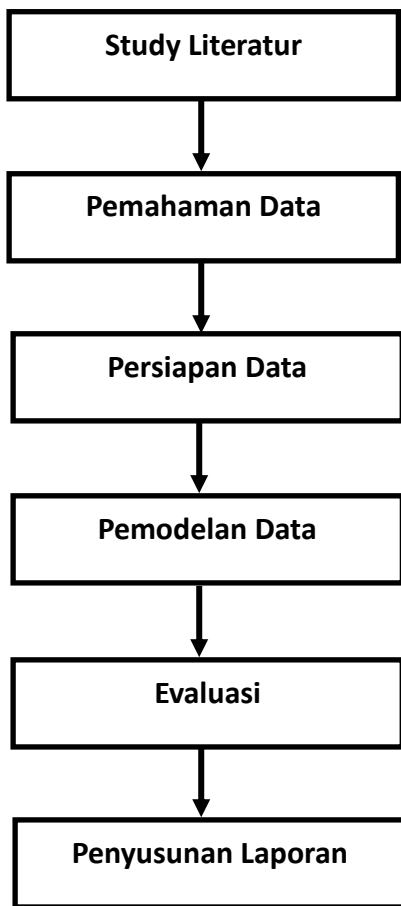


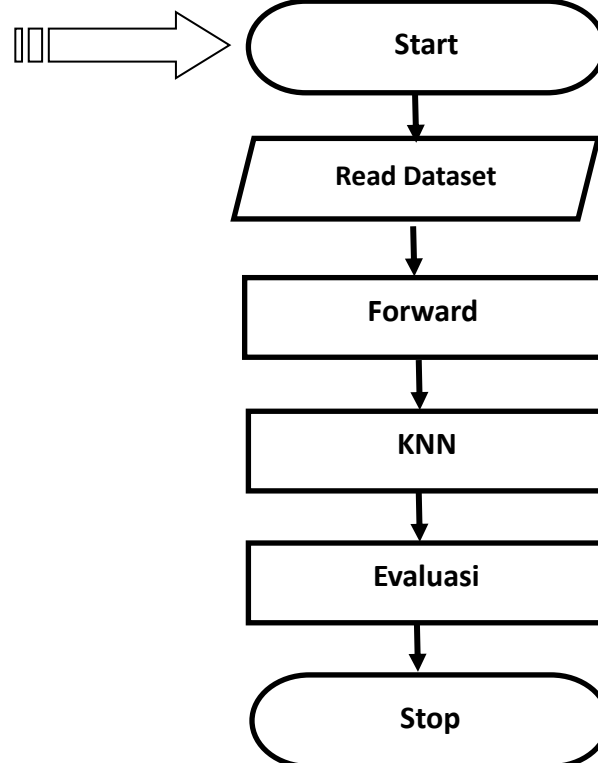
BAB III METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset dalam memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan yang diinginkan maka dibuat alur dalam tahapan penelitian yang akan dilakukan berikut:



Gambar 3.1. Flowchart Tahapan Penelitian



Gambar 3.2. Flowchart Tahapan Pemodelan

Berikut ini merupakan penjelasan dari gambar 3.1 yang merupakan alur dari penelitian ini :

1. Tahap Pemahaman bisnis juga dapat disebut fase pemahaman penelitian yang meliputi tujuan dan persyaratan proyek dengan jelas dalam hal bisnis atau unit penelitian secara keseluruhan, menterjemahkan tujuan dan batasan ke dalam perumusan definisi masalah data mining, menyiapkan strategi awal untuk mencapai tujuan tersebut.
2. Tahap pemahaman data dengan melakukan pengumpulan data menggunakan analisis data eksplorasi untuk membiasakan diri dengan data dan menemukan wawasan awal, mengevaluasi kualitas data, apabila diinginkan pilih subset menarik yang mungkin berisi pola yang dapat ditindaklanjuti.
3. Tahap persiapan data adalah mempersiapkan data mentah awal kumpulan data akhir yang akan digunakan untuk semua tahapan berikutnya. Pada tahapan ini memilih kasus dan variabel yang diinginkan, menganalisis yang sesuai untuk analisis. Melakukan transformasi pada variabel tertentu, jika diperlukan, membersihkan data mentah sehingga siap untuk alat pemodelan. Data pada penelitian ini terdiri dari 858 record, 36 atribut.
4. Tahap pemodelan, Tahapan ini dilakukan dengan memilih dan menerapkan teknik pemodelan yang sesuai. kalibrasi pengaturan model untuk mengoptimalkan hasil, beberapa teknik berbeda dapat digunakan untuk masalah data mining yang sama, apabila diperlukan mengulangi kembali ke tahapan persiapan data untuk membawa bentuk data sesuai dengan persyaratan khusus dari teknik data mining tertentu. Pemodelan pada penelitian ini menggunakan teknik data mining yaitu algoritma KNN dan Forward Selection.
5. Tahap selanjutnya evaluasi, yaitu tahapan penelitian untuk mengukur kinerja model prediksi yang telah dibangun. Evaluasi memberikan

wawasan tentang sejauh mana model lebih efektif dalam melakukan prediksi pada data yang tidak pernah dilihat sebelumnya.

- Tahap terakhir yaitu penyusunan laporan berupa laporan dari hasil pengujian yang mengacu pada rumusan masalah serta tujuan penelitian.

3.2 Tahapan Pemodelan Data

tahap awal data yang ada akan dilakukan pemodelan dengan algoritma Forward Selection, yaitu dengan menggunakan model-model basis yang telah ditraining untuk melakukan klasifikasi pada testing dataset

1. Read Dataset

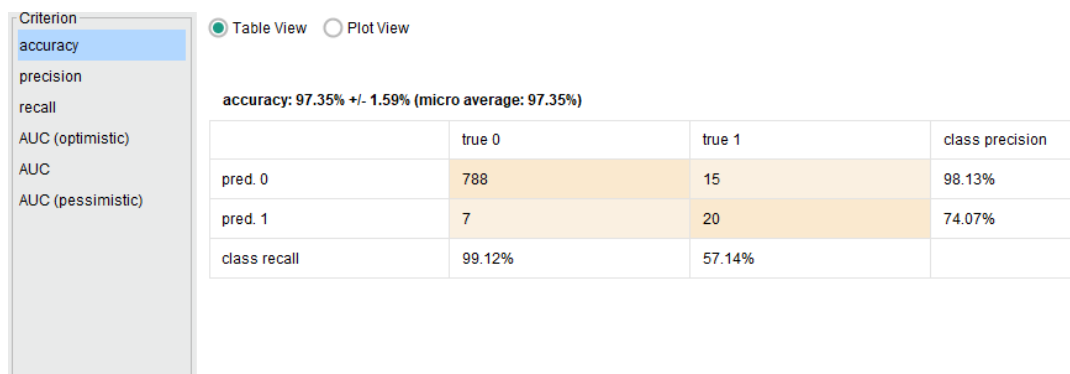
TIPE DATA PADA TIAP KOLOM PADA DATASET					
Kolom	Tipe Data	Kolom	Tipe Data	Kolom	Tipe Data
<i>Age</i>	<i>Integer</i>	<i>STDs (number)</i>	<i>Boolean</i>	<i>STDs: HPV</i>	<i>Boolean</i>
<i>Number of sexual partners</i>	<i>Integer</i>	<i>STDs: condylomatosis</i>	<i>Boolean</i>	<i>STDs: Number of diagnosis</i>	<i>Integer</i>
<i>First sexual intercourse</i>	<i>Integer</i>	<i>STDs: cervical condylomatosis</i>	<i>Boolean</i>	<i>STDs: Time since first diagnosis</i>	<i>Integer</i>
<i>Number of pregnancies</i>	<i>Integer</i>	<i>STDs: vaginal condylomatosis</i>	<i>Boolean</i>	<i>STDs: Time since last diagnosis</i>	<i>Integer</i>
<i>Smokes</i>	<i>Boolean</i>	<i>STDs: vulvo-perineal condylomatosis</i>	<i>Boolean</i>	<i>Dx: Cancer</i>	<i>Boolean</i>
<i>Smokes (years)</i>	<i>Integer</i>	<i>STDs: syphilis</i>	<i>Boolean</i>	<i>Dx: CIN</i>	<i>Boolean</i>
<i>Smokes (packs/year)</i>	<i>Boolean</i>	<i>STDs: pelvic inflammatory disease</i>	<i>Boolean</i>	<i>Dx: HPV</i>	<i>Boolean</i>
<i>Hormonal Contraceptives</i>	<i>Boolean</i>	<i>STDs: genital herpes</i>	<i>Boolean</i>	<i>Dx</i>	<i>Boolean</i>
<i>Hormonal Contraceptives (years)</i>	<i>Integer</i>	<i>STDs: molluscum contagiosum</i>	<i>Boolean</i>	<i>Hinselmann (Label)</i>	<i>Boolean</i>
<i>IUD</i>	<i>Boolean</i>	<i>STDs: AIDS</i>	<i>Boolean</i>	<i>Schiller (Label)</i>	<i>Boolean</i>
<i>IUD (years)</i>	<i>Integer</i>	<i>STDs: HIV</i>	<i>Boolean</i>	<i>Citology (Label)</i>	<i>Boolean</i>
<i>STDs</i>	<i>Boolean</i>	<i>STDs: Hepatitis B</i>	<i>Boolean</i>	<i>Biopsy (Label)</i>	<i>Boolean</i>

Gambar 3.3 dataset berdasarkan label

Algoritma	Forward & KNN
Jenis Data	Publick
Alamat Dataset	https://www.kaggle.com/datasets/loveall/cervical-cancer-risk-classification

2. Evaluasi

Eksperimen dengan menerapkan teknik Forward Selection, yang menggabungkan beberapa model base yang beragam untuk meningkatkan akurasi prediksi, selama eksperimen ini. Hasil dari eksperimen ini yang merupakan informasi kunci nuntuk evaluasi kualitas prediksi kanker serviks. Dapat dilihat pada gambar dibawah ini:



Gambar 3.4 Hasil Akurasi Forward-KNN

3.3 Alat dan Bahan

3.3.1 Alat

Berikut adalah rincian spesifikasi perangkat keras dan perangkat lunak yang akan digunakan:

a) Perangkat Lunak

Berikut adalah perangkat lunak yang digunakan untuk menganalisa data penyakit kanker serviks:

- 1) Microsoft Excel
- 2) Rapid Miner

b) Perangkat Keras

Berikut adalah perangkat keras yang digunakan untuk mengoperasikan seluruh software yang dibutuhkan didalam penelitian ini:

- 1) Processor Intel i3
- 2) RAM 16Gb
- 3) Kapasitas SSD 256Gb
- 4) Perangkat mouse dan keyboard

c) Bahan Penelitian

1. Dataset

<https://www.kaggle.com/datasets/loveall/cervical-cancer-risk-classification>

2. Jurnal Nasional dan Internasional