

BAB I PENDAHULUAN

1.1 Latar Belakang

Dunia memerlukan pendidikan, dari peradaban kuno hingga era revolusi industri 4.0 sekarang ini pendidikan merupakan hak azasi bagi setiap individu, masyarakat dan negara. Pendidikan merupakan suatu kebutuhan yang penting bagi kehidupan setiap individu, sehingga akses terhadap layanan pendidikan merupakan keniscayaan yang mutlak ada diseluruh dunia. Kebutuhan pendidikan bagi masyarakat sangat menjadi perhatian lembaga-lembaga dunia, pada tahun 2019 UNICEF mengadvokasi pendidikan dasar yang berkualitas tinggi dan ramah anak untuk semua, sejalan dengan ambisi Agenda Pendidikan Global 2030. Untuk memenuhi visi yang dibawa oleh Kerangka Aksi Pendidikan 2030 dan SDG4, UNICEF merilis Strategi Pendidikan 2019–2030 nya sendiri, 'Setiap Anak Belajar', yang menguraikan tiga tujuan berbeda:

1. Akses yang setara ke kesempatan belajar;
2. Peningkatan pembelajaran dan keterampilan untuk semua; dan
3. Peningkatan pembelajaran dan perlindungan bagi anak-anak dalam konteks darurat dan rapuh.

Bagi negara-negara berkembang, pendidikan sangat penting untuk memenuhi tujuan pembangunan berkelanjutan, kemajuan teknologi informasi berkembang sedemikian pesatnya yang secara tidak langsung membuat akses masyarakat terhadap layanan pendidikan menjadi semakin maju. Namun terjadi juga di beberapa negara berkembang bahwa tingkat kinerja atau prestasi siswa sangat dipengaruhi faktor – faktor tertentu yang berkait erat terhadap siswa itu sendiri antara lain faktor : gender (jenis kelamin), ethnicity(suku bangsa), parental level of education, lunch(pemenuhan kebutuhan makan yang berkait dengan tingkat gizi), test preparation course, math score, reading score, writing score. Umumnya bahwa sejumlah dataset siswa yang berkaitan dengan student performance kurang lebih tersusun dari atribut – atribut tersebut. Salah satu data set yang ada di situs www.Kaggle.com - “Students-performance-in-exams data set. Data set ini memuat semua atribut di atas dan terdiri dari kurang lebih seribu data, sehingga dari data set

tersebut kita dapat melakukan kegiatan data mining untuk menghasilkan suatu prediksi. Decision Tree merupakan algoritma yang mampu mengolah data set “student performance” yang tersusun dari beberapa kelompok atribut : gender, ethnicity, parental level of education, lunch, test preparation course, math score, reading score, writing score yang dapat digunakan untuk memprediksi kinerja siswa, dimana hal tersebut sangat mempengaruhi keberhasilan atau kegagalan seorang siswa.

Dari data set yang ada telah dilakukan prediksi Students Performance dengan menggunakan metode decision tree dan metode split data secara murni ternyata hanya menghasilkan akurasi prediksi sebesar 59,14% dan 79 %. Angka 59,14% dan 79 % sesungguhnya kurang signifikan. Oleh karenanya diperlukan suatu terobosan metode data mining guna meningkatkan hasil akurasi prediksi student performance. Dalam tulisan ini, kita hanya akan fokus pada prediksi. Penelitian ini akan menggunakan dataset yang tersedia dalam pusat kumpulan data internasional. Sementara algoritma yang digunakan adalah algoritma clustering dan algoritma klasifikasi dalam data mining yaitu Decision Tree C4.5 dengan dua metode optimasi data mining diharapkan menjadi kombinasi terbaik untuk mendapatkan akurasi terbaik.

1.2 Ruang Lingkup

Peneliti membatasi ruang lingkup penelitian yang hanya dibatasi pada clustering dan prediksi kinerja/prestasi siswa menggunakan teknik data mining dengan algoritma K-Means dan Algoritma Decision Tree C4.5. Data yang digunakan dalam penelitian ini adalah dataset public Kaggle yaitu data “Student Performance in Exam”, serta penggunaan tool Rapid Miner(RapidMiner, 2012).

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah bagaimana cara meningkatkan akurasi prediksi kinerja/prestasi siswa dengan menggunakan algoritma K-Means dan Algoritma Decision Tree C4.5 dengan menggunakan tool Rapid Miner.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah : mengklustering dataset publik (*student performs in exams*) untuk menghasilkan kelompok siswa dengan nilai tinggi dan nilai rendah, dan menghasilkan model algoritma yang memiliki akurasi tinggi untuk akurasi prediksi kinerja/prestasi siswa dengan menggunakan algoritma K-Means dan algoritma Decision Tree C4.5 menggunakan tool Rapid Miner.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah :

1. Penelitian ini merupakan penelitian Data Mining yang memberikan gambaran bagaimana mensegmentasi/ clustering dan prediksi kinerja/prestasi siswa dengan penggunaan algoritma K-Means dan Decision Tree C4.5 dengan menggunakan aplikasi Rapid Miner, dan menguji beberapa variabel yang diperoleh dari dataset publik.
2. Bagi institusi khususnya program studi Magister Teknik Informatika dapat digunakan sebagai referensi ilmiah dalam penelitian penerapan data mining.
3. Serta memberikan masukan dalam ilmu pengetahuan untuk penelitian selanjutnya dalam pengembangan ilmu pengetahuan tentang data mining.

1.6 Sistematika Penulisan

Laporan penelitian tesis ini terdiri dari lima bab dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Dalam bab ini tercantum antara lain latar belakang, ruang lingkup, rumusan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini memuat tentang teori-teori dan penelitian yang terkait yang mendukung penelitian yang akan dilakukan oleh penulis/peneliti. Penelitian yang menggunakan

analisis statistik, bab ini memuat kerangka pikir dan hipotesis (bila diperlukan).

BAB III METODOLOGI PENELITIAN

Dalam bab ini berisi objek penelitian, alat dan bahan, metode yang dilakukan dalam pengumpulan data, prosedur penelitian, pengukuran variabel dan metode analisis (metode-metode pendekatan penyelesaian permasalahan yang dipakai dan metode analisis data).

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini disajikan hasil dari penelitian, implementasi penelitian, analisis penelitian dan pembahasan penelitian. Hasil dan implementasi dapat berupa gambar alat/program dan aplikasinya. Analisis dan pembahasan berupa hasil pengolahan data.

BAB V SIMPULAN DAN SARAN

Dalam bab ini disajikan simpulan dan saran dari hasil pembahasan.