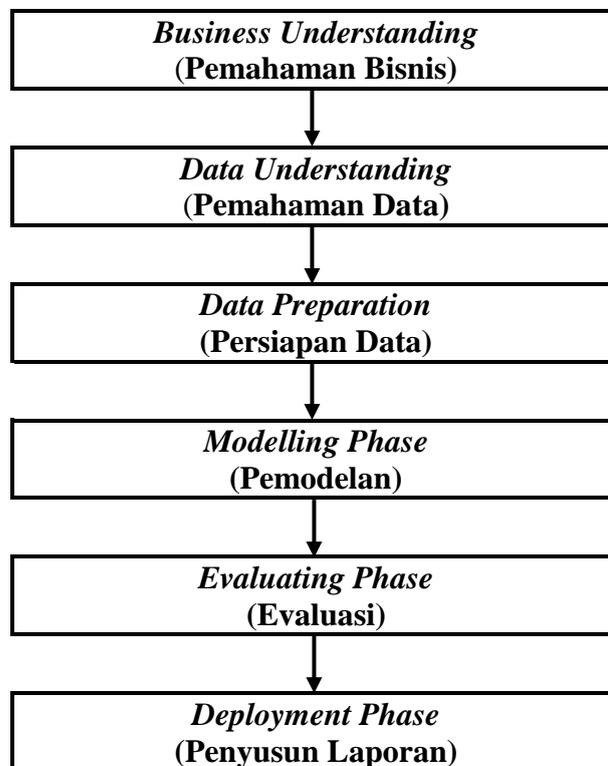


BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

Pada Gambar 3.1 akan dijelaskan metodologi penelitiannya. Alur penelitian menggambarkan tahapan mulai dari awal sampai selesai. Tahapan alur penelitian dapat dilihat sebagai berikut:



Gambar 3.1 Alur Dalam Tahapan Penelitian

Berikut penjelasannya berdasarkan Gambar 3.1 Alur Tahapan Penelitian :

1. *Business Understanding* (Tahap Pemahaman bisnis), Fase ini, yang juga dikenal sebagai fase pemahaman penelitian, mendefinisikan tantangan penambangan data, merumuskan pendekatan awal untuk mencapai tujuan, dan membicarakan tentang tujuan dan kebutuhan proyek untuk unit bisnis atau penelitian.

2. *Data Understanding* (Tahap pemahaman data), Anda dapat membiasakan diri dengan data dan memperoleh pemahaman awal dengan melakukan analisis data eksplorasi selama proses pengumpulan data. Selain itu, Anda dapat mengumpulkan kualitas data dan memilih subgrup menarik yang mungkin memiliki *tren* berguna. 400 *records*, 26 atribut, dan target kelas adalah penyakit ginjal kronik (PGK) pada dataset publik pada kondisi yang diakses melalui tautan berikut: <https://www.kaggle.com/datasets/mansoordaku-/ckdisease>. Atribut yang digunakan dalam penelitian ini dijelaskan pada Tabel 3.1:

Tabel 3.1 Atribut-Atribut yang Digunakan pada Penelitian

Atribut	Deskripsi	Nilai	Type Data
<i>ID</i>	Nomor identifikasi	1,2,3,4,5	<i>Integer</i>
<i>Age</i>	Umur	Umur dalam tahun	<i>Integer</i>
<i>Blood Pressure (bp)</i>	Ukuran tekanan yang dihasilkan oleh darah	<i>mm/Hg</i>	<i>Integer</i>
<i>Specific Gravity (sg)</i>	Seriusnya suatu situasi atau kondisi	1.005,1.010,1.015, 1.020,1.025	<i>Real</i>
<i>Albumin (al)</i>	Protein plasma darah	0,1,2,3,4,5	<i>Integer</i>
<i>Sugar (su)</i>	Kadar gula darah	0,1,2,3,4,5	<i>Integer</i>
<i>Red Blood Cells (rbc)</i>	Pengukuran sel darah merah	<i>Normal, abnormal</i>	<i>Binominal</i>
<i>Pus Cell (pc)</i>	Pengukuran sel nanah dalam urin	<i>Normal, abnormal</i>	<i>Binominal</i>
<i>Pus Cell clumps (pcc)</i>	Pengukuran gumpalan sel nanah dalam urin	<i>Present, not present</i>	<i>Binominal</i>
<i>Bacteria (ba)</i>	Pengukuran bakteri yang terdapat dalam tubuh	<i>Present, not present</i>	<i>Binominal</i>
<i>Blood Glucose Random (bgr)</i>	Pengukuran kadar glukosa	<i>mgs/dl</i>	<i>Real</i>
<i>Blood Urea (bu)</i>	Pengukuran kadar urea dalam darah	<i>mgs/dl</i>	<i>Integer</i>
<i>Serum Creatinine (sc)</i>	Pengukuran kreatin pada darah	<i>mgs/dl</i>	<i>Real</i>
<i>Sodium (sod)</i>	Pengukuran natrium pada darah	<i>mEq/L</i>	<i>Real</i>
<i>Potassium (pot)</i>	Pengukuran kalium pada darah	<i>mEq/L</i>	<i>Real</i>
<i>Hemoglobin (hemo)</i>	Menghitung jumlah hemoglobin	<i>gms</i>	<i>Real</i>
<i>Packed Cell Volume</i>	Parameter pengukuran sel darah	<i>numeric</i>	<i>Integer</i>
<i>White Blood Cell</i>	Menghitung jumlah total sel	<i>Millions/cmm</i>	<i>Integer</i>

Atribut	Deskripsi	Nilai	Tipe Data
<i>Count (wc)</i>	darah merah		
<i>Red Blood Cell Count (rc)</i>	Menghitung jumlah total sel darah merah	<i>Millions/cmm</i>	<i>Real</i>
<i>Hypertension (htn)</i>	Memiliki masalah hipertensi	<i>Yes,no</i>	<i>Binominal</i>
<i>Diabetes Mellitus (dm)</i>	Memiliki masalah diabetes	<i>Yes,no</i>	<i>Binominal</i>
<i>Coronary Artery Disease (cad)</i>	Coronary Artery Disease	<i>Yes,no</i>	<i>Binominal</i>
<i>Appetite (appet)</i>	Memiliki masalah nafsu makan	<i>Good,poor</i>	<i>Binominal</i>
<i>Pedal Edema (pe)</i>	Pengukuran cairan pada betis	<i>Yes,no</i>	<i>Binominal</i>
<i>Anemia (ane)</i>	Pengukuran darah merah	<i>Yes,no</i>	<i>Binominal</i>
<i>Class (class)</i>	Target Class	<i>CKD, not-CKD</i>	<i>Binominal</i>

Setiap karakteristik data penelitian dijelaskan pada tabel di atas. Kumpulan data ini bertujuan untuk mengumpulkan informasi tentang gejala dan menentukan apakah seseorang mungkin menderita Penyakit Ginjal Kronis (PGK).

3. Data Preparation (Tahap persiapan data), mempersiapkan data mentah putaran pertama untuk kumpulan data akhir yang akan digunakan di semua fase berikutnya. Untuk membuat data mentah digunakan alat pemodelan aplikasi seperti RapidMiner, Anda harus memilih kasus dan variabel yang ingin Anda pelajari. Kemudian, jika diperlukan, Anda dapat melakukan transformasi pada variabel tertentu. Agar penambangan data berhasil, langkah persiapan data sangat penting. Pertama, prosedur pemilihan data, yaitu menghilangkan atribut-atribut yang tidak diperlukan. Selanjutnya, prosedur integrasi diterapkan untuk menggabungkan informasi dari banyak *Database*. Karena data yang digunakan dalam penelitian ini hanya berasal dari satu *Database*, penulis tidak melakukan integrasi data, untuk atribut ID tidak digunakan karena tidak memberikan informasi yang relevan atau bermakna untuk analisis. Setelah itu menambahkan operator *Replace* yang berfungsi untuk mengganti nilai-nilai tertentu dalam suatu dataset dengan nilai yang baru, di penelitian ini atribut kelas nya berupa "CKD" (*Chronic Kidney Disease*) dan "not-CKD" (tidak- *Chronic Kidney Disease*), dengan keterangan atribut kelas sebagai berikut:

- a. Nilai 1.0 dapat diinterpretasikan sebagai prediksi bahwa seseorang memiliki CKD (“CKD”).
- b. Nilai 0.0 dapat diinterpretasikan sebagai prediksi bahwa seseorang tidak memiliki CKD (“not-CKD”).

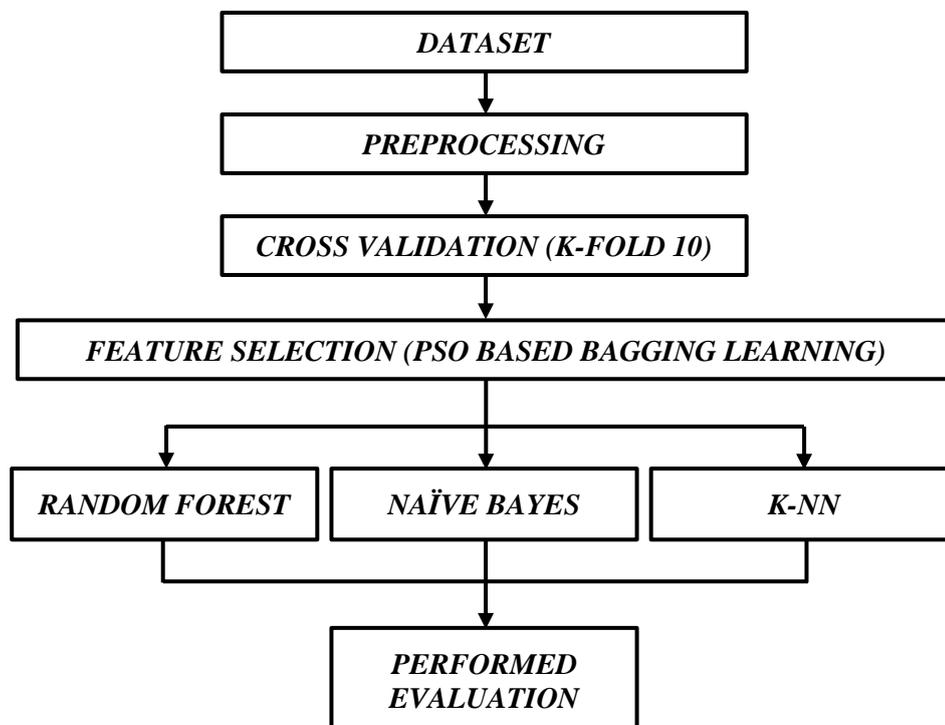
Operator *Replace Missing Values* kemudian diterapkan dalam kasus ini. Tujuannya adalah untuk menangani nilai-nilai yang kosong atau hilang dalam kumpulan data dan pembersihan data, yaitu tindakan menghilangkan data yang berlebihan, tidak lengkap, atau kosong. Karena data yang diunduh masih ada yang *missing*, maka penulis melakukan pembersihan data. Hasilnya, nilai rata-rata setiap atribut yang digunakan diterapkan pada data kosong. Data kosong dapat Anda lihat pada tabel 3.2 di bawah ini.

Tabel 3.2 Data-Data yang *Missing*

Atribut	Missing
<i>Age</i>	9
<i>Blood Pressure (bp)</i>	12
<i>Specific Gravity (sg)</i>	47
<i>Albumin (al)</i>	46
<i>Sugar (su)</i>	49
<i>Red Blood Cells (rbc)</i>	152
<i>Pus Cell (pc)</i>	65
<i>Pus Cell clumps (pcc)</i>	4
<i>Bacteria (ba)</i>	4
<i>Blood Glucose Random (bgr)</i>	44
<i>Blood Urea (bu)</i>	19
<i>Serum Creatinine (sc)</i>	17
<i>Sodium (sod)</i>	87
<i>Potassium (pot)</i>	88
<i>Hemoglobin (hemo)</i>	52
<i>Packed Cell Volume (pcv)</i>	71
<i>White Blood Cell Count (wc)</i>	106
<i>Red Blood Cell Count (rc)</i>	131
<i>Hypertension (htn)</i>	2
<i>Diabetes Mellitus (dm)</i>	2
<i>Coronary Artery Disease (cad)</i>	2
<i>Appetite (appet)</i>	1
<i>Pedal Edema (pe)</i>	1
<i>Anemia (ane)</i>	1

Setelah data dibersihkan, selanjutnya di tambahkan operator *Map* yang bertujuan untuk mentransformasi nilai-nilai dalam suatu kolom atau atribut *dataset* ke nilai-nilai baru sesuai dengan aturan yang ditentukan, atribut-atribut yang diubah nilainya adalah (ane, appet, ba, cad, dm, htn, pc, pcc, pe, rbc) misal dipenelitian ini untuk mengonversi nilai kategori "Yes" dan "No", ke nilai-nilai *numerik* seperti 1 dan 0. Setelah itu data ditransformasikan kedalam format yang sama dengan menggunakan operator *Store*. Penulis dalam penelitian ini mengubah data menjadi format tabular dengan menggunakan *Microsoft Excel*.

4. *Modeling Phase* (Tahap pemodelan), menetapkan metodologi serta merancang prosedur pengujian yang akan diterapkan guna menjalankan proses klasifikasi menggunakan teknik data mining, dengan tujuan mencapai hasil akurasi optimal. Metode yang digunakan yaitu *Random Forest*, *Naïve Bayes* dan *k-NN* dengan menggunakan metode *Bagging* berbasis *Particel Sward Optimazion* (PSO). Gambar berikut menunjukkan diagram model yang digunakan dalam penelitian ini.



Gambar 3.2 Bagan Pemodelan

5. *Evaluation Phase* (Tahap evaluasi), melakukan eksperimen atau dalam rangka menemukan hasil akurasi terbaik melalui evaluasi yang teliti dan cermat.
6. *Deployment Phase* (Tahap penyusunan laporan), merumuskan kesimpulan dalam bentuk laporan berdasarkan hasil pengujian, dengan merujuk pada pernyataan masalah dan tujuan penelitian yang telah ditetapkan.

3.2 Alat dan Bahan

Laptop *Acer Intel(R) Core(TM) i5 8250U* dengan RAM 8.00 GB dan CPU @ 1.60GHz hingga 1.80GHz merupakan perangkat keras yang digunakan dalam penelitian ini. *Microsoft Excel* dan *RapidMiner Studio 10* adalah program perangkat lunaknya.