

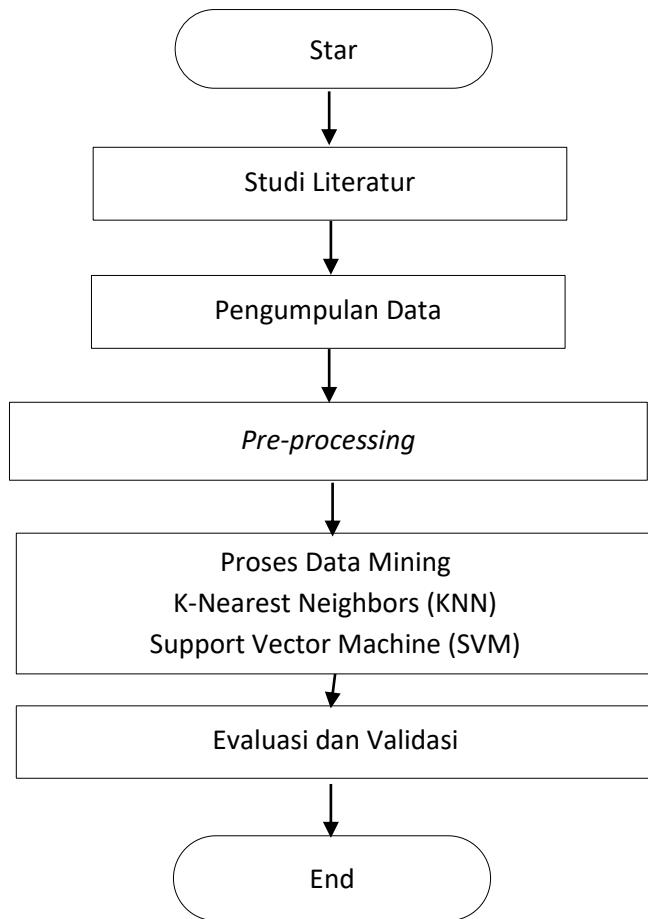
BAB III

METODOLOGI PENELITIAN

Bab ini dimulai dengan memperkenalkan skema penelitian dan alat yang digunakan dalam penelitian ini, yang merupakan alat Machine Learning (ML) open-source yang siap pakai. Selain itu, dijelaskan pula mengenai dataset yang digunakan dalam penelitian. Penelitian ini bertujuan untuk membandingkan performa antara algoritma *K-Nearest Neighbors* (KNN) dan *algoritma Dalam* bab ini, akan dibahas mengenai rancangan penelitian, data dan sumber data yang digunakan, teknik pengolahan data, serta analisis data untuk mendapatkan hasil penelitian yang akurat.

3.1 Tahapan 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 dalam memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan, maka dibuat alur dalam tahapan penelitian yang akan dilakukan sebagai berikut:



Gambar 3. 1 Alur Dalam Tahapan Penelitian Prediksi Kinerja Pegawai dengan Algoritma KNN dan SVM

1. Studi Literatur:

Tahap ini melibatkan pengumpulan dan analisis literatur yang relevan dengan topik penelitian. Tujuannya adalah untuk memahami konsep-konsep dasar, teknik-teknik yang telah digunakan sebelumnya, dan tantangan yang mungkin ditemui dalam penelitian serupa. Dengan melakukan studi literatur, peneliti dapat membangun landasan teori yang kuat untuk penelitiannya.

2. Pengumpulan Data:

Setelah memiliki pemahaman yang baik tentang topik penelitian, langkah selanjutnya adalah mengumpulkan data yang diperlukan. Data dapat diperoleh dari adatabase sistem kepegawaian Dinas Pendidikan dan Kebudayaan Provinsi Lampung, adapun data tersebut berjumlah 215 data pegawai yang memiliki atribut (pegawai, usia, lama bekerja, pendidikan, jabatan, kehadiran, kepuasan kerja dan kinerja sebagai label). Kualitas dan relevansi data yang dikumpulkan akan sangat mempengaruhi hasil akhir dari proses data mining. Kualitas dan relevansi data yang dikumpulkan sangat mempengaruhi hasil akhir dari proses data mining. Oleh karena itu, pemilihan data yang tepat menjadi tahap penting dalam menunjang keakuratan analisis model klasifikasi. Tabel 3.1 berikut menyajikan deskripsi atribut yang digunakan dalam penelitian ini:

Tabel 3.1 Keterangan Atribut

Atribut	Keterangan
Usia	Usia pegawai dalam tahun dan bulan (format: “XX THN YY BLN”)
Lama Bekerja	Durasi pengalaman kerja pegawai dalam tahun dan bulan
Pendidikan	Tingkat pendidikan terakhir yang dicapai (contoh: Sarjana, Pascasarjana, Doktor)
Jabatan	Posisi jabatan pegawai dalam organisasi atau instansi
Kehadiran (%)	Persentase kehadiran pegawai selama periode tertentu
Kepuasan Kerja (1-5)	Skala tingkat kepuasan kerja pegawai (1 = sangat tidak puas, 5 = sangat puas)
Rating Perilaku Kerja	Penilaian terhadap sikap dan perilaku kerja pegawai (misalnya: Baik, Sangat Baik)
Rating Hasil Kinerja	Evaluasi terhadap hasil kerja pegawai berdasarkan ekspektasi organisasi
Rating Perilaku Kerja.1	Penilaian tambahan atau lanjutan terhadap perilaku kerja pegawai
Predikat Kinerja Periodik	Predikat kinerja pegawai berdasarkan penilaian periodik (misal: Baik, Sangat Baik)
Kinerja	Status pencapaian target kerja pegawai (Tercapai / Tidak Tercapai)

3. Pre-processing:

Tahap ini merupakan salah satu tahap yang paling penting dalam proses data mining. Data yang telah dikumpulkan biasanya tidak siap untuk langsung diproses. Oleh karena itu, perlu dilakukan pembersihan data (data cleaning) untuk menghilangkan data yang tidak konsisten

4. Proses Data Mining:

Inilah inti dari proses data mining. Pada tahap ini, berbagai teknik dan algoritma data mining diterapkan pada data yang telah diproses. Beberapa teknik yang umum digunakan antara lain:

1. K-Nearest Neighbors (KNN)

- a. Menghitung jarak antara data yang ingin diklasifikasikan dengan data lain dalam dataset (misalnya, menggunakan jarak Euclidean).
- b. Memilih **k** tetangga terdekat berdasarkan jarak yang dihitung.
- c. Menentukan kelas data berdasarkan mayoritas kelas dari **k** tetangga terdekat.

2. Support Vector Machine (SVM)

- a. Membangun **hyperplane** terbaik yang memisahkan data berdasarkan kelasnya dengan margin maksimum.
- b. Jika data tidak dapat dipisahkan secara linear, digunakan fungsi kernel untuk memetakan data ke dimensi yang lebih tinggi agar dapat dipisahkan.
- c. Data baru diklasifikasikan berdasarkan sisi hyperplane tempat data tersebut berada.

Kedua algoritma ini diterapkan untuk melakukan klasifikasi atau prediksi berdasarkan pola yang ditemukan dalam database.

5. Evaluasi dan Validasi:

Setelah mendapatkan hasil dari proses data mining, langkah selanjutnya adalah mengevaluasi dan memvalidasi model yang telah dibangun. Tujuannya adalah untuk mengukur kinerja model dan memastikan bahwa model tersebut dapat generalisasi ke data baru. Beberapa metrik yang umum digunakan untuk evaluasi model antara lain akurasi, presisi, recall, dan F1-score.

3.2 Alat dan Bahan

Adapun alat dan bahan yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Hardware

Kebutuhan perangkat keras (hardware) yang digunakan:

1. Laptop Asus Vivobook gen 11
2. Processor: Intel(R) Core(TM) i3-1005G1 CPU @ 1.20GHz 1.19 GHz.
RAM 8,00 GB (7,74 GB usable).
3. System Type: 64-bit operating system, x64-based processor, Harddisk dengan kapasitas 237 GB.

2. Software

Kebutuhan perangkat lunak (software) yang digunakan:

1. Rapid Miner 10.1
2. Microsoft Excel 2019
3. Microsoft Windows 11