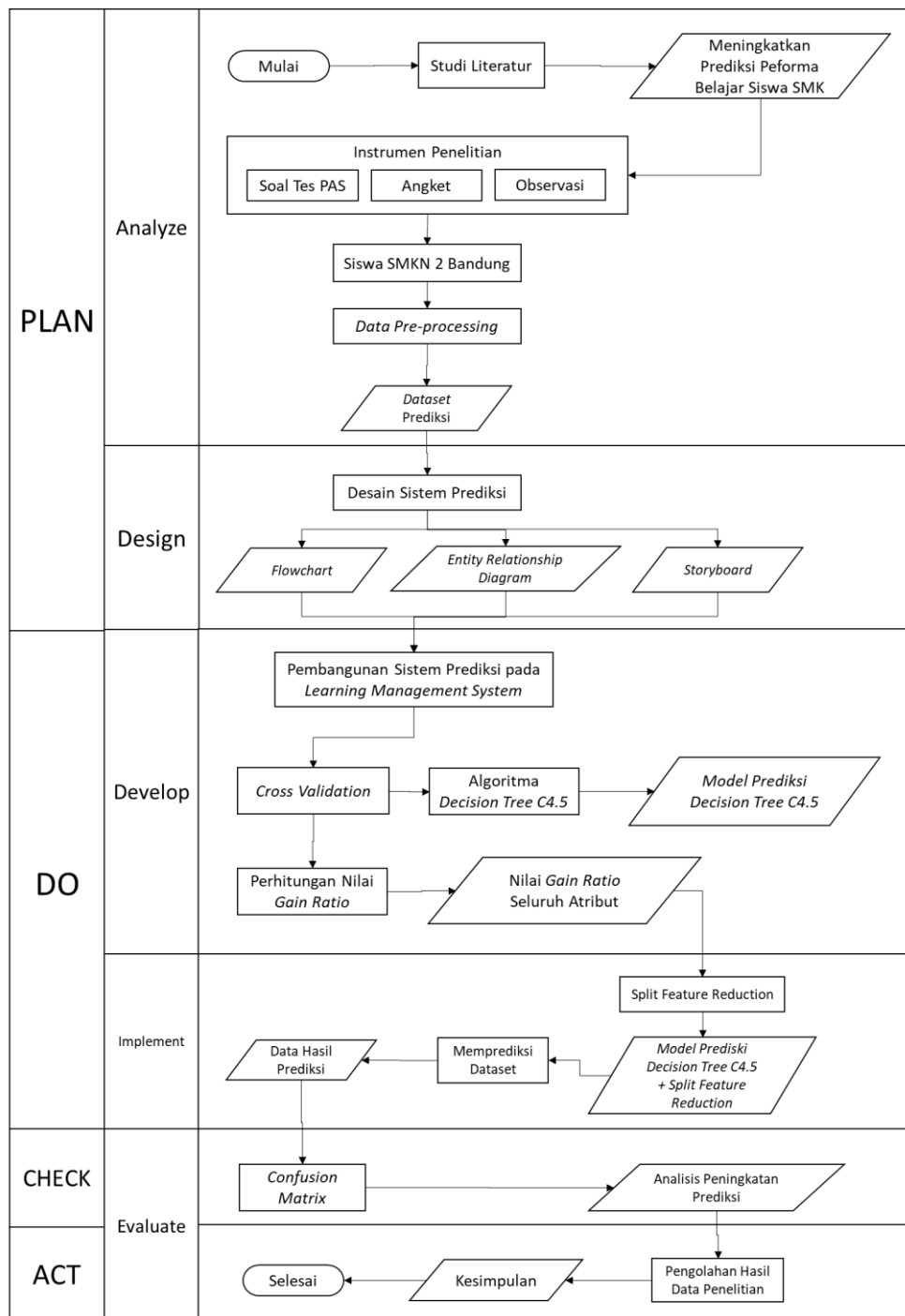


BAB III

METODE PENELITIAN

Prosedur penelitian ini akan dipaparkan dalam bentuk *Smart Learning Environment Establishment Guideline (SLEEG)* berdasarkan *Analyze-Design-Develop-Implement-Evaluate (ADDIE)* yang dikutip dari Rosmansyah, Putro, Putri, Utomo, dan Suhardi (2022).



Gambar 3.1 Prosedur Penelitian

Penjelasan untuk gambar prosedur penelitian di atas adalah sebagai berikut:

3.1 *Analyze*

Pada tahap *analyze* penulis melakukan studi terhadap artikel jurnal internasional terkait *Educational Data Mining* yang meneliti mengenai prediksi performa belajar siswa. Pada tahap ini juga dilakukan perancangan instrumen penelitian untuk pengumpulan data siswa yang akan digunakan dalam pembuatan model prediksi performa belajar siswa SMK. Instrumen penelitian yang digunakan untuk mengumpulkan data siswa adalah Soal Tes Penilaian Akhir Semester (PAS), Angket, dan Observasi. Instrumen Soal Tes PAS digunakan untuk mengumpulkan nilai PAS siswa yang digunakan sebagai acuan performa belajar siswa, sedangkan angket digunakan untuk mengumpulkan data latar belakang siswa seperti jenis kelamin, umur, akses *smartphone*, akses internet, akses komputer, jumlah saudara, wali di rumah, minat terhadap jurusan kelas, pendidikan dan pekerjaan orang tua, dan observasi dilakukan untuk mengumpulkan data akademik siswa selama kegiatan pembelajaran di sekolah seperti kehadiran, tugas, dan nilai PTS. Setelah instrumen penelitian diberlakukan kepada siswa kelas 10 TMP 2,3, dan 4 SMKN 2 Bandung, data yang terkumpul diproses menjadi bentuk yang tertata untuk dianalisis.

3.1.1 Instrumen Penelitian

Instrumen penelitian adalah suatu alat yang digunakan untuk mengumpulkan data atau mengukur objek dari suatu variabel penelitian (Yusup, 2018). Alat pengumpul data yang digunakan dalam penelitian ini adalah soal tes, pedoman angket, dan pedoman observasi.

3.1.1.1 Soal Tes PAS

Soal tes PAS digunakan untuk mengetahui nilai PAS siswa sebagai pengukur performa belajar siswa selama satu semester. Soal tes tersebut merupakan tes tipe objektif pilihan berganda yang terdiri dari 40 butir soal yang diberlakukan kepada seluruh kelas 10 SMKN 2 Bandung pada mata pelajaran Informatika. Materi soal tes PAS Informatika yang diujikan adalah sebagai berikut:

- Berpikir Komputasional
 - Abstraksi
 - Algoritma

- Dekomposisi
- Pengenalan pola
- Teknologi Informasi dan Komunikasi
 - Aplikasi Pengolah Kata
 - Aplikasi Pengolah Angka
 - Aplikasi Presentasi
- Sistem Komputasi
 - Perangkat Keras Komputer
 - Perangkat Lunak Komputer
- Jaringan Komputer dan Internet
 - Komponen Jaringan
 - *Layout* Jaringan

Uji validitas yang digunakan untuk instrumen soal tes PAS ini adalah uji validitas banding tes menggunakan rumus korelasi produk momen Pearson (Yusup, 2018). Rumus yang dimaksud adalah sebagai berikut:

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{(n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2)(n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2)}}$$

Rumus 3.1 Korelasi *Product Moment*

Keterangan:

- r_{xy} = koefisien korelasi antara variabel x dan variabel y
- x_i = nilai data ke-i untuk kelompok variabel x
- y_i = nilai data ke-i untuk kelompok variabel y
- n = banyak data

Instrumen soal dikatakan valid ketika nilai koefisien korelasi Pearson lebih besar dari nilai r pada tabel Pearson dengan taraf signifikansi $\alpha = 0,05$ dan $n = 110$ sesuai dengan banyak data (Yusup, 2018). Maka kriteria validitas instrumen dapat diuraikan sebagai berikut:

- Instrumen valid, jika $r_{xy} \geq r_{tabel}$
- Instrumen tidak valid, jika $r_{xy} < r_{tabel}$

Validitas instrumen lalu dapat dikategorikan menjadi beberapa tingkatan dengan ketentuan sebagai berikut:

Tabel 3.1 Kategori Validitas Instrumen Soal Tes PAS

$0,80 < r_{xy} \leq 1,00$	Sangat Tinggi
$0,60 < r_{xy} \leq 0,80$	Tinggi
$0,40 < r_{xy} \leq 0,60$	Sedang
$0,20 < r_{xy} \leq 0,40$	Rendah
$0,00 < r_{xy} \leq 0,20$	Sangat Rendah
$r_{xy} \leq 0,00$	Tidak Valid

Setelah dilakukan validitas, instrumen soal tes PAS diuji reliabilitasnya untuk diketahui konsistensinya dalam mengumpulkan data (Yusup, 2018). Rumus yang digunakan adalah *Spearman-Brown*.

$$r_{11} = \frac{2 * r_{xy}}{1 + r_{xy}}$$

Rumus 3.2 *Spearman-Brown*

Keterangan:

r_{11} = Koefisien Reliabilitas *Spearman-Brown*

r_{xy} = koefisien korelasi antara variabel x dan variabel y

Berikut adalah ketentuan dalam menentukan kategori nilai reliabilitas *Spearman-Brown*:

Tabel 3.2 Kategori Reliabilitas Instrumen Soal Tes PAS

$0,80 < r_{11} \leq 1,00$	Sangat Tinggi
$0,60 < r_{11} \leq 0,80$	Tinggi
$0,40 < r_{11} \leq 0,60$	Sedang
$0,20 < r_{11} \leq 0,40$	Rendah
$0,00 < r_{11} \leq 0,20$	Sangat Rendah

Berikutnya adalah menentukan indeks kesukaran untuk mengetahui seberapa sulit atau mudahnya siswa peserta tes menjawab soal (Arifin, 2017). Rumus untuk menentukan indeks kesukaran soal adalah sebagai berikut:

$$P = \frac{B}{JS}$$

Rumus 3.3 Indeks Kesukaran

Keterangan:

P = Indeks Kesukaran

B = Jumlah siswa menjawab soal dengan betul

JS = Jumlah seluruh siswa peserta tes

Berikut adalah ketentuan dalam menentukan kategori nilai reliabilitas indeks kesukaran:

Tabel 3.3 Kategori Indeks Kesukaran Instrumen Soal Tes PAS

$P < 0,30$	Sukar
$0,30 \leq P \leq 0,70$	Sedang
$0,70 < P$	Mudah

Lalu dicari nilai indeks daya pembeda untuk mengetahui seberapa baik sebuah soal membedakan siswa peserta tes dalam kelompok tinggi dan rendah (Arifin, 2017). Rumus untuk menentukan nilai daya pembeda adalah sebagai berikut:

$$DP = \frac{B_A}{J_A} - \frac{B_B}{J_B}$$

Rumus 3.4 Daya Pembeda

Keterangan:

DP = Daya Pembeda

B_A = Banyak siswa peserta tes kelompok atas menjawab soal dengan benar

B_B = Banyak siswa peserta tes kelompok bawah menjawab soal dengan benar

J_A = Banyak siswa peserta tes kelompok atas

J_B = Banyak siswa peserta tes kelompok bawah

Berikut adalah ketentuan dalam menentukan kategori nilai daya pembeda:

Tabel 3.4 Kategori Daya Pembeda Instrumen Soal Tes PAS

$0,70 \leq DP$	Baik Sekali
$0,40 \leq DP < 0,70$	Baik
$0,20 \leq DP < 0,40$	Cukup
$DP < 0,20$	Jelek

3.1.1.2 Angket

Angket merupakan cara pengumpulan data dengan cara menyebarkan daftar pertanyaan kepada responden dengan harapan responden akan memberikan respons terhadap daftar pertanyaan tersebut. Angket yang dibuat merupakan angket tertutup yang sebagian besarnya menggunakan skala *Guttman* dalam

pengukurannya. Instrumen angket digunakan untuk mengumpulkan data siswa yang berupa data pribadi dan latar belakang siswa seperti jenis kelamin, umur, akses *smartphone*, akses internet, akses komputer, jumlah saudara, wali di rumah, minat terhadap jurusan kelas, pendidikan dan pekerjaan orang tua.

3.1.1.3 Pedoman Observasi

Observasi merupakan cara pengumpulan data dengan cara melakukan pengamatan terhadap sampel penelitian mengenai sifat yang berkaitan dengan penelitian. Pedoman observasi digunakan untuk mengumpulkan data yang bersifat akademik yang sudah dilakukan pada proses pembelajaran. Atribut-atribut siswa yang diamati adalah kehadiran, tugas, dan nilai PTS.

3.1.2 Populasi dan Sampel

Populasi adalah kumpulan unit yang akan diteliti karakteristiknya atau keseluruhan obyek yang akan dijadikan sasaran penelitian. Sampel adalah bagian dari populasi yang diseleksi dengan harapan bagian tersebut dapat merefleksikan seluruh karakteristik yang ada pada populasi (Abdullah, 2015). Pada umumnya, jumlah sampel yang dipandang cukup besar untuk penelitian korelasional adalah sebanyak 30 orang (Sukmadinata, 2012).

Populasi dalam penelitian ini adalah siswa kelas 10 SMKN 2 Bandung yang mempelajari mata pelajaran Informatika dengan jumlah 648 orang siswa. Sampel yang akan digunakan dalam penelitian ini adalah siswa kelas 10 X TMP 2, 3, dan 4 yang berjumlah 110 orang. Penentuan sampel yang digunakan pada penelitian ini menggunakan jenis *non probability sampling* dengan teknik *convenience sampling*. *Convenience sampling* digunakan berdasarkan kemudahan pengambilan sampel yang diperlukan. Sampel siswa yang diambil merupakan siswa kelas yang diajar oleh penulis pada mata pelajaran Informatika.

3.2 Design

Pada tahap ini dilakukan desain sistem prediksi dan sistem *early warning* yang akan diimplementasikan pada sebuah sistem pengolahan data yang menyerupai sebuah *Learning Management System* (LMS). Desain yang dibuat berupa *flowchart*, *entity relationship diagram* (ERD) dan *storyboard*. *Flowchart* didesain untuk menggambarkan alur kerja sistem yang dirancang dalam bentuk diagram tertata, ERD didesain untuk menggambarkan hubungan antar entitas data

yang ada pada sistem, dan *Storyboard* didesain untuk menggambarkan visual alur kerja sistem berupa desain antarmuka.

3.3 *Develop*

Pada tahap *develop* penulis membangun model prediksi berdasarkan *dataset* yang dihasilkan pada tahap *design* menggunakan teknik *Cross Validation fold 10*. *Dataset* mula-mula dihitung nilai *entropy*-nya untuk diketahui homogenitas kumpulan data, lalu perhitungan nilai *gain* dan *split info* yang nantinya akan digunakan untuk perhitungan nilai *gain ratio* dalam menentukan signifikansi atribut terhadap performa belajar siswa sehingga alur model prediksi *Decision Tree C4.5* terbentuk. Proses ini diulang sebanyak 10 kali dengan perbandingan jumlah *train data* dan *test data* sebanyak 9:1 yang bertukar-tukar peran pada setiap *fold*-nya. *Train dataset* digunakan sebagai acuan dalam memodelkan model prediksi *Decision Tree C4.5*, dan *test dataset* akan digunakan dalam menguji dan mengevaluasi model prediksi *Decision Tree C4.5* yang dihasilkan. Pada tahap ini juga sistem prediksi dan sistem *early warning* dikembangkan menggunakan kerangka kerja *PHP Laravel*.

3.4 *Implement*

Pada tahap *implement* penulis menerapkan teknik *Split Feature Reduction* untuk meningkatkan model prediksi *Decision Tree C4.5* yang sebelumnya dihasilkan. Teknik *Split Feature Reduction* ini memanfaatkan nilai *gain ratio* yang sebelumnya didapatkan untuk menghilangkan atribut yang tidak berarti sama sekali terhadap performa belajar, dan membagi atribut yang signifikan menjadi beberapa bagian yang sama besar untuk meningkatkan kemampuan model prediksi. Pada tahap ini juga model prediksi yang dihasilkan diimplementasikan pada sistem prediksi untuk mengklasifikasikan siswa yang memiliki performa baik atau buruk, jika siswa memiliki performa buruk maka sistem *early warning* akan bekerja dan mengingatkan siswa dalam bentuk notifikasi bahwa siswa tersebut memiliki performa rendah sehingga siswa dan guru dapat berkegas dalam mengatasi masalah yang dialami.

3.5 *Evaluate*

Pada tahap *evaluate* penulis menerapkan teknik *Confusion Matrix* untuk mengevaluasi model prediksi yang dihasilkan. *Confusion Matrix* digunakan untuk mengetahui jumlah data yang berhasil diprediksi dan data yang gagal diprediksi

untuk menghasilkan nilai *accuracy*, *precision*, *recall*, *f-measure*, dan *error-rate* model prediksi yang nantinya digunakan untuk menentukan peningkatan yang dihasilkan teknik *Split Feature Reduction* dari model prediksi *Decision Tree C4.5* biasa. Pada tahap ini juga dilakukan perbandingan hasil prediksi yang dihasilkan system yang dibuat dengan hasil prediksi yang dihasilkan WEKA untuk mengetahui kebenaran implementasi model prediksi pada LMS. Hasil evaluasi yang didapat lalu diolah menjadi kesimpulan.

Data yang telah diterima instrumen penelitian akan membentuk model prediksi dan hasil pengujiannya akan dianalisis untuk menemukan peningkatan antara model prediksi *Decision Tree C4.5* biasa dengan model prediksi penerapan *Split Feature Reduction* dengan langkah-langkah sebagai berikut:

- a) Menggunakan *Confusion Matrix* untuk menentukan kemampuan model prediksi berdasarkan jumlah data yang terprediksi benar dan salah terhadap aktual data (Yuliansyah, Imaniati, Wirasto, dan Wibowo, 2021). *True Positive* (TP) adalah nilai banyak data positif yang terprediksi benar oleh model prediksi. *True Negative* (TN) adalah nilai banyak data negatif yang terprediksi benar oleh model prediksi. *False Positive* (FP) adalah nilai banyak data negatif yang terprediksi salah oleh model prediksi. *False Negative* (FN) adalah nilai banyak data positif yang terprediksi salah oleh model prediksi.

Tabel 3.5 *Confusion Matrix*

Klasifikasi	Prediksi = Baik	Prediksi = Buruk
Aktual = Baik	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Aktual = Buruk	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

- b) Pertama, nilai yang dicari adalah nilai *accuracy*. Nilai *accuracy* adalah nilai ukur yang menentukan kemampuan model prediksi memprediksi data dengan benar terhadap keseluruhan data (Yuliansyah dkk., 2021). Rumus yang digunakan adalah sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Rumus 3.5 Nilai *Accuracy* Model Prediksi

- c) Nilai kedua yang dicari adalah nilai *precision*. Nilai *precision* adalah nilai ukur yang menentukan kemampuan model prediksi dalam memprediksi data dengan

benar tanpa salah mengklasifikasikan data (Yuliansyah dkk., 2021). Rumus yang digunakan adalah sebagai berikut:

$$Precision (positive) = \frac{TP}{TP + FP}$$

Rumus 3.6 Nilai *Precision (Positive)* Model Prediksi

$$Precision (negative) = \frac{TN}{TN + FN}$$

Rumus 3.7 Nilai *Precision (Negative)* Model Prediksi

- d) Ketiga adalah nilai *recall*. Nilai *recall* adalah nilai ukur yang menentukan kemampuan model prediksi dalam memprediksi data dengan benar sebanyak-banyaknya sesuai data aktual yang ada (Yuliansyah dkk., 2021). Rumus yang digunakan adalah sebagai berikut:

$$Recall (positive) = \frac{TP}{TP + FN}$$

Rumus 3.8 Nilai *Recall (Positive)* Model Prediksi

$$Recall (negative) = \frac{TN}{TN + FP}$$

Rumus 3.9 Nilai *Recall (Negative)* Model Prediksi

- e) Nilai keempat adalah nilai *F1-Score*. *F1-Score* adalah nilai rerata harmonik terhadap nilai *precision* dan nilai *recall*. Nilai ini lebih cocok digunakan untuk mengukur peningkatan sebuah model prediksi ketika jumlah data kelas atribut tujuannya tidak seimbang (Czibula, Mihai, dan Crivei, 2019). Rumus yang digunakan adalah sebagai berikut:

$$F1 - Score (positive) = 2 * \frac{Precision (positive) * Recall (positive)}{Precision (positive) + Recall (positive)}$$

Rumus 3.10 Nilai *F1-Score (Positive)* Model Prediksi

$$F1 - Score (negative) = 2 * \frac{Precision (negative) * Recall (negative)}{Precision (negative) + Recall (negative)}$$

Rumus 3.11 Nilai *F1-Score (Negative)* Model Prediksi

- f) Nilai kelima adalah nilai *error-rate*. Nilai *error-rate* ini mirip dengan nilai *accuracy* tetapi untuk mengukur data-data yang salah diprediksi oleh model prediksi terhadap keseluruhan data (Okfalisa dkk., 2022). Rumus yang digunakan adalah sebagai berikut:

$$Error - Rate = \frac{FP + FN}{TP + FP + TN + FN}$$

Rumus 3.12 Nilai *Error-Rate* Model Prediksi