BAB I
PENDAHULUAN

A. Latar belakang

Pemanfaatan teknologi prediktif dalam sistem informasi telah menjadi salah satu kebutuhan bagi organisasi yang mengelola data baik dalam jumlah kecil maupun besar, termasuk institusi pendidikan yang ingin mengoptimalkan proses pengambilan keputusan berbasis data. Salah satu implementasi pentingnya ialah model prediksi berbasis machine learning, yang memungkinkan sistem mengestimasi nilai suatu variabel berdasarkan pola pada data historis Sukmaningtyas dkk,(2024).

SMKN 1 Sinunukan sebagai sekolah kejuruan memiliki data siswa telah terdokumentasi dalam sistem administrasi sekolah, termasuk data terkait kondisi ekonomi keluarga dan catatan kedisiplinan. Dalam disiplin ilmu Sistem Informasi, atribut tersebut dapat diolah sebagai fitur (features) dalam model prediksi. Pemanfaatan variabel tersebut sangat relevan dalam membangun sistem pendukung keputusan dalam memprediksi prestasi siswa.

Selain itu, berdasarkan data resmi SMKN 1 Sinunukan tahun ajaran 2024/2025, jumlah peserta didik (PD) tercatat sebanyak 374 siswa, terdiri dari 222 siswa laki-laki dan 152 siswa perempuan. Data nilai yang diperoleh dari dokumen rapor menunjukkan bahwa distribusi capaian prestasi siswa berada pada variasi yang cukup lebar. Berdasarkan rekapitulasi nilai rata-rata siswa pada salah satu kompetensi keahlian, sekitar 30% siswa memiliki nilai rata-rata pada rentang 70–75, 65% siswa memiliki nilai rata-rata pada rentang 76–85, dan hanya sekitar 5% siswa yang memiliki nilai rata-rata di atas 85. Hal ini menunjukkan bahwa sebagian besar siswa memiliki capaian sedang hingga tinggi, namun tetap terdapat kelompok siswa yang membutuhkan perhatian akademik tambahan. Sekolah ini menjadi lokasi penelitian karena memiliki data administratif siswa yang lengkap meliputi nilai rapor, catatan kedisiplinan (kehadiran), serta informasi status ekonomi keluarga yang dibutuhkan untuk pemodelan prediksi prestasi belajar.

UIN IMAM BONJOL PADANG

Selain faktor ekonomi, tingkat kedisiplinan siswa juga menunjukkan variasi yang cukup signifikan. Berdasarkan data kehadiran siswa yang tercatat secara administratif setiap bulan, terlihat adanya perbedaan tingkat kehadiran antar siswa, baik dari sisi jumlah kehadiran maupun frekuensi ketidakhadiran selama periode pembelajaran. Sebagian siswa menunjukkan tingkat kehadiran yang relatif stabil, sementara sebagian lainnya menunjukkan ketidakkonsistenan kehadiran antar bulan. Perbedaan tersebut menunjukkan adanya variasi tingkat kedisiplinan siswa dalam mengikuti proses belajar mengajar, yang berpotensi memengaruhi capaian prestasi belajar.

Selain faktor eksternal dan perilaku tersebut, faktor internal siswa juga memiliki peran penting dalam menentukan prestasi belajar, salah satunya adalah motivasi belajar. Motivasi belajar merupakan dorongan internal yang memengaruhi tingkat usaha, ketekunan, dan konsistensi siswa dalam mengikuti proses pembelajaran. Penelitian oleh Utami dkk,(2026) menunjukkan bahwa motivasi belajar berpengaruh positif dan signifikan terhadap prestasi siswa berdasarkan analisis regresi linier berganda. Selain itu, penelitian oleh Juliani dkk,(2024) menemukan bahwa motivasi belajar bersama dengan minat belajar mampu menjelaskan lebih dari 50% variasi hasil belajar siswa. Penelitian lain oleh Rachmahdiyah dkk,(2023) juga menunjukkan bahwa motivasi berprestasi memiliki pengaruh dominan terhadap capaian akademik siswa. Hal ini menunjukkan bahwa motivasi belajar tidak hanya sebagai faktor pendukung, tetapi juga sebagai variabel penting yang secara langsung memengaruhi prestasi siswa. Oleh karena itu, memasukkan variabel motivasi belajar ke dalam model prediksi menjadi penting untuk menghasilkan analisis yang lebih komprehensif dan akurat.

Oleh karena itu variasi kedisiplinan dan motivasi belajar tersebut memperkuat urgensi penerapan model prediktif berbasis data mining untuk menganalisis keterkaitan antara status ekonomi, kedisiplinan, dan prestasi siswa secara lebih objektif dan berbasis data.

UIN IMAM BONJOL PADANG

Pada banyak penelitian sebelumnya, regresi linier berganda digunakan sebagai model dasar dalam tugas prediksi karena model ini sederhana, mudah digunakan, dan hasilnya mudah dijelaskan Saragih dkk.(2023), Pradita, (2022). Namun, karakteristik data siswa dapat menunjukkan hubungan antarvariabel yang beragam, sehingga hubungan antara variabel prediktor dan prestasi siswa tidak selalu dapat dijelaskan secara optimal hanya dengan pendekatan linier. Beberapa penelitian menunjukkan adanya variasi hubungan antarvariabel serta perbedaan karakteristik data siswa Lyu, (2023).

Kebutuhan akan algoritma yang mampu menangkap pola data yang kompleks ini menjadikan perbandingan regresi linier berganda dan support vector regression relevan dalam konteks SMK. Sekolah kejuruan memiliki data siswa yang heterogen, berbeda antar program keahlian, serta cenderung membentuk hubungan fitur yang tidak sepenuhnya linier. Kondisi tersebut menuntut pengujian apakah regresi linier berganda yang bekerja dengan asumsi linier masih mampu memberikan akurasi prediksi yang memadai. Sebaliknya, support vector regression (SVR) dapat digunakan untuk memodelkan pola data yang kompleks melalui penggunaan fungsi kernel. Oleh karena itu, perbandingan kedua metode ini diperlukan untuk menentukan algoritma yang paling sesuai dengan karakteristik data siswa di SMKN 1 Sinunukan.

UIN IMAM BONJOL PADANG

Support Vector Regression (SVR) dapat digunakan sebagai salah satu metode regresi dalam prediksi prestasi siswa.Ying & Ma (2024)menerapkan SVR dalam pendekatan regresi untuk memprediksi prestasi siswa dan menunjukkan bahwa pendekatan regresi memberikan kinerja prediksi yang lebih baik dibandingkan pendekatan klasifikasi pada dataset yang digunakan. Sejumlah penelitian menunjukkan bahwa SVR dapat memberikan performa prediksi yang baik pada data pendidikan. Hutajulu & Wulandhari (2024) membandingkan SVR, Random Forest Regression, AdaBoost Regression, dan XGBoost Regression dalam prediksi prestasi mahasiswa. Hasil penelitian menunjukkan bahwa SVR memperoleh

performa terbaik pada dataset mahasiswa Informatics Management (MI) dengan nilai R^2 sebesar 0,58,sehingga menjadi kandidat kuat untuk dibandingkan dengan regresi linier berganda.

Namun, hingga saat ini masih terdapat kesenjangan penelitian, terutama di Indonesia dan khususnya pada sekolah kejuruan diantaranya, sebagian besar penelitian hanya menggunakan satu model tanpa melakukan penerapan atau perbandingan langsung antar algoritma, fitur administratif seperti status ekonomi dan kedisiplinan jarang digunakan secara bersamaan dalam model prediksi, berdasarkan literatur yang di telaah dalam penelitian ini , berdasarkan literatur yang di telaah dalam penelitian ini, belum di temukan penelitian yang secara khusus menerapkan atau membandingkan regresi linier berganda dan SVR untuk memprediksi prestasi siswa berdasarkan status ekonomi dan kedisiplinan dan motivasi belajar.

Berdasarkan kondisi tersebut, penelitian ini menerapkan data mining menggunakan regresi linier berganda dan support vector regression untuk memprediksi prestasi siswa berdasarkan status ekonomi dan kedisiplinan dan motivasi belajar. Hasil penelitian diharapkan untuk digunakan sebagai referensi dalam menentukan metode prediksi yang paling sesuai dengan karakteristik data siswa.

UIN IMAM BONJOL PADANG

B. Rumusan masalah

1. Bagaimana proses penerapan data mining dilakukan untuk membangun model prediksi prestasi siswa di SMKN 1 Sinunukan?
2. Bagaimana penerapan metode regresi linier berganda dalam melakukan prediksi berdasarkan status ekonomi, kedisiplinan dan motivasi belajar?
3. Bagaimana penerapan metode *support vector regression* (SVR) dalam melakukan prediksi berdasarkan status ekonomi, kedisiplinan dan motivasi belajar?
4. Bagaimanakah performa regresi linier berganda dan *support vector regression* (SVR) dalam melakukan prediksi berdasarkan status ekonomi, kedisiplinan dan motivasi belajar?

C. Batasan masalah

Agar penelitian ini lebih terarah, maka ditetapkan batasan masalah sebagai berikut:

1. Variabel yang digunakan meliputi
    a. Status ekonomi siswa (penghasilan orang tua) sebagai variabel input (X1)
    b. Kedisiplinan siswa (kehadiran) sebagai variabel input (X2)
    c. Motivasi belajar sebagai variabel input (X3)
    d. Prestasi siswa (nilai rapor) sebagai variabel target (Y)
2. Data status ekonomi dan kedisiplinan diperoleh dari data administratif sekolah, sedangkan data motivasi belajar diperoleh melalui penyebaran kuesioner kepada siswa menggunakan skala Likert.
3. Model prediksi yang dibandingkan hanya menggunakan dua algoritma meliputi regresi linier berganda dan *support vector regression* (SVR).
4. Ruang lingkup penelitian berfokus pada pemodelan prediksi, dan sampai pada pembuatan aplikasi menggunakan streamlit yang bersifat prototipe sederhana.
5. Evaluasi performa model terbatas pada tiga metrik meliputi *mean squared error* (MSE), *root mean squared error* (RMSE), *coefficient of determination* (R^2).
6. Seluruh proses perhitungan pelatihan model dan evaluasi dilakukan menggunakan *google colab* dengan *library python* seperti *pandas*, *numPy*, dan *scikit-learn*.

D. Tujuan penelitian

1. Mendeskripsikan proses pembangunan model prediksi prestasi siswa menggunakan pendekatan data mining.
2. Menerapkan metode regresi linier berganda dalam memprediksi prestasi siswa berdasarkan status ekonomi, kedisiplinan dan motivasi belajar.

UIN IMAM BONJOL PADANG

3. Menerapkan metode *support vector regression* (SVR) dalam memprediksi prestasi siswa berdasarkan status ekonomi, kedisiplinan dan motivasi belajar.
4. Mengetahui performa metode regresi linier berganda dan *support vector regression* (SVR) dalam melakukan prediksi prestasi siswa berdasarkan status ekonomi, kedisiplinan, dan motivasi belajar menggunakan metrik evaluasi MSE, RMSE, dan R^2.

E. Manfaat penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Manfaat teoretis
   a. Memberikan kontribusi terhadap pengembangan ilmu Sistem Informasi, khususnya dalam penerapan *data mining* untuk membangun model prediksi berdasarkan variabel Status Ekonomi, Kedisiplinan, dan Motivasi Belajar..
   b. Referensi akademik mengenai performa antara algoritma regresi linier berganda dan *support vector regression* dalam konteks pengolahan data pendidikan pada sistem informasi.
2. Manfaat praktis
   a. Rekomendasi algoritma terbaik yang dapat digunakan sebagai komponen analitik dalam sistem informasi sekolah untuk memprediksi prestasi siswa.
   b. Mendukung pihak sekolah dalam memperoleh informasi prediktif sebagai bahan pendukung pemantauan Prestasi Siswa, berdasarkan Status Ekonomi, Kedisiplinan, dan Motivasi Belajar.
   c. Memberikan informasi tambahan melalui fitur peringatan dini (*early warning*) dan rekomendasi tindak lanjut yang dapat digunakan sebagai bahan pendukung dalam melakukan pemantauan kondisi siswa.

UIN IMAM BONJOL PADANG