

Penerapan Data Mining Algoritma Decision Tree Berbasis PSO

Sri Farida Utami

Magister Sistem Informasi, STMIK LIKMI, Bandung, Indonesia

Email: utamigeulis@gmail.com

Abstrak—Penelitian ini bertujuan untuk membuat prediksi prestasi belajar siswa berdasarkan nilai akhir semester menggunakan metode data mining dengan algoritma Decision Tree berbasis Particle Swarm Optimization. Data penelitian yang digunakan adalah nilai semester dua siswa SDIT KAMPUS MASJID AL-UMMAH berjumlah 24 siswa. Hasil penelitian menunjukkan bahwa analisis prediksi menggunakan decision tree memiliki akurasi sebesar 81,67 %, sedangkan analisis prediksi menggunakan Decision tree berbasis PSO memiliki tingkat akurasi 96,67%. Berdasarkan hasil tersebut bisa disimpulkan bahwa metode Decision tree berbasis PSO lebih baik dibandingkan dengan metode Decision tree.

Kata Kunci: Data Mining, Decision Tree, PSO

1. PENDAHULUAN

Pendidikan adalah salah satu faktor kemajuan dan kemandirian bangsa. Semakin maju pendidikan suatu bangsa, maka akan semakin maju dan mandiri bangsa tersebut. Melalui pendidikan para generasi penerus bangsa dibentuk kualitasnya. Pendidikan nasional berfungsi mengembangkan kemampuan dan membentuk watak serta peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa, bertujuan untuk mengembangkan potensi peserta didik agar menjadi manusia yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, cakap, kreatif, mandiri dan menjadi warga negara yang demokratis serta bertanggungjawab [1]. Seiring dengan tujuan pendidikan tersebut maka kualitas dan manajemen pembelajaran di sekolah atau lembaga pendidikan formal perlu ditingkatkan. Salah satu indikator kualitas dan manajemen sekolah atau lembaga pendidikan dapat dilihat dari prestasi belajar peserta didik.

SDIT Kampus Masjid Al-Ummah yang berdiri sejak tahun 1999 merupakan salah satu satuan pendidikan yang berusaha menyusun dan mengembangkan Kurikulum berdasarkan Standar yang telah ditentukan. Kurikulum SDIT AL-Ummah disusun bersama-sama oleh guru, komite sekolah, dan narasumber dengan kepala sekolah sebagai ketua merangkap anggota.

Penyusunan Kurikulum SDIT Al-Ummah didasarkan pada analisis internal dan eksternal dengan mengacu kepada analisis SWOT (Strength, Weakness, Opportunities, and Threats /Kekuatan, Kelemahan, peluang dan tantangan/Ancaman). Hal tersebut dilakukan sebagai salah satu bentuk usaha untuk meningkatkan kualitas lulusan, dibutuhkan juga kerjasama dan sinkronisasi yang baik antara pemerintah, manajemen sekolah, kompetensi guru, orang tua dan peserta didik itu sendiri. Beberapa elemen tersebut sangat berpengaruh baik secara langsung maupun tidak langsung terhadap peningkatan mutu peserta didik, yang pada akhirnya akan berimbas pada peningkatan kualitas pendidikan. Prestasi belajar peserta didik dipengaruhi oleh banyak faktor. Salah satu faktor penting dalam pendidikan agar prestasi belajar siswa meningkat adalah manajemen dan kurikulum sekolah. Semakin baik manajemen pembelajaran yang ditawarkan sekolah kepada peserta didik semakin besar pula kemungkinan prestasi belajar peserta didik akan baik.

Menyadari pentingnya kualitas pembelajaran dalam rangka mencerdaskan kehidupan bangsa, maka pemerintah bersama sektor swasta sama-sama telah dan terus berupaya mewujudkan amanat tersebut melalui berbagai usaha pembangunan pendidikan yang lebih berkualitas antara lain melalui pengembangan dan perbaikan kurikulum dan sistem evaluasi, perbaikan sarana pendidikan, pengembangan dan pengadaan materi bahan ajar, serta pelatihan bagi guru dan tenaga kependidikan lainnya. Tetapi pada kenyataannya upaya pemerintah tersebut belum terlalu signifikan dalam upaya meningkatkan kualitas pendidikan. [1]. Proses pembelajaran merupakan salah satu faktor penentu keberhasilan pendidikan di sekolah. Rendahnya kualitas pendidikan adalah akibat dari rendahnya kualitas proses pembelajaran yang dilakukan di sekolah.

Banyak hal yang belum dilakukan oleh lembaga pendidikan salah satu diantaranya adalah melakukan antisipasi potensi peserta didik yang mengalami hambatan atau berprestasi dalam belajar. Hal ini penting karena lembaga pendidikan yang lebih awal menyadari potensi setiap peserta didik yang mungkin mengalami hambatan belajar, lembaga pendidikan atau sekolah dapat melakukan langkah-langkah antisipatif.

Berdasarkan uraian di atas, maka penelitian ini bertujuan untuk membuat prediksi sejak dini terhadap siswa yang berpotensi mengalami hambatan dalam belajarnya, sehingga dapat dilakukan langkah-langkah kongkret dari sekolah untuk menghindari resiko dari kemungkinan tidak naik kelas bahkan peserta didik yang rawan DO dari sekolah. Langkah yang bisa diambil sekolah setelah mengetahui adanya siswa yang berpotensi tidak berprestasi adalah dengan melakukan pendampingan khusus terhadap siswa tersebut. Harapan akhirnya adalah semua siswa dari berbagai latar belakang faktor masing-masing dapat maksimal dalam prestasi belajar siswanya.

Belajar adalah suatu proses atau upaya yang dilakukan oleh setiap orang demi menghasilkan suatu perubahan. Perubahan tersebut antara lain tingkah laku, pengetahuan dan keterampilan sikap dan nilai positif sebagai pengalaman dari berbagai materi pembelajaran yang telah dipelajari di sekolah, namun demikian kadang, peserta didik jenuh dan akhirnya memilih untuk tidak belajar. Untuk itulah mereka perlu diberi dorongan dari segala pihak. Pengaruh motivasi terhadap prestasi belajar peserta didik sangat besar karena itu orang – orang di sekitar lingkungan peserta didik baik itu di sekolah maupun diluar rumah hendaknya memberi motivasi yang tepat.

Faktor disiplin dalam pengelolaan ajaran adalah hal yang sangat penting. Tanpa kesadaran akan kebutuhan untuk menerapkan aturan yang telah ditetapkan maka mengajar mungkin tidak mencapai target maksimal. Seorang pelajar harus

memiliki disiplin untuk melakukan latihan yang memperkuat dirinya untuk selalu terbiasa patuh dan mempertinggi kekuatan pengendalian diri. Disiplin yang muncul dari kesadaran sendiri akan lebih terpacu dan tahan lama dibandingkan dengan disiplin yang timbul karena pengawasan dari orang lain.

Disiplin dapat tumbuh dan dibina melalui latihan, pendidikan atau pembiasaan yang harus dimulai sejak dalam lingkungan keluarga di rumah mulai pada masa kanak-kanak dan terus tumbuh berkembang sehingga menjadi disiplin yang semakin kuat. Yang telah disebutkan oleh Tulus [2] bahwa dengan disiplin yang muncul untuk kesadaran diri, peserta didik dapat meraih keberhasilan dalam belajarnya, tanpa disiplin yang baik suasana sekolah dan juga kelas menjadi kurang kondusif bagi kegiatan pembelajaran, secara positif disiplin memberi dukungan lingkungan yang tenang dan tertib bagi proses pembelajaran, disiplin merupakan jalan bagi peserta didik untuk sukses dalam belajar dan kelak ketika bekerja karena kesadaran pentingnya norma, aturan, kepatuhan dan ketaatan merupakan dasar menuju kesuksesan seseorang

Penelitian-penelitian tentang faktor yang mempengaruhi prestasi belajar peserta didik sebelumnya, banyak yang menggunakan pengolahan data statistika. Pada penelitian ini, peneliti menggunakan data mining. Data mining merupakan cabang ilmu baru di bidang komputer, cukup banyak penerapan yang dapat matematika dan pengolahan citra) membuat penerapan data mining menjadi makin luas. Alasan utama mengapa data mining sangat menarik perhatian industri informasi dalam beberapa tahun belakangan ini adalah karena tersedianya data dalam jumlah yang besar dan semakin besarnya kebutuhan untuk mengubah data tersebut menjadi informasi dan pengetahuan yang berguna.

Definisi sederhana dari data mining adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di database yang besar. Dalam jurnal ilmiah, data mining juga dikenal dengan nama Knowledge Discovery in Databases (KDD). Hal itu ditunjang kekayaan dan keanekaragaman berbagai bidang ilmu (arti- kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran/berjumlah besar, informasi inilah yang nantinya sangat berguna untuk pengembangan ficial intelligence, database, statistik, pemodelan.

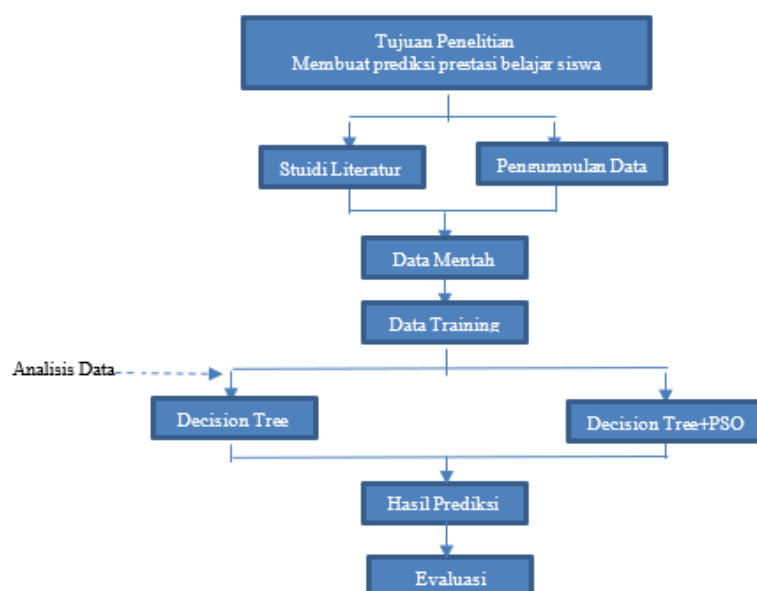
Lembaga pendidikan pada umumnya dan Sekolah dasar khususnya adalah institusi yang memiliki data explosion yang belum dimanfaatkan dengan baik untuk sesuatu yang berguna. Data yang biasa dimiliki sekolah dalam jumlah besar dan akan selalu bertambah setiap tahun.

Data - data tersebut belum termanfaatkan dengan baik sebagai bahan untuk dijadikan pertimbangan, kajian dan penelitian untuk membuat suatu informasi yang berguna bagi sekolah. Oleh karena itu, dengan adanya data mining bisa menjadi salah satu solusi untuk menambang tumpukan data-data tersebut untuk membuat nilai atau informasi yang lebih berguna. Berdasarkan uraian di atas, penelitian ini menggunakan metode data mining untuk membuat suatu prediksi tentang prestasi belajar siswa dengan memanfaatkan tumpukan data yang telah dimiliki sekolah tersebut.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif. Tempat Penelitian ini adalah SDIT KAMPUS MASJID AL-UMMAH. Target/subjek penelitian adalah siswa kelas 3 SDIT KAMPUS MASJID AL-UMMAH Pelajaran 2017/2018 sejumlah 24 siswa.

Prosedur penelitian menggunakan tahapan-tahapan KDD (Knowledge Data Discov- ery). Tahapan-tahapannya adalah: (1) Data Selection yaitu tahapan ini dilakukan untuk memilih data yang sesuai dengan variabel yang dibutuhkan dalam penelitian; (2) Pre Prosesing Data (Cleaning) sebelum proses data mining.



Gambar 1. Kerangka pikir penelitian

Dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang data yang terduplikasi, memeriksa data yang imajiner, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi); (3) Transformation, coding adalah proses trans- formasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining; (4) Analisis data; (5) Pattern Evaluation, merupakan tahapan evaluasi untuk

mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan sumber data yang ada. Kerangka pikir yang digunakan dalam penelitian ini dapat dilihat pada gambar 1.

2.1 Teknik Analisis Data

Penelitian ini menggunakan teknik Deci- sion Tree dan Decision Tree Berbasis PSO untuk melakukan prediksi prestasi belajar siswa SDIT KAMPUS MASJID AL-UMMAH berdasarkan Nilai Siswa

2.2 Decision Tree

Decision Tree akan memperlihatkan faktor-faktor kemungkinan (probabilitas) yang akan mempengaruhi alternatif-alternatif prestasi belajar siswa, disertai dengan prediksi hasil akhir yang akan didapat bila faktor-faktor dalam Decision Tree terpenuhi. Decision Tree akan mengubah data kedalam bentuk visual berupa diagram pohon keputusan dan aturan-aturan keputusan.

Data dalam Decision Tree dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan tree. Salah satu atribut yang merupakan atribut yang menyatakan data solusi per-item data yang disebut dengan target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan instance. Alur proses analisis dalam decision tree adalah mengubah bentuk data (table) menjadi model tree, mengubah model tree menjadi rule dan menyederhanakan rule (pruning). Data yang diambil dalam penelitian ini adalah populasi sejumlah 24 siswa akan digunakan untuk membuat model prediksi Decision Tree. Model yang telah dibuat kemudian akan dihitung tingkat akurasi prediksinya.

Ada banyak metodologi Data Mining, salah satu yang populer adalah Pohon Keputusan (Decision Tree). Pohon keputusan merupakan salah satu metode klasifikasi yang sangat menarik yang melibatkan konstruksi pohon keputusan yang terdiri dari node keputusan yang di hubungkan dengan cabang-cabang dari simpul akar sampai ke node daun (akhir). Pada node keputusan atribut akan diuji, dan setiap hasil akan menghasilkan cabang. Setiap cabang akan diarahkan ke node lain atau ke node akhir untuk menghasilkan suatu keputusan[3].

Menurut Maimon pohon keputusan adalah salah satu metode klasifikasi yang dinyatakan sebagai partisi rekursif. Pohon keputusan terdiri dari node yang membentuk pohon yang berakar, semua node memiliki satu masukan. Node yang keluar disebut node tes. Node yang lain disebut node keputusan atau sering disebut node daun. Setiap simpul internal membagi dua atau lebih sub-ruang sesuai dengan kategori atribut dan akan dipartisi sesuai dengan nilai kategori kasus. Kasus-kasus tersebut membentuk pohon keputusan yang menghasilkan problem solving[4].

Klasifikasi data mining adalah penempatan objek-objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Klasifikasi banyak digunakan untuk mendiksi kelas pada suatu label tertentu, yaitu dengan mengklasifikasi data (membangun model) berdasarkan training set dan nilai-nilai (label kelas) dalam mengklasifikasikan atribut tertentu dan menggunakannya dalam mengklasifikasi data yang baru.

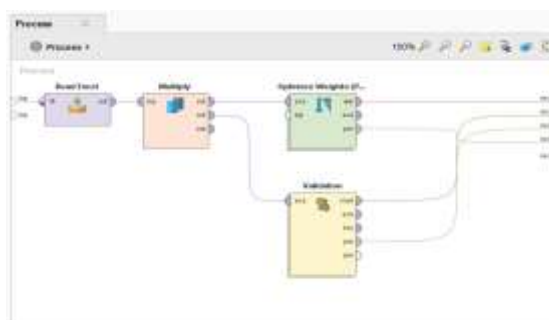
Pohon keputusan biasanya digunakan untuk mendapatkan informasi untuk tujuan pengambilan sebuah keputusan. Pohon keputusan dimulai dengan sebuah root node (titik awal) yang digunakan oleh user untuk mengambil tindakan. Berdasarkan node root ini, user memecahkan leaf node sesuai dengan algoritma decision tree.

Hasil akhir dari penyusunan node root dan leaf node tersebut adalah sebuah pohon keputusan dengan setiap cabangnya menunjukkan kemungkinan skenario dari keputusan yang diambil serta hasilnya. Konsep pohon keputusan adalah mengubah data menjadi sebuah pohon keputusan (decision tree) dan aturan-aturan keputusan (rule).

3. ANALISA DAN PEMBAHASAN

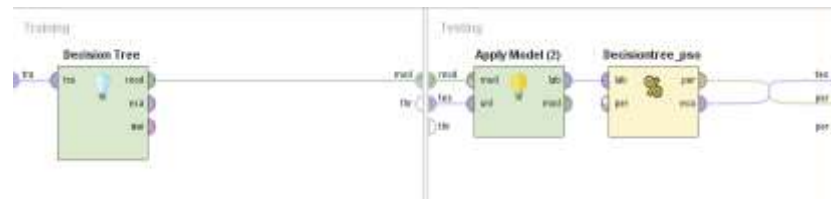
Hasil penelitian berdasarkan tahapan- tahapan proses dalam KDD (*Knowledge Data Discovery*) sebagai berikut: (1) *Data Selection*, tahapan ini dilakukan untuk memilih data yang sesuai dengan variabel yang dibutuhkan dalam penelitian. Caranya adalah dengan memilih atau menentukan atribut-atribut data mana yang akan digunakan dalam penelitian dari sekelompok data operasional yang ada. (2) *Pre- processing/Cleaning*, proses *cleaning* tersebut dilakukan terhadap keseluruhan data yang diteliti yang berjumlah 24 siswa. Setelah di- lakukan proses *cleaning* data sejumlah 24, di- hasilkan data bersih sebanyak 24 record data yang digunakan untuk proses analisis berikut- nya; (3) *Transformation*, tahap ini menghasil- kan satu *recordset* data yang siap untuk analisis data; (4) Analisis data.

Pertama yang dilakukan adalah menguji dataset pada aplikasi rapidminer dengan algoritma Decision tree dan Decision Tree berbasis PSO dan hasilnya dapat dilihat pada gambar 2.



Gambar 2. Proses Sumber Hasil penelitian 2019

Didalam validasi terbagi menjadi dua frame yaitu Training dan Testing. Pada frame training peneliti menggunakan algoritma Decision tree C4.5 berbasis PSO yang kemudian dihungkan dengan application model dan performance pada frame testing. Gambar 2. merupakan gambar pengujian algoritma Decision tree C4.5 berbasis PSO menggunakan rapidminer pada frame training dan testing



Gambar 3. Pengujian algoritma Decision tree berbasis PSO
Sumber hasil penelitian 2019



Gambar 4. Pengujian algoritma Decision tree
Sumber hasil penelitian 2019

a. Confusion matrix

Pada Tabel 4 menunjukkan hasil dari confusion matrix metode C4.5 berbasis PSO.
accuracy: 96.67% +/- 10.00% (micro average: 93.79%)

Tabel 1. Hasil Dari Confusion Matrix Metode C4.5 Berbasis PSO

	true TIDAK	true YA	class precision
pred. TIDAK	14	1	93.33%
pred. YA	0	9	100.00%
class recall	100.00%	90.00%	

Sumber hasil penelitian 2019

Keterangan : Dari tabel diatas diperoleh jumlah True Negative (TN) sebanyak 14 sebagai false dan sesuai dengan klasifikasi, False Positive(FP) sebanya 1 diprediksi false ternyata hasil prediksi True Postive(TP) sebanyak 9 diklasifikasi sebagai True dan sesuai dengan prediksi yang digunakan menggunakan cross validation dan False Negative(FN) sebanyak 0 dan klasifikasinya true ternyata hasil klasifikasinya false. Tingkat akurasi yang diperoleh menggunakan algoritma C4.5 berbasis PSO dapat dilihat dari tabel 4.1 adalah sebesar 96.67 % dan dapat dihitung untuk mencari nilai accuracy, sensitivity, specificity, PPV dan NPV pada persamaai dibawah ini:

$$\text{Accurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{9 + 14}{9 + 14 + 1 + 0} = 96.67\%$$

b. Receiver Operating Curve (ROC)



Gambar 5. Kurva ROC metode Decision tree berbasis PSO
Sumber hasil penelitian 2019

c. Confusion matrix

Pada Tabel 4 menunjukkan hasil dari confusion matrix metode Decision tree.
accuracy: 96.67% +/- 10.00% (micro average: 93.79%)

Tabel 2. Hasil Dari Confusion Matrix Metode Decision tree

	true TIDAK	true YA	class precision
pred. TIDAK	14	1	93.33%
pred. YA	0	9	100.00%
class recall	100.00%	90.00%	

Sumber hasil penelitian 2019

d. Receiver Operating Curve (ROC)



Gambar 6. Kurva ROC metode Decision tree

Sumber hasil penelitian 2019

Dari hasil pengujian / eksperimen yang telah dilakukan terhadap data siswa dapat disimpulkan bahwa data dengan nilai Confidence = 0.25 , Minimal Gain = 0.1 dan Populasi Size = 5 memiliki akurasi tertinggi sebesar 96.67% dan AUC sebesar 10.00%.

Perbandingan akurasi klasifikasi antara metode Decision tree, dan Decision tree berbasis PSO adalah sebagai berikut: Tingkat akurasi atau signifikansi variabel prediktor terhadap prestasi belajar siswa dengan menggunakan metode Decision tree berbasis PSO sebesar 96.67%. Tingkat akurasi atau signifikansi variabel prediktor terhadap prestasi belajar siswa dengan menggunakan metode Decision tree sebesar 81,67%.

4. KESIMPULAN

Berdasarkan analisis data menggunakan decision tree data mining untuk memprediksi prestasi belajar siswa berdasarkan nilai menggunakan metode data mining diperoleh hasil sebagai berikut : (1) variabel motivasi adalah variabel yang menentukan potensi seorang siswa berhasil atau tidak prestasi belajarnya di waktu yang akan datang. Hal ini menunjukkan bahwa aspek knowledge atau kecerdasan siswa sangat berpengaruh terhadap keberhasilan belajarnya. Sebaliknya, walaupun siswa terprediksi secara knowledge kurang tetapi dengan motivasi tinggi tetap bisa berprestasi minimal pada kategori B atau C.

REFERENCES

- [1] Kurniawan, Deny. (2008). Regresi linier (linear regression). Vienna, Austria: R Foundation for Statistical Computing
- [2] Depdiknas. (2003). Undang-Undang Republik Indonesia Nomor 20 Tahun 2003, tentang Sistem Pendidikan Nasional
- [3] Depdiknas. (1990). Peraturan Pemerintah RI No. 29, Tahun 1990, tentang Pendidikan Menengah
- [4] Kass G.V. (1980). An exploratory technique for investigating large quantities of categorical data. Appl. Statist. 29 No.2, pp 119-127
- [5] Larose, & Daniel T. (2005). Discovering knowledge in data: an introduction to data mining. USA: John Wiley and Sons
- [6] Lior Rokach, & Oded Maimon. (2005). Data mining with decision tree. World Scientific Publishing Co. Pte. Ltd. Series in Machine Perception Artificial Intelligence Volume 69
- [7] Umaedi. (2001). Manajemen peningkatan mutu berbasis sekolah. Jakarta: Departemen Pendidikan Nasional Direktorat Jendral Pendidikan Dasar dan Menengah Direktorat Sekolah Lanjutan Tingkat Pertama
- [8] Tulus. (2004). Peran disiplin pada perilaku dan prestasi siswa. Jakarta: Grasindo
- [9] Xin Yan, & Xiao Gang Su. (2009). Linear regression analysis. London: World Scientific Publishing Co. Pte. Ltd., Covent Garden