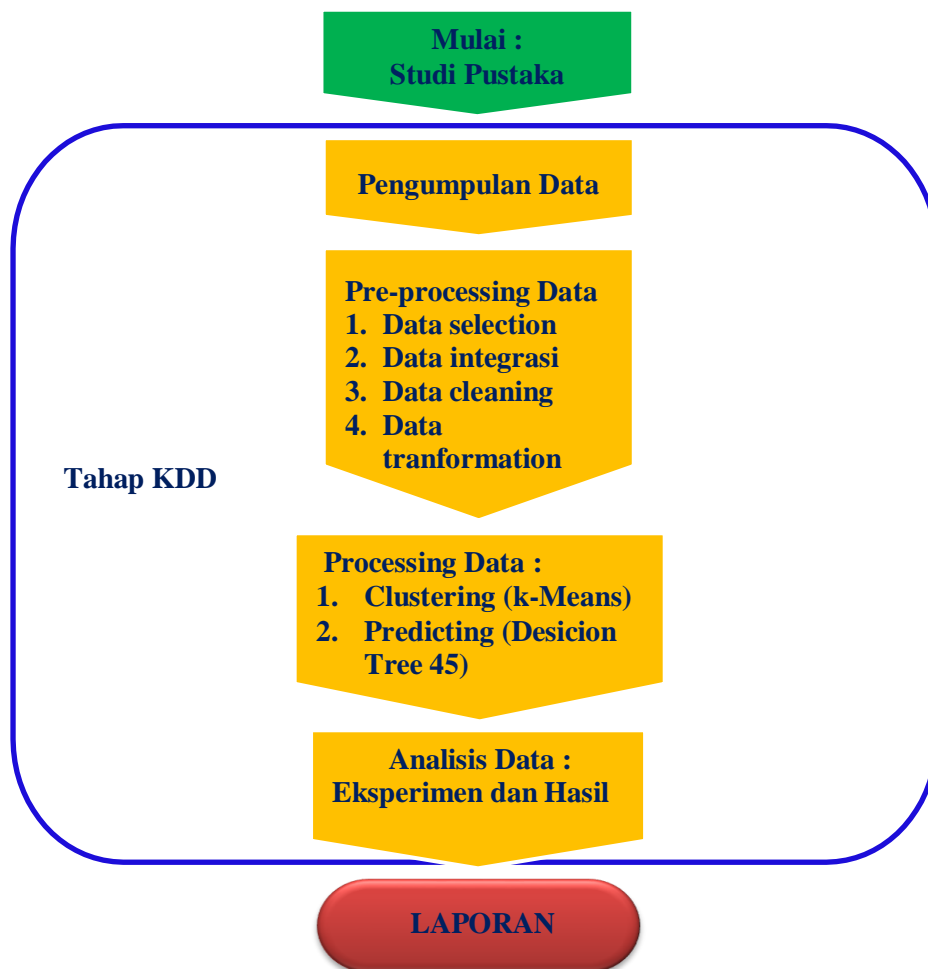


## BAB III METODOLOGI PENELITIAN

### 3.1 Metode Penelitian

Pada Bab ini akan membahas langkah-langkah dari proses penelitian yang akan dilaksanakan, dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset untuk memudahkan penelitian dan dapat berjalan dengan sistematis sehingga memenuhi tujuan yang diinginkan maka dibuat alur tahapan penelitian secara detail dilihat pada gambar 3.1 :



Gambar 3.1. Alur Tahapan Penelitian

Berikut penjelasan dari gambar 3.1 :

1. Tahap pertama hal yang dilakukan adalah membuat suatu identifikasi masalah mengenai prediksi kinerja siswa menggunakan teknik data mining.
2. Tahap selanjutnya melakukan studi pustaka yang berkaitan dengan penelitian-penelitian sebelumnya tentang penggunaan teknik data mining serta untuk menguatkan penelitian ini berdasarkan teori yang digunakan.
3. Tahap selanjutnya pengumpulan data, data yang digunakan berasal dari datasetpublic yaitu data siswa di <https://www.kaggle.com/spscientist/student-performance-in-exams/data>.
4. Tahap selanjutnya : memasuki tahap Penemuan Pengetahuan dalam Basis Data (KDD) Pre - processing yang terdiri dari beberapa tahap :
  - a. Database ; merupakan dasar dalam data mining sebagai sumber data. Database yang kita dapatkan dalam penelitian ini adalah datasetpublic yaitu data siswa di <https://www.kaggle.com/spscientist/student-performance-in-exams/data>.
  - b. Data Integration. Dari dataset yang kita dapatkan selanjutnya data-data yang ada di dalamnya akan mengalami proses penggabungan atau penyatuan dua atau lebih data dengan menyamakan semua formatnya. Hasil integrasi data ini akan disimpan di dalam gudang data (Data Warehouse) dan akan melalui tahap selanjutnya.
  - c. Data Cleaning. Setelah melakukan integrasi data, selanjutnya dilakukan lagi tahap data cleaning atau pembersihan data. Jadi, data-data yang ada akan mengalami pembersihan terhadap data yang tidak diperlukan atau mungkin bernilai null. Dalam tahap ini bukan berarti langsung melakukan pembersihan dengan menghapus beberapa data, tapi tepatnya akan dilakukan pengecekan terhadap semua data yang ada. Dalam tahap ini juga akan terjadi data transformation, yaitu melakukan perubahan-pengubahan data agar sesuai dengan hasil data mining yang diinginkan.
  - d. Data Selection. Setelah melalui tahap data cleaning termasuk data transformation, selanjutnya data akan diseleksi dengan memilih atribut apa yang diinginkan. Jadi, pada tahap ini akan dilakukan penyeleksian data-data yang relevan dengan apa yang diinginkan (Task-relevant Data).

- e. Transformasi data , dalam tahap ini diketahui bahwa dataset yang ada tidak memiliki label/class sehingga dataset terkategori sebagai unsupervised learning dataset, sehingga karena dataset tersebut akan digunakan sebagai dataset prediksi maka dalam tahap selanjutnya dataset tersebut harus dirubah dari **Unsupervised learning dataset** menjadi **Supervised dataset** dengan menggunakan algoritma untuk kasus-kasus klustering, asosiasi dan dimensionality reduction dan juga algoritma untuk tugas-tugas klasifikasi dan regresi, misal dalam kasus object recognition, predictive analysis dan sentiment analysis.
5. Tahap selanjutnya adalah Processing data, dalam tahap ini kita menentukan metode dan pengujian metode yang akan digunakan untuk melakukan clustering (algoritma **K-Means** ) dan prediksi (algoritma **Decision Tree C 4.5**) dengan teknik data mining untuk menemukan hasil akurasi terbaik.
  6. Tahap selanjutnya melakukan eksperimen atau pengujian untuk menemukan hasil akurasi terbaik,
  7. Tahap terakhir yaitu membuat kesimpulan berupa laporan (*Darmajaya 2021, 2021*) dari hasil pengujian yang mengacu pada rumusan masalah serta tujuan penelitian.

### 3.2. Alat dan Bahan

#### a. Alat

Penelitian ini menggunakan perangkat keras Laptop Lenovo IdeaPad S145 Intel Core i5 8th Gen up to 3,2 Ghz, RAM 20 Gigabyte, VGA Nvidia 2Gb dengan Ruang penyimpanan berupa Solid State Disk kapasitas 265 Gb, sedangkan perangkat lunak yang terlibat adalah system operasi Windows 10, Microsoft Excel dan Rapidminer Studio sebagai aplikasi yang digunakan untuk pengolahan data.

#### b. Bahan

Sebagai bahan penelitian : adalah data set yang berasal dari datasetpublic yaitu data siswa di <https://www.kaggle.com/spscientist/student-performance-in-exams/data>, berisi data sejumlah 1000 data record, berupa data siswa dari berbagai latar belakang yang beraneka ragam, yang terbagi menjadi 8 atribut yang terdiri dari :

*Tabel 3.2 Atribut Dataset*

No	Attribute	Domain	type	Keterangan
1	Gender	M/F	Binominal	Male/Female
2	Race/Ethnicity	A/B/C	Polynominal	Group A, Group B, Group C
3	Parental Level of Education	A/B/C/D/E	Polynominal	Associate's degree, Bachelor's degree, High school, Master's degree, some college
4	Lunch	Y/N	Binominal	Standard, Free/reduced
5	Test Preparation Course	Y/N	Binominal	None, Complete
6	Math Score	1 - 100	Integer	Number
7	Reading Score	1 - 100	Integer	Number
8	Writing Score	1 - 100	Integer	Number

Gender, Race/Ethnicity, Parental Level Education, Lunch, Test Preparation Course, Math Score, Reading Score, Writing Score yang semuanya merupakan faktor-faktor yang dimiliki oleh para siswa dengan penjelasan arti setiap grup dalam kolom sebagai berikut :

1. Gender (jenis kelamin) : menentukan jenis kelamin siswa (laki-laki/perempuan).
2. Race/Ethnicity (ras) : menentukan ras siswa (grup A, grup B, grup C).

3. Parental Level Education (tingkat pendidikan orang tua) : menentukan kualifikasi pendidikan tertinggi dari setiap orang tua dari setiap siswa.
4. Lunch (tipe\_makan siang) : standar/dikurangi, jenis paket makan siang yang dipilih untuk siswa.
5. Test Preparation Course (kursus test\_persiapan) : menentukan apakah kursus persiapan ujian diselesaikan oleh siswa atau tidak.
6. Math\_score: menentukan skor dalam matematika (variabel target kami).
7. Reading\_score: menentukan skor dalam membaca.
8. Writing\_score: menentukan skor secara tertulis.

Tabel dataset menunjukkan karakter data pada penelitian ini, dimana data tersebut belum memiliki *classlabel* yang berarti data ini merupakan data Unsupervised dataset. Sehingga sebelum dataset prediksi diperlukan langkah clustering dataset terlebih dahulu untuk menentukan kelas labelnya.