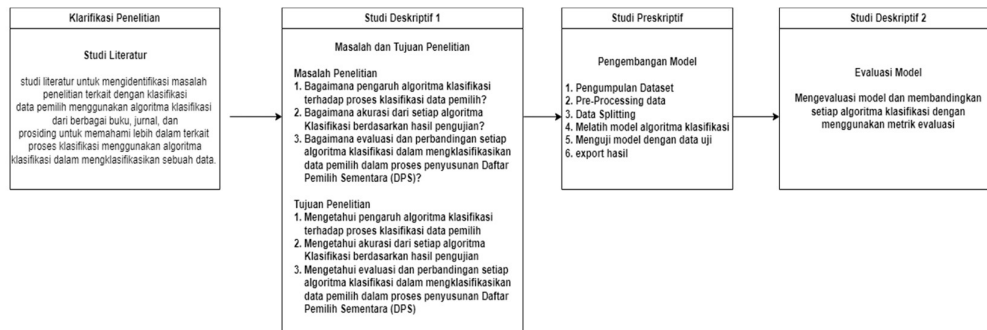


## BAB III METODE PENELITIAN

### 3.1 Desain Penelitian

Metode penelitian yang digunakan dalam penelitian ini yaitu metode Design Research methodology (DRM). Metode Design Research Methodology adalah suatu pendekatan yang digunakan untuk mengkaji, mengembangkan, dan menguji berbagai desain dalam konteks inovasi dan pemecahan masalah. Metode ini terutama digunakan dalam bidang desain, seni, dan teknik, tetapi juga dapat diterapkan pada berbagai disiplin ilmu lainnya. Design Research Methodology (DRM) menekankan proses desain sebagai inti penelitian. Tujuannya adalah untuk menciptakan sesuatu yang baru, menyelesaikan masalah, atau meningkatkan desain produk, layanan, sistem, atau pengalaman pengguna. Tahapan dalam metode penelitian Design Research Methodology (DRM) terdiri dari klarifikasi penelitian, studi deskriptif 1, studi preskriptif, dan terakhir studi deskriptif 2 (Chakrabarti dan Blessing, 2009).



Gambar 3.1 Desain Penelitian

#### 3.1.1 Klarifikasi Penelitian

Pada tahap ini dilakukan studi literatur untuk mengidentifikasi masalah penelitian terkait dengan klasifikasi data pemilih menggunakan algoritma klasifikasi dari berbagai buku, jurnal, dan prosiding untuk memahami lebih dalam terkait proses klasifikasi menggunakan algoritma klasifikasi dalam mengklasifikasikan sebuah data. Serta pada tahap ini, perlunya memahami

bagaimana alur penyusunan Daftar Pemilih Sementara (DPS) berdasarkan undang-undang pemilu yang berlaku di Indonesia (PKPU no.7 Tahun 2022).

### **3.1.2 Studi Deskriptif 1**

Tahapan ini menjelaskan tentang permasalahan dan tujuan penelitian berdasarkan studi literatur yang telah dilakukan sebelumnya. Hasil dari langkah ini akan digunakan sebagai dasar untuk membangun solusi untuk masalah penelitian ini.

### **3.1.3 Studi Preskriptif**

Dalam tahap ini, tahapan pengembangan model algoritma klasifikasi diterapkan dengan merujuk pada penelitian sebelumnya sebagai referensi dengan berbagai penyesuaian mengikuti permasalahan penelitian dalam penelitian ini. Kemudian model tersebut dievaluasi dengan menggunakan metrik evaluasi.

### **3.1.4 Studi Deskriptif 2**

Tahapan terakhir yaitu mengevaluasi model yang telah dikembangkan dengan berbagai metrik evaluasi. Hasil evaluasi ini kemudian dibuat kesimpulan dari penelitian ini.

## **3.2 Populasi dan Sampel**

Populasi merupakan keseluruhan aspek dari subjek penelitian yang akan diamati dan diteliti oleh peneliti dalam waktu tertentu, sedangkan sampel merupakan bagian dari jumlah dan karakteristik yang berasal dari populasi. Populasi dari penelitian ini adalah data potensial pemilih yang berasal dari Data Penduduk Potensial Pemilih Pemilu (DP4). Sedangkan sampel yang diambil yakni data potensial pemilih dari DP4 yang berdomisili di Desa Cisempur, Kec. Jatinangor, Kab. Sumedang, yang mana data tersebut sudah dilakukan pencocokan dan penelitian (Coklit) oleh pantarlih Desa Cisempur.


## **3.3 Alat dan Bahan Penelitian**

Dalam penelitian ini, model algoritma klasifikasi dibangun menggunakan bahasa pemrograman *python* dengan Google colab sebagai *code editor*-nya. Adapun library dan module yang digunakan adalah sebagai berikut:

- 1) *Pandas* : library yang digunakan untuk manipulasi dan analisis data
- 2) *Scikit-learn* : library untuk membuat model algoritma klasifikasi
- 3) *Train\_test\_split* : untuk memisahkan data menjadi data training dan data testing
- 4) *DecisionTreeClassifier*: untuk membuat model *Decision Tree*
- 5) *KNeighborsClassifier* : untuk membuat model *k-NN*
- 6) *GaussianNB* : untuk membuat model *Naive Bayes*
- 7) *SVC* : untuk membuat model *SVM Classifier*
- 8) *OneHotEncoder* : untuk mengubah data kategorikal menjadi numerikal.
- 9) *Accuracy\_score* : untuk menghitung nilai *accuracy*
- 10) *Precision\_score* : untuk menghitung nilai *precision*
- 11) *Recall\_score* : untuk menghitung nilai *recall*
- 12) *F1\_score* : untuk menghitung nilai *f1-score*
- 13) *Sms.heatmap* : untuk membuat *confusion matrix* dalam bentuk heatmap

### 3.4 Instrumen Penelitian

MODEL A-Daftar Pemilih



**DAFTAR PEMILIH  
PEMILIHAN UMUM  
TAHUN 2024**

PROVINSI : ..... Kecamatan : .....

KABUPATEN/ : ..... Kelurahan/Desa : .....

KOTA : ..... TPS : .....

No	No KK	NIK	Nama	Tempat Lahir	Tanggal Lahir	Status Perkawinan B/S/P	Jenis Kelamin L/P	Alamat KTP-el			Disabilitas	Status Kepemilikan KTP-el S/B	Keterangan
								Jalan/Dukuh	Rt	Rw			
1	2	3	4	5	6	7	8	9	10	11	12	13	14

Gambar 3.2 Contoh Model-A Daftar Pemilih

Sumber : [jdih.kpu.go.id](http://jdih.kpu.go.id)

Rencana target eksperimen atau studi kasus dari penelitian yang akan dibuat adalah meneliti bagaimana penggunaan algoritma klasifikasi dapat mempengaruhi proses penyusunan Daftar Pemilih Sementara. Sumber data diperoleh dari PPS Desa Cisempur dan Pantarlih Desa Cisempur berupa lembar Model-A Daftar Pemilih dan Model-A Daftar Potensial Pemilih dari salahsatu TPS yang ada di Desa Cisempur.

Aspek yang diteliti adalah seberapa akurat algoritma klasifikasi dalam mengklasifikasi data potensial pemilih apabila dibandingkan dengan cara konvensional. Juga berapa banyak data yang memenuhi syarat dan tidak memenuhi syarat yang berhasil diklasifikasikan oleh algoritma klasifikasi.

### 3.4.1 Dataset

Dalam penelitian ini, data yang didapatkan berupa data pemilih yang berdomisili di Desa Cisempur yang terdiri dari variabel NKK, NIK, Nama, Tempat lahir, Tanggal Lahir, Status Perkawinan, Jenis Kelamin, Alamat, RT, RW, Disabilitas, dan Status e-KTP. Berikut informasi variabel dari dataset tersebut.

Tabel 3.1

*Tabel Informasi Variabel*

No	Variabel	Deskripsi	Tipe Data
1	NKK	Nomor Kartu Keluarga	<i>integer</i>
2	NIK	Nomor Induk Kependudukan	<i>integer</i>
3	Nama	Nama lengkap pemilih	<i>string</i>
4	Tempat_lahir	Tempat lahir pemilih	<i>string</i>
5	Tanggal_lahir	Tanggal lahir pemilih	<i>string</i>
6	Usia	Usia Pemilih (<17 Tahun/≥17 Tahun)	<i>integer</i>
6	Status_Perkawinan	Status perkawinan (Belum/Sudah/Pernah)	<i>string</i>
7	Jenis_Kelamin	Jenis kelamin (Laki-laki/Perempuan)	<i>string</i>
9	Alamat	Alamat sesuai KTP pemilih	<i>string</i>
10	RT	RT sesuai KTP pemilih	<i>string</i>
11	RW	RW sesuai KTP pemilih	<i>string</i>

No	Variabel	Deskripsi	Tipe Data
12	Disabilitas	Status disabilitas (disabilitas fisik/mental/intelektual, sensorik wicara/rungu/netra)	<i>string</i>
13	Status_e-KTP	Status kepemilikan e-KTP (Belum/Sudah)	<i>string</i>
14	Keterangan	Keterangan pemilih (Pemilih Sesuai/Pemilih Tersaring/Pemilih Ubah/Pemilih Baru)	<i>string</i>

Adapun jumlah data pemilih yang diperoleh dari PPS Desa Cisempur sebanyak 7310 data pemilih, terdiri dari 3634 pemilih berjenis kelamin laki-laki dan 3676 pemilih berjenis kelamin perempuan yang berasal dari delapan Dusun di Desa Cisempur yaitu Dusun Baturumpil, Dusun Cisempur, Dusun Lunjuk, Dusun Karasak, Dusun Sirahcai, Dusun Cintamanah, Dusun Wates, dan Dusun Calamkuning, serta satu kawasan perumahan yaitu Perumahan Taman Bukit Makmur.

### 3.5 Tahapan Pengembangan

#### 3.5.1 Selection

Tahapan ini bertujuan untuk memilih dataset yang akan digunakan dalam penelitian ini. Data yang digunakan harus sesuai dengan batasan penelitian yang telah ditetapkan dalam penelitian ini yaitu dataset hanya berasal dari data pemilih yang berdomisili di Desa Cisempur, Kec. Jatinangor, Kab. Sumedang dengan jumlah data yang diperoleh sebanyak 7310 data pemilih yang dibagi menjadi 6665 data latih dan 645 data uji.

#### 3.5.2 Pre-Processing

Tahapan pre-processing bertujuan untuk mencari data yang memiliki missing value yang berarti data yang telah diperoleh belum normal. Dalam tahapan pre-processing ini juga terdapat proses membuat label/Class. Dalam penelitian ini, class yang dibuat diberi label “Memenuhi Syarat”, dan “Tidak Memenuhi Syarat”. Proses pelabelan dilakukan bersama dengan Pantarlih Desa Cisempur, data latih

dilabeli oleh Pantarlih Desa Cisempur sedangkan data uji dilabeli oleh peneliti sendiri.

Dalam tahapan ini, dataset akan dibagi menjadi data latih dan data uji. Data latih akan dibagi lagi menjadi data training dan data validasi dengan rasio 80:20. Dengan menggunakan rasio 80:20, maka memungkinkan untuk mendapatkan hasil yang valid, dalam artian tidak melebih-lebihkan akurasi model, serta lebih akurat diantara estimasi rasio yang valid dimana overestimasi dari perkiraan error pada rasio ini adalah yang paling kecil (Gholamy dkk., 2018).

### 3.5.3 Pengembangan Model

Dalam penelitian ini, model *Decision Tree* dikembangkan dengan menerapkan library scikit-learn. Penggunaan library scikit-learn memberikan beberapa kemudahan yaitu memberikan kemudahan dalam melakukan pre-processing data, kemudahan dalam mengukur kinerja model dengan menyediakan berbagai metrik kinerja, dan kemudahan dalam memvisualisasikan model *Decision Tree* menggunakan library eksternal seperti Graphviz.

Proses selanjutnya adalah melatih model *Decision Tree* dengan menggunakan dataset training yang dibagi menjadi 80% data training, dan 20% data validasi untuk mengevaluasi akurasi algoritma ketika algoritma sedang dilatih.

### 3.5.4 Evaluasi Model algoritma klasifikasi

Tahapan ini bertujuan untuk memperoleh hasil menggunakan model algoritma yang telah ditentukan sebelumnya. Tahapan ini adalah langkah kunci dalam pembuatan model. Data yang digunakan didapat dari PPS Desa Cisempur dan Pantarlih Desa Cisempur berupa lembar Model-A Daftar Pemilih dan Model-A Daftar Potensial Pemilih dari seluruh TPS yang ada di Desa Cisempur yang telah dikonversi kedalam format .csv yang selanjutnya akan dibagi menjadi dataset untuk melatih model dan dataset untuk menguji model.

Untuk menunjukkan kinerja dari algoritma yang digunakan menggunakan persamaan berikut ini:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100\%$$

Keterangan:

TP = Benar Positif (True Positive)

FP = Benar Negatif (False Positive)  
 TN = Salah Positif (True Negative)  
 FN = Salah Negatif (False Negative)

*Precision* adalah perbandingan antara benar positif dengan keseluruhan hasil yang dianggap positif.

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%$$

*Recall* adalah perbandingan antara data yang benar positif dengan jumlah dari data yang benar positif.

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\%$$

*F1-score* adalah hasil dari perbandingan rata-rata *Precision* dan *Recall* yang dibobotkan.

$$\text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})}$$

*Confusion matrix* adalah tabel yang digunakan untuk mengevaluasi kinerja klasifikasi untuk masalah klasifikasi pembelajaran mesin yang keluarannya dapat menjangkau dua kelas atau lebih. *Confusion matrix* memiliki empat elemen yang digambarkan oleh tabel dibawah ini:

Tabel 3.2

*Confusion matrix*

Observasi	Prediksi Positif	Prediksi Negatif
Positif	True positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

- True positive (TP): jumlah instance yang diklasifikasikan dengan benar sebagai positif
- True Negative (TN): jumlah instance yang diklasifikasikan dengan benar sebagai negatif
- False Positive (FP): jumlah instance yang diklasifikasikan salah sebagai positif
- False Negative (FN): jumlah instance yang diklasifikasikan salah sebagai negatif

Dengan adanya *Confusion matrix*, kita dapat menilai sejauh mana kemampuan model dalam mengklasifikasikan data pemilih, dan apakah model memiliki kecenderungan tertentu dalam membuat kesalahan tertentu dengan cara memasukkan empat elemen *Confusion matrix* ini.

### 3.6 Analisis Data

Dalam penelitian ini, analisis data dilakukan dengan melakukan evaluasi terhadap data pemilih yang telah mengalami proses klasifikasi oleh model algoritma klasifikasi yang telah dibuat. Data pemilih yang telah diklasifikasikan oleh setiap model algoritma klasifikasi akan dibandingkan satu sama lain. Langkah ini bertujuan untuk mengukur tingkat akurasi model algoritma klasifikasi dalam melakukan klasifikasi data pemilih. Dengan membandingkan data hasil klasifikasi model dengan data yang dipilih secara manual, penelitian ini akan mengungkap sejauh mana efektivitas dan ketepatan setiap model algoritma klasifikasi dalam mengklasifikasikan data pemilih. Selanjutnya, proses penyusunan daftar pemilih sementara dapat dilaksanakan dengan mempertimbangkan data dari hasil analisis sebelumnya. Data tersebut dapat membantu mengarahkan proses pengumpulan dan pengaturan informasi yang diperlukan untuk menghasilkan daftar pemilih sementara yang akurat dan sesuai dengan kriteria yang telah ditetapkan.