

Project Probabilistic Reasoning (Penalaran Probabilistik): Prediksi Performa Mahasiswa Berbasis Bayesian Network

Performa akademik mahasiswa dipengaruhi oleh berbagai faktor yang saling berkaitan dan memiliki unsur ketidakpastian. Institusi pendidikan tinggi menghadapi tantangan dalam mengidentifikasi faktor-faktor yang memengaruhi performa akademik mahasiswa, sehingga dapat memberikan dukungan dan intervensi yang tepat. Pendekatan penalaran probabilistik, khususnya Bayesian Network, menawarkan kerangka yang kuat untuk memodelkan ketidakpastian dan hubungan sebab-akibat antar variabel yang memengaruhi performa akademik mahasiswa.

Tujuan Project

Mengembangkan model prediksi berbasis penalaran probabilistik yang dapat menganalisis dan memprediksi performa akademik mahasiswa (GPA dan GradeClass) berdasarkan berbagai faktor demografis, sosial, dan perilaku belajar menggunakan dataset `Student_performance_data.csv`, serta menyediakan dasar untuk rekomendasi intervensi dan dukungan yang dipersonalisasi.

Spesifikasi Project

1. Dataset dan Variabel

- **Sumber Data:** <https://www.kaggle.com/code/joelknapp/student-performance-analysis/notebook>
- **Variabel dalam Dataset:**
 - **Demografis:** StudentID, Age, Gender, Ethnicity, ParentalEducation
 - **Akademik & Perilaku Belajar:** StudyTimeWeekly, Absences, Tutoring
 - **Dukungan & Aktivitas:** ParentalSupport, Extracurricular, Sports, Music, Volunteering
 - **Performa (Output):** GPA, GradeClass
- **Persyaratan Preprocessing Dataset:**
 - Analisis statistik deskriptif untuk memahami distribusi data
 - Penanganan missing values (jika ada)
 - Normalisasi atau diskretisasi variabel jika diperlukan
 - Analisis korelasi untuk mengidentifikasi hubungan awal antar variabel
 - Visualisasi distribusi data untuk memahami pola

2. Model Penalaran Probabilistik

1. Bayesian Network (Wajib)

- Pemodelan struktur jaringan (directed acyclic graph) yang menggambarkan hubungan antar variabel dalam dataset
- Estimasi parameter (conditional probability tables) untuk setiap node dalam jaringan
- Implementasi inferensi probabilistik untuk memprediksi GPA dan GradeClass
- Analisis kausal dan hubungan antar variabel

2. Metode Tambahan (Pilih Minimal Satu)

- Hidden Markov Models
- Naive Bayes Classifier
- Probabilistic Neural Networks
- Markov Random Fields
- Dynamic Bayesian Networks
- Gaussian Mixture Models

3. Fitur Wajib:

- Struktur grafik yang menunjukkan hubungan antar variabel dalam dataset
- Tabel probabilitas bersyarat untuk setiap node
- Kemampuan inferensi untuk memprediksi GPA dan GradeClass berdasarkan faktor lain
- Analisis sensitivitas untuk mengidentifikasi faktor-faktor kunci performa
- Kemampuan untuk memberi rekomendasi peningkatan performa

4. Fitur Tambahan (Pilih Minimal Dua):

- Parameter learning dari data (berbanding dengan expert judgement)
- Structure learning (menemukan struktur grafik optimal dari data)
- Analisis intervensi (what-if analysis)
- Clustering mahasiswa berdasarkan profil performa
- Penalaran dengan bukti parsial (missing values)
- Visualisasi interaktif hubungan antar variable

3. Implementasi Sistem

1. Pengembangan Model:

- Pemrosesan dan persiapan dataset `Student_performance_data.csv`
- Pembagian data menjadi training dan testing set (misalnya 80:20)
- Pengembangan struktur Bayesian Network berdasarkan variabel dalam dataset
- Pembelajaran parameter model dari data
- Validasi model dengan metrik yang sesuai

2. Antarmuka dan Visualisasi:

- Dashboard untuk menampilkan hasil prediksi dan analisis
- Visualisasi struktur Bayesian Network dari variabel dataset
- Kemampuan untuk memasukkan data mahasiswa baru
- Tampilan prediksi GPA dan GradeClass berdasarkan faktor-faktor input
- Visualisasi faktor-faktor yang paling berpengaruh terhadap performa

4. Evaluasi dan Validasi

1. Metrik Evaluasi Wajib:

- Mean Absolute Error (MAE) atau Root Mean Squared Error (RMSE) untuk prediksi GPA

- Akurasi, presisi, dan recall untuk prediksi GradeClass (jika digunakan sebagai kategori)
 - Area Under ROC Curve (AUC) untuk klasifikasi performa
 - Kalibrasi prediksi probabilistik
2. **Metode Validasi Wajib:**
- k-fold Cross-validation (k=5 atau k=10)
 - Sensitivity analysis terhadap parameter model
 - Comparison dengan baseline models (e.g., regresi linear, random forest)

Yang dikumpulkan

1. **Implementasi Model**
 - Source code lengkap dengan dokumentasi
 - Dataset yang digunakan (raw dan preprocessed)
 - File model yang dapat dijalankan untuk prediksi
 - Panduan instalasi dan penggunaan
2. **Dashboard Analisis Performa**
 - Antarmuka untuk input data mahasiswa baru
 - Visualisasi struktur model dan hubungan antar variabel dari dataset
 - Visualisasi prediksi performa (GPA dan GradeClass)
 - Kemampuan what-if analysis untuk simulasi intervensi
3. **Laporan Teknis (15-20 halaman)**
 - Landasan teori penalaran probabilistik
 - Analisis komprehensif dataset Student_performance_data.csv
 - Metodologi pengembangan model
 - Deskripsi struktur Bayesian Network dan justifikasi
 - Metode estimasi parameter dan inferensi
 - Hasil evaluasi dan analisis kinerja
 - Analisis faktor-faktor yang paling memengaruhi GPA dan GradeClass
 - Diskusi kelebihan dan keterbatasan model
 - Rekomendasi praktis untuk meningkatkan performa mahasiswa
4. **Presentasi (15-20 menit)**
 - Penjelasan dataset dan pendekatan model
 - Demo dashboard dan visualisasi
 - Hasil analisis dan insights utama
 - Diskusi implikasi praktis untuk institusi pendidikan

Kriteria Penilaian Detil

1. **Kualitas Model Probabilistik (35%)**
 - Ketepatan struktur Bayesian Network (10%)
 - Akurasi estimasi parameter (10%)
 - Implementasi algoritma inferensi (10%)
 - Kinerja prediktif model pada GPA dan GradeClass (5%)
2. **Implementasi Fitur Tambahan (20%)**
 - Kualitas implementasi fitur tambahan (10%)

- Integrasi dengan model utama (5%)
- Nilai tambah untuk analisis dan prediksi (5%)
- 3. **Antarmuka dan Visualisasi (15%)**
 - Kegunaan dashboard (5%)
 - Kualitas visualisasi struktur dan hasil (5%)
 - Interaktivitas dan user experience (5%)
- 4. **Evaluasi dan Analisis (15%)**
 - Ketepatan metodologi evaluasi (5%)
 - Kedalaman analisis hasil (5%)
 - Interpretasi dan insights dari dataset (5%)
- 5. **Dokumentasi dan Presentasi (15%)**
 - Kelengkapan dokumentasi teknis (5%)
 - Kualitas presentasi (5%)
 - Kemampuan menjawab pertanyaan (5%)