

**PROPOSAL**  
**KLASIFIKASI MAHASISWA BERDASARKAN TINGKAT**  
**PEMAHAMAN MATERI MATA KULIAH DENGAN MENGGUNAKAN**  
**ALGORITMA C4.5 DAN K-MEANS CLUSTERING**



**RISET INFORMATIKA D081**

**Dosen pengampu :**

Dr. Basuki Rahmat, S.Si., M.T.

**Dibuat Oleh:**

Athaya Aqilah 21081010306

**PROGRAM STUDI TEKNIK INFORMATIKA**  
**FAKULTAS ILMU KOMPUTER**  
**UPN "VETERAN" JAWA TIMUR**

**2024**

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pendidikan tinggi merupakan salah satu pilar utama dalam pengembangan sumber daya manusia yang berkualitas. Di era globalisasi dan kemajuan teknologi informasi saat ini, tantangan dalam dunia pendidikan semakin kompleks. Salah satu tantangan yang paling signifikan adalah memastikan bahwa mahasiswa tidak hanya hadir dalam perkuliahan, tetapi juga memahami materi yang diajarkan dengan baik. Pemahaman yang mendalam terhadap materi kuliah sangat penting, karena berpengaruh langsung terhadap prestasi akademik mahasiswa dan kemampuan mereka untuk menerapkan pengetahuan dalam konteks nyata.

Namun, fenomena rendahnya tingkat pemahaman mahasiswa terhadap materi kuliah sering kali terjadi. Berbagai penelitian menunjukkan bahwa banyak mahasiswa yang mengalami kesulitan dalam memahami konsep-konsep dasar dalam mata kuliah tertentu, yang dapat disebabkan oleh berbagai faktor, seperti metode pengajaran yang kurang efektif, kurangnya motivasi, dan keterbatasan sumber daya belajar. Hal ini menimbulkan kebutuhan mendesak untuk mengidentifikasi dan menganalisis faktor-faktor yang mempengaruhi pemahaman mahasiswa agar dapat dilakukan intervensi yang tepat.

Dalam konteks ini, teknologi analisis data menjadi alat yang sangat berguna. Dengan memanfaatkan algoritma pembelajaran mesin, institusi pendidikan dapat menganalisis data mahasiswa secara lebih mendalam untuk mendapatkan wawasan tentang tingkat pemahaman mereka. Dua metode yang menarik untuk diterapkan dalam penelitian ini adalah algoritma C4.5 dan K-Means Clustering. Algoritma C4.5 adalah salah satu algoritma pembelajaran mesin yang digunakan untuk membangun model klasifikasi berdasarkan data pelatihan. Metode ini mampu menghasilkan pohon keputusan yang jelas dan mudah dipahami, sehingga memungkinkan pengajar untuk mengidentifikasi karakteristik mahasiswa dengan pemahaman rendah.

Di sisi lain, K-Means Clustering merupakan metode pengelompokan yang efektif untuk mengelompokkan mahasiswa berdasarkan kesamaan karakteristik tertentu. Dengan menggunakan K-Means Clustering, peneliti dapat mengidentifikasi kelompok-kelompok mahasiswa dengan pola pemahaman yang serupa, sehingga dapat merancang strategi pembelajaran yang lebih sesuai dengan kebutuhan masing-masing kelompok.

Meskipun telah banyak penelitian yang menggunakan algoritma C4.5 dan K-Means Clustering secara terpisah dalam konteks pendidikan, masih sedikit studi yang mengintegrasikan kedua metode ini untuk menganalisis tingkat pemahaman mahasiswa terhadap materi kuliah secara bersamaan. Penelitian ini bertujuan untuk mengisi celah tersebut dengan melakukan klasifikasi mahasiswa berdasarkan tingkat pemahaman mereka menggunakan algoritma C4.5 dan mengelompokkan mereka menggunakan K-Means Clustering.

Dengan demikian, hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan bagi dunia pendidikan, terutama dalam upaya meningkatkan kualitas pembelajaran di perguruan tinggi. Selain itu, penelitian ini juga diharapkan dapat memberikan rekomendasi bagi dosen dan institusi pendidikan untuk merancang metode

pengajaran yang lebih efektif berdasarkan analisis data pemahaman mahasiswa. Pengembangan latar belakang di atas memberikan konteks yang lebih mendalam tentang pentingnya penelitian ini serta relevansi penggunaan algoritma C4.5 dan K-Means Clustering dalam meningkatkan pemahaman mahasiswa terhadap materi kuliah.

## **1.2 Rumusan Masalah**

1. Bagaimana tingkat pemahaman mahasiswa terhadap materi kuliah?
2. Apa saja faktor yang mempengaruhi tingkat pemahaman mahasiswa?
3. Bagaimana efektivitas algoritma C4.5 dan K-Means Clustering dalam mengklasifikasikan dan mengelompokkan mahasiswa berdasarkan tingkat pemahaman mereka?

## **1.3 Tujuan**

1. tingkat pemahaman mahasiswa terhadap materi kuliah.
2. Menganalisis faktor-faktor yang mempengaruhi tingkat pemahaman mahasiswa.
3. Mengimplementasikan algoritma C4.5 dan K-Means Clustering untuk mengklasifikasikan dan mengelompokkan mahasiswa berdasarkan tingkat pemahaman mereka.

## **1.4 Manfaat**

- Memberikan wawasan bagi dosen dan institusi pendidikan tentang faktor-faktor yang mempengaruhi pemahaman mahasiswa.
- Menyediakan model klasifikasi yang dapat digunakan untuk meningkatkan proses pembelajaran.
- Menjadi referensi bagi penelitian selanjutnya terkait analisis data pendidikan.

## **BAB II**

### **Tinjauan Pustaka**

#### **2.1 Teori Pembelajaran**

Pembelajaran adalah proses yang kompleks dan multifaset yang melibatkan perubahan perilaku, pengetahuan, atau keterampilan sebagai hasil dari pengalaman. Dalam konteks pendidikan tinggi, pemahaman materi kuliah menjadi kunci utama dalam mencapai tujuan pembelajaran. Berbagai teori pembelajaran telah dikembangkan untuk menjelaskan bagaimana individu belajar dan memahami informasi baru. Teori konstruktivisme, misalnya, menekankan bahwa pembelajaran adalah proses aktif di mana mahasiswa membangun pengetahuan mereka sendiri melalui interaksi dengan lingkungan dan pengalaman belajar yang relevan. Sementara itu, teori behaviorisme berfokus pada hubungan antara stimulus dan respons, di mana penguatan positif dapat meningkatkan motivasi mahasiswa untuk belajar. Memahami teori-teori ini penting untuk merancang strategi pengajaran yang efektif yang dapat meningkatkan pemahaman mahasiswa terhadap materi kuliah.

#### **2.2 Algoritma C4.5**

C4.5 adalah algoritma pembelajaran mesin yang dikembangkan oleh Ross Quinlan untuk membangun model klasifikasi berbasis pohon keputusan. Algoritma ini merupakan pengembangan dari algoritma ID3 dan memiliki kemampuan untuk menangani atribut numerik dan kategorikal dengan baik. Proses kerja C4.5 dimulai dengan memilih atribut terbaik untuk membagi dataset menjadi subset berdasarkan kriteria tertentu, seperti informasi gain atau gain ratio. Dengan cara ini, C4.5 menghasilkan pohon keputusan yang dapat digunakan untuk memprediksi kelas dari data baru berdasarkan nilai atributnya. Salah satu keunggulan dari algoritma ini adalah kemampuannya untuk menghasilkan model yang transparan dan mudah dipahami, sehingga pengguna dapat dengan mudah menginterpretasikan hasil klasifikasi. Dalam konteks penelitian ini, C4.5 akan digunakan untuk mengklasifikasikan mahasiswa berdasarkan tingkat pemahaman mereka terhadap materi kuliah, memberikan wawasan tentang karakteristik mahasiswa yang memiliki pemahaman rendah.

#### **2.3 K-Means Clustering**

K-Means Clustering adalah metode pengelompokan yang populer dalam analisis data, digunakan untuk mengelompokkan data ke dalam sejumlah cluster berdasarkan kesamaan fitur-fitur tertentu. Proses K-Means dimulai dengan menentukan jumlah cluster yang diinginkan ( $k$ ) dan kemudian menginisialisasi centroid untuk setiap cluster secara acak. Selanjutnya, setiap data akan dikelompokkan ke dalam cluster terdekat berdasarkan jarak Euclidean dari centroid masing-masing. Proses ini diulang hingga tidak ada perubahan dalam pengelompokan atau hingga kriteria konvergensi lainnya terpenuhi. Kelebihan dari K-Means adalah kemampuannya dalam menangani dataset besar dengan cepat dan efisien, serta kemudahan interpretasi hasilnya. Dalam penelitian ini, K-Means Clustering akan digunakan untuk mengelompokkan mahasiswa berdasarkan karakteristik tertentu yang mempengaruhi pemahaman mereka terhadap materi kuliah, sehingga memungkinkan identifikasi pola-pola dalam data yang dapat membantu dalam merancang strategi pembelajaran yang lebih efektif.

## **2.4 Penelitian Terkait**

Sejumlah penelitian sebelumnya telah mengeksplorasi penggunaan algoritma C4.5 dan K-Means Clustering dalam konteks pendidikan, namun masih terdapat kekurangan dalam integrasi kedua metode tersebut secara bersamaan. Beberapa studi menggunakan C4.5 untuk menganalisis faktor-faktor yang mempengaruhi prestasi akademik mahasiswa tanpa mempertimbangkan pengelompokan berdasarkan karakteristik mahasiswa itu sendiri. Di sisi lain, penelitian yang menggunakan K-Means sering kali tidak melakukan klasifikasi mendalam terhadap tingkat pemahaman mahasiswa setelah pengelompokan dilakukan. Dengan demikian, penelitian ini bertujuan untuk mengisi celah tersebut dengan mengintegrasikan kedua metode dalam satu kerangka analisis untuk memberikan gambaran yang lebih komprehensif tentang tingkat pemahaman mahasiswa terhadap materi kuliah serta faktor-faktor yang mempengaruhinya. Pengembangan di atas memberikan penjelasan yang lebih mendalam mengenai teori pembelajaran, algoritma C4.5, K-Means Clustering, serta relevansi penelitian.

## **BAB III**

### **METODELOGI**

#### **3.1 Desain Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif dengan desain deskriptif. Pendekatan kuantitatif dipilih karena memungkinkan pengumpulan dan analisis data numerik yang dapat diukur secara objektif. Desain deskriptif memungkinkan peneliti untuk menggambarkan karakteristik populasi mahasiswa serta menganalisis hubungan antara variabel-variabel yang mempengaruhi tingkat pemahaman mereka terhadap materi kuliah. Dengan menggunakan desain ini, penelitian bertujuan untuk memberikan gambaran yang jelas mengenai tingkat pemahaman mahasiswa serta faktor-faktor yang berkontribusi terhadap pemahaman tersebut, sehingga hasilnya dapat digunakan untuk merumuskan strategi pembelajaran yang lebih efektif.

#### **3.2 Populasi dan Sampel**

Populasi dalam penelitian ini adalah mahasiswa UPN Veteran Jawa Timur. Untuk memastikan representativitas, sampel diambil secara acak dari berbagai jurusan dan program studi, sehingga mencakup beragam latar belakang pendidikan dan pengalaman belajar. Pengambilan sampel dilakukan dengan menggunakan kuisinier yang dikirimkan kepada mahasiswa. Dengan cara ini, penelitian dapat mencerminkan keragaman mahasiswa dan memberikan hasil yang lebih generalizable mengenai tingkat pemahaman mereka terhadap materi kuliah.

#### **3.3 Pengumpulan Data**

Data dikumpulkan melalui kuesioner yang dirancang khusus untuk mengukur berbagai aspek yang mempengaruhi pemahaman mahasiswa terhadap materi kuliah. Kuesioner terdiri dari beberapa bagian: bagian pertama mencakup informasi demografis mahasiswa, seperti usia, jenis kelamin, dan jurusan; bagian kedua berisi pertanyaan tentang motivasi belajar, metode pembelajaran yang digunakan dosen, serta ketersediaan sumber daya belajar. Setiap pertanyaan dalam kuesioner menggunakan skala Likert untuk mengukur tingkat setuju atau tidak setuju responden terhadap pernyataan yang diberikan. Kuesioner ini disebarluaskan secara online melalui platform survei untuk memudahkan pengumpulan data dan menjangkau lebih banyak responden.

#### **3.4 Proses Cleaning Data**

Setelah pengumpulan data selesai, langkah selanjutnya adalah melakukan cleaning data untuk memastikan bahwa dataset siap untuk analisis. Proses ini meliputi beberapa tahap penting: pertama, identifikasi dan penghapusan entri duplikat untuk menghindari bias dalam analisis; kedua, penanganan missing values dengan cara mengisi nilai yang hilang menggunakan metode imputasi seperti rata-rata atau median dari kolom tersebut; ketiga, transformasi data jika diperlukan, seperti mengubah variabel kategorikal menjadi format numerik agar dapat digunakan dalam analisis statistik. Proses cleaning data sangat penting untuk memastikan integritas dan kualitas data yang akan dianalisis.

### **3.5 Analisis Data**

Analisis data dilakukan dalam dua tahap menggunakan dua metode berbeda: algoritma C4.5 untuk klasifikasi dan K-Means Clustering untuk pengelompokan. Pada tahap pertama, algoritma C4.5 diterapkan untuk membangun model klasifikasi tingkat pemahaman mahasiswa berdasarkan atribut-atribut yang telah ditentukan sebelumnya dalam kuesioner. Model ini akan membantu dalam mengidentifikasi karakteristik mahasiswa dengan pemahaman rendah dan memberikan wawasan tentang faktor-faktor penyebabnya. Pada tahap kedua, K-Means Clustering digunakan untuk mengelompokkan mahasiswa ke dalam kategori berdasarkan kesamaan karakteristik mereka. Hasil dari analisis ini akan memberikan gambaran tentang pola-pola dalam data yang dapat digunakan untuk merancang strategi pembelajaran yang lebih sesuai dengan kebutuhan kelompok-kelompok mahasiswa tertentu.

### **3.6 Alat yang Digunakan**

Analisis data akan dilakukan menggunakan Visual Studio Code (VSCode) sebagai Integrated Development Environment (IDE) utama bersama dengan bahasa pemrograman Python. VSCode dipilih karena kemampuannya dalam mendukung berbagai bahasa pemrograman serta fitur-fitur seperti debugging, kontrol versi, dan ekstensi yang memudahkan pengembangan aplikasi berbasis data. Dalam konteks analisis ini, Python akan digunakan bersama dengan pustaka-pustaka penting seperti Pandas untuk manipulasi data, NumPy untuk perhitungan numerik, Matplotlib dan Seaborn untuk visualisasi data, serta Scikit-learn untuk penerapan algoritma C4.5 dan K-Means Clustering. Dengan kombinasi alat-alat ini, peneliti dapat melakukan eksperimen dengan berbagai parameter algoritma dan mendapatkan hasil analisis yang akurat serta mudah dipahami. Pengembangan di atas memberikan penjelasan lebih mendalam mengenai desain penelitian, populasi dan sampel, pengumpulan data, proses cleaning data, analisis data, serta alat yang digunakan dalam penelitian Anda dengan menekankan penggunaan Visual Studio Code dan bahasa Python.