

## DATA MINING UNTUK MEMPREDIKSI PRESTASI SISWA BERDASARKAN SOSIAL EKONOMI, MOTIVASI, KEDISIPLINAN DAN PRESTASI MASA LALU

Heri Susanto  
SMK Negeri 4 Surakarta  
heri.smk4@gmail.com

Sudiyatno  
Universitas Negeri Yogyakarta  
sudiyatno@uny.ac.id

### Abstrak

*Penelitian ini bertujuan untuk membuat prediksi prestasi belajar siswa berdasarkan status sosial ekonomi orang tua, motivasi, kedisiplinan siswa dan prestasi masa lalu menggunakan metode data mining dengan algoritma J48. Sebagai perbandingan, data penelitian dianalisis juga dengan CHAID (Chi Squared Automatic Interaction Detection) dan regresi ganda. Pendekatan penelitian yang digunakan adalah kuantitatif. Subyek penelitian ini adalah siswa tingkat X SMK Negeri 4 Surakarta berjumlah 416 siswa. Teknik pengumpulan data yang digunakan adalah dokumentasi dan angket. Hasil penelitian menunjukkan bahwa analisis prediksi menggunakan decision tree algoritma J48 memiliki akurasi sebesar 95,7%, sedangkan analisis prediksi menggunakan CHAID memiliki tingkat akurasi 82,1% dan analisis regresi ganda menghasilkan tingkat signifikansi sebesar 90,6%. Berdasarkan hasil tersebut bisa disimpulkan bahwa metode J48 lebih baik dibandingkan dengan metode CHAID dan regresi ganda.*

**Kata kunci:** data mining

## DATA MINING TO PREDICT STUDENT'S ACHIEVEMENT BASED ON SOCIO-ECONOMIC, MOTIVATION, DISCIPLINE AND ACHIEVEMENT OF THE PAST

### Abstract

*This study aims to make student achievement prediction based on socio-economic status of parents, motivation, discipline students and past achievements using data mining methods with the J48 algorithm. For comparison, the data were analyzed also with CHAID (Chi Squared Automatic Interaction Detection) and multiple regression. The research approach is quantitative. The subjects of this study were student-first level at SMK Negeri 4 Surakarta totaled 416 students. Data collection techniques used are documentation and questionnaires. The results showed that the predictive analysis using J48 decision tree algorithm has an accuracy of 95.7%, while the predictive analysis using CHAID has the rank of an accuracy of 82.1% and a multiple regression analysis resulted in a significance level of 90.6%. Based on these results it can be concluded that the J48 method is better than the CHAID and multiple regression methods.*

**Keywords:** data mining

## PENDAHULUAN

Pendidikan merupakan 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 (Depdiknas, 2003, p.4). Berdasarkan tujuan pendidikan tersebut maka kualitas dan manajemen pembelajaran di sekolah atau lembaga pendidikan perlu ditingkatkan. Salah satu indikator kualitas dan manajemen sekolah atau lembaga pendidikan dapat dilihat dari prestasi belajar siswa.

Sekolah Menengah Kejuruan (SMK) merupakan salah satu lembaga pendidikan yang berupaya menciptakan sumber daya manusia yang memiliki kemampuan, keterampilan, dan keahlian sehingga lulusannya dapat mengembangkan keterampilannya agar dapat terjun dalam dunia kerja maupun untuk melanjutkan ke jenjang pendidikan yang lebih tinggi. Menurut Peraturan Pemerintah No. 29 Tahun 1990 pasal 2 ayat (1) menyebutkan bahwa sekolah kejuruan bertujuan untuk meningkatkan pengetahuan siswa dalam pengembangan diri dan untuk meningkatkan kemampuan siswa sebagai anggota masyarakat. Menurut Peraturan Pemerintah No. 29 Tahun 1990 pasal 3 ayat (2) disebutkan bahwa sekolah kejuruan bertujuan untuk menyiapkan siswa dalam memenuhi lapangan kerja, menyiapkan siswa agar mampu memiliki karir, dan menyiapkan tamatan agar menjadi warga Negara yang produktif, adaptif, dan normatif.

Secara garis besar tujuan diselenggarakan sekolah kejuruan adalah untuk membekali lulusan dengan kompetensi yang berguna bagi diri sendiri dalam karir dan kehidupan bermasyarakat. Tujuan sekolah menengah kejuruan akan lebih terarah jika kurikulum yang digunakan tepat dan dilaksanakan dengan baik.

Usaha untuk meningkatkan kualitas lulusan SMK, dibutuhkan kerjasama dan sinkronisasi yang baik antara pemerintah, manajemen sekolah, kemitraan dengan dunia industri, kompetensi guru, orang tua dan siswa sendiri. Beberapa elemen tersebut sangat berpengaruh baik secara langsung maupun tidak langsung terhadap peningkatan mutu siswa, yang pada akhirnya akan berimbas pada peningkatan kualitas pendidikan. Prestasi belajar siswa dipengaruhi oleh banyak faktor. Salah satu faktor penting dalam pendidikan agar prestasi belajar siswa meningkat adalah manajemen pembelajaran di sekolah. Semakin baik manajemen pembelajaran yang ditawarkan sekolah kepada siswa semakin besar pula kemungkinan prestasi belajar siswa akan baik.

Satu permasalahan yang sering muncul di kalangan peserta didik SMK adalah masih banyak yang mementingkan kompetensi mata pelajaran produktif daripada mata pelajaran adaptif dan normatif. Banyak peserta didik SMK yang piawai menghasilkan karya sesuai dengan kompetensi kejuruannya tetapi mereka lemah dalam mata pelajaran adaptif dan normatif. Contohnya adalah banyak peserta didik jurusan Tata Busana mampu membuat dan mendesain busana yang bagus tapi dia lemah pada mata pelajaran Matematika yang menjadi salah satu syarat kelulusannya. Oleh karena itu, sangat diperlukan perhatian oleh satuan pendidikan untuk mampu menyelenggarakan pembelajaran yang bisa mengatasi masalah tersebut.

Menyadari pentingnya kualitas pembelajaran dalam rangka mencerdaskan kehidupan bangsa, maka pemerintah bersama kalangan 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 ajar, serta pelatihan bagi guru dan tenaga kependidikan lainnya. Tetapi pada kenyataannya upaya pemerintah tersebut belum cukup berarti dalam meningkatkan kualitas pendidikan. (Umaedi, 2001, p.3). Proses belajar mengajar merupakan salah satu faktor penentu keberhasilan pendidikan di

sekolah. Rendahnya mutu pendidikan merupakan akibat dari rendahnya kualitas proses pembelajaran yang dilakukan di sekolah.

Satu hal yang belum banyak dilakukan oleh lembaga-lembaga pendidikan atau sekolah kejuruan adalah melakukan antisipasi terhadap peserta didik yang berpotensi mengalami hambatan atau kurang berprestasi dalam belajarnya. Hal ini dianggap penting karena semakin awal lembaga-lembaga pendidikan atau sekolah kejuruan mengetahui adanya potensi siswa yang kemungkinan akan mengalami hambatan dalam belajarnya, maka lembaga-lembaga pendidikan atau sekolah bisa melakukan langkah-langkah antisipatif.

Dampak dari kurangnya antisipasi atau pencegahan sejak dini terhadap siswa atau calon siswa yang berpotensi mengalami hambatan dalam belajarnya atau kurang berprestasi adalah fakta masih banyak siswa yang tidak naik kelas karena terhambat prestasinya bahkan ada yang dikeluarkan dari sekolah. Kedua hal tersebut sering kita dengar dan pasti dialami oleh semua sekolah.

Berdasarkan uraian di atas, maka penelitian ini bertujuan untuk membuat prediksi sejak dini terhadap siswa yang berpotensi tidak berprestasi atau mengalami hambatan dalam belajarnya, sehingga dapat dilakukan langkah-langkah antisipatif dari sekolah untuk menghindarkan diri dari kemungkinan tidak naik kelas bahkan dikeluarkannya siswa 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.

Beberapa faktor yang mempengaruhi prestasi belajar siswa Sekolah Menengah Kejuruan antara lain sosial ekonomi orang tua, fasilitas belajar sekolah, motivasi, kedisiplinan siswa dan prestasi masa lalu siswa. Hal tersebut dibuktikan dengan banyaknya penelitian yang telah dilakukan sebelumnya.

Faktor status sosial ekonomi orang tua yang mewujudkan pada kemampuan finansialnya. Kemampuan finansial yang berbeda-beda sedikit banyak akan berpengaruh terhadap

prestasi belajar siswa. Dengan kemampuan finansial orang tua, tentunya akan mempengaruhi fasilitas belajar yang disediakan oleh orang tua terhadap sarana dan prasarana yang dibutuhkan oleh seorang siswa untuk meningkatkan prestasi belajarnya.

Berdasarkan Undang-Undang Sistem Pendidikan Nasional No. 20 Tahun 2003 Bab III Pasal 45 tentang sarana dan prasarana pendidikan, dinyatakan bahwa (1) setiap satuan pendidikan formal maupun non formal menyediakan sarana dan prasarana yang memenuhi keperluan pendidikan sesuai dengan pertumbuhan dan perkembangan secara fisik, kecerdasan intelektual sosial, emosional, dan kejiwaan peserta didik; (2) ketentuan mengenai penyediaan sarana dan prasarana pendidikan pada semua satuan pendidikan sebagaimana dimaksud pada ayat (1) diatur lebih lanjut dengan peraturan pemerintah.

Dari kedua ayat di atas dimaksudkan agar tiap-tiap sekolah menyediakan sarana dan prasarana pembelajaran yang memadai semua keperluan pendidikan agar siswa dapat memantulkannya sebagai penunjang belajar siswa.

Tulus (2004, pp.81-83) mengungkapkan bahwa sarana belajar biasanya menjadi penunjang prestasi belajar, namun demikian bila kelengkapan fasilitas belajar sebagai sarana penunjang belajar di sekolah memadai, sebaliknya dapat menjadi faktor penghambat apabila kelengkapan fasilitas belajar di sekolah kurang memadai.

Sekolah Menengah Kejuruan hendaknya menyediakan sarana dan prasarana sesuai dengan kebutuhan dan perkembangan siswa agar dapat menumbuhkan, mengembangkan dirinya sesuai dengan bakat dan kemampuan sebagai manusia seutuhnya. Dengan adanya fasilitas belajar yang lengkap, akan menumbuhkan rasa bangga dan rasa memiliki. Pemeliharaan fasilitas belajar di sekolah merupakan tanggung jawab semua pihak yang bersangkutan. Hal ini bertujuan agar fasilitas belajar dapat dimanfaatkan sesuai fungsinya dan dapat bertahan dengan jangka waktu yang lama. Pengadaan fasilitas belajar sangat penting bagi siswa dan kurikulum pada saat itu.

Belajar sudah menjadi kewajiban yang harus dilakukan oleh seorang siswa. Kadang,

siswa jenuh dan akhirnya memilih untuk tidak belajar. Untuk itulah mereka perlu diberi dorongan baik dari segala pihak. Pengaruh motivasi terhadap prestasi belajar siswa sangat besar karena itu orang – orang di sekitar hendaknya memberi motivasi yang tepat.

Faktor kedisiplinan dalam pengelolaan pengajaran merupakan suatu hal yang sangat penting. Tanpa adanya kesadaran akan keharusan melaksanakan aturan yang sudah ditentukan sebelumnya pengajaran tidak mungkin mencapai target yang maksimal. Seorang siswa perlu memiliki sikap disiplin dengan melakukan latihan yang memperkuat dirinya sendiri untuk selalu terbiasa patuh dan mempertinggi daya kendali diri. Sikap disiplin yang timbul dari kesadarannya sendiri akan dapat lebih memacu dan tahan lama dibandingkan dengan sikap disiplin yang timbul karena adanya pengawasan dari orang lain.

Disiplin dapat tumbuh dan dibina melalui latihan, pendidikan atau penanaman kebiasaan yang harus dimulai sejak dalam lingkungan keluarga, mulai pada masa kanak-kanak dan terus tumbuh berkembang sehingga menjadi disiplin yang semakin kuat. Seperti halnya disebutkan oleh Tulus (2004, p.37) bahwa dengan disiplin yang muncul karena kesadaran diri, siswa berhasil 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 siswa untuk sukses dalam belajar dan kelak ketika bekerja karena kesadaran pentingnya norma, aturan, kepatuhan dan ketaatan merupakan kesuksesan seseorang.

Berdasarkan uraian mengenai faktor-faktor yang mempengaruhi prestasi belajar siswa di atas, maka penelitian ini dilakukan dengan tujuan untuk mengetahui seberapa besar kontribusi dari berbagai faktor tersebut terhadap prestasi belajar siswa dan faktor-faktor mana yang paling dominan kontribusinya terhadap prestasi belajar siswa. Harapan dari hasil penelitian ini adalah adanya hasil rumusan faktor-faktor yang mempengaruhi prestasi belajar siswa. Berdasarkan faktor-faktor tersebut, siswa yang diprediksi memiliki faktor-faktor

kuat mengalami hambatan dalam prestasi belajarnya bisa dilakukan langkah-langkah antisipasi sejak dini terhadap siswa tersebut.

Penelitian-penelitian tentang faktor-faktor yang mempengaruhi prestasi belajar siswa sebelumnya, banyak yang menggunakan pengolahan data statisika. Pada penelitian ini, peneliti menggunakan data mining. Data mining merupakan cabang ilmu baru di bidang komputer, cukup banyak penerapan yang dapat dilakukannya. Hal itu ditunjang kekayaan dan keanekaragaman berbagai bidang ilmu (*artificial intelligence, database, statistik, pemodelan 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. Data mining adalah kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran/berjumlah besar, informasi inilah yang nantinya sangat berguna untuk pengembangan.

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).

Banyak perusahaan, lembaga atau instansi yang menggunakan hasil penelitian menggunakan data mining untuk mendukung tujuannya. Beberapa penelitian yang telah menggunakan data mining dalam berbagai bidang antara lain : (1) penelitian tentang peramalan cuaca menggunakan data mining yang bisa digunakan oleh BMG (Badan Meteorologi dan Geofisika). Judul penelitian “Peramalan Cuaca Menggunakan Data Mining”, Fandianto & Helmy (2009), berdasarkan hasil penelitian, dapat disimpulkan bahwa SVM dengan fungsi kernel polynomial adalah metode data mining terbaik untuk peramalan cuaca interval satu jam dengan tingkat kesalahan yang rendah dan akurat; (2) Judul penelitian “Menerapkan Pendekatan Data Mining Penelitian Tanah di Indonesia” oleh Balai Besar Penelitian dan

Pengembangan Sumber Daya Lahan Pertanian, Yiyi Sulaeman dan Rizatus Shofiyati (2008), berdasarkan penelitian ini, data mining membantu peneliti tanah untuk mengembangkan model (seperti persamaan, aturan, dan *decision tree*) dan untuk menggali dan menemukan pola data. Selain itu, model dan pola data dapat digunakan untuk memformulasikan hipotesis yang mudah dimengerti untuk penelitian tanah dan untuk membangun perangkat (*expert systems* dan *decision support system*) untuk pengelolaan lahan dan tanah. *Framework* untuk mengaplikasikan data mining pada penelitian tanah nasional diusulkan meliputi pengembangan database tanah, seleksi dataset, aplikasi algoritma, dan penyajian dan interpretasi hasil; (3) Judul penelitian “Prediksi Status Keaktifan Studi Mahasiswa dengan algoritma C5.0 dan *K-Nearest Neighbor*”, Iin Ernawati (2008), berdasarkan penelitian ini disimpulkan bahwa Rata-rata keberhasilan algoritma C5.0 dan KNN dalam melakukan klasifikasi data mencapai akurasi di atas 90%. Hal ini menunjukkan bahwa keduanya memiliki performa yang handal dalam melakukan klasifikasi. Algoritma C5.0 tetap dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data yang diklasifikasi dapat diperoleh dengan jelas baik dalam bentuk struktur pohon keputusan maupun aturan *if-then*, sehingga memudahkan pengguna dalam melakukan penggalan informasi terhadap data yang bersangkutan. Berdasarkan ketiga contoh penelitian menggunakan data mining di atas, dapat disimpulkan bahwa penggunaan metode data mining sangat layak dan terpercaya untuk digunakan dalam penelitian khususnya dalam membuat suatu prediksi.

Kehadiran data mining dilatar belakangi dengan problema *data explosion* yang dialami akhir-akhir ini dimana banyak organisasi telah mengumpulkan data sekian tahun lamanya. Kaitannya dengan penelitian ini adalah adanya penumpukan data seperti data sosial ekonomi orang tua, kehadiran siswa (disiplin), data nilai siswa, prestasi masa lalu siswa dan lain-lain. Data-data tersebut biasa didapat oleh sekolah pada saat PPDB (Penerimaan Peserta Didik Baru). Betapa besarnya ukuran data tersebut jika nanti telah berjalan beberapa tahun. Pertanyaannya sekarang, apakah data tersebut

akan dibiarkan menggunung, tidak berguna lalu dibuang, ataukah kita dapat me-‘nambang’-nya untuk mencari ‘emas’, ‘berlian’ yaitu informasi yang berguna untuk meningkatkan prestasi anak didik kita. Banyak di antara kita yang kebanjiran data tapi miskin informasi.

Lembaga pendidikan pada umumnya dan Sekolah Menengah Kejuruan 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 tahunnya antara lain daftar hadir siswa, data nilai calon siswa pada saat penerimaan siswa baru dan data status sosial ekonomi orang tua yang biasa diinput pada saat penerimaan siswa baru juga atau bisa juga dari data yang dimiliki BP/BK di sekolah.

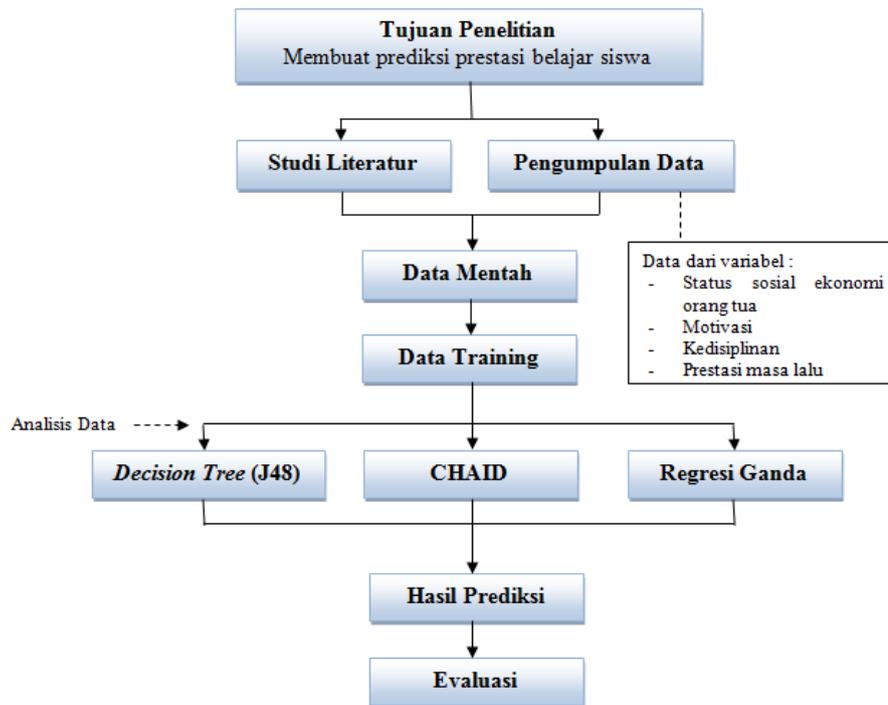
Data - data tersebut belum termanfaatkan dengan baik sebagai bahan pertimbangan, kajian dan penelitian untuk membuat suatu informasi yang berguna bagi sekolah. Oleh karena itu, dengan adanya data mining bisa menjadi salah satu solusi untuk menambang tumpukan data-data tersebut untuk membuat nilai atau 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.

## METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif. Tempat Penelitian ini adalah SMK Negeri 4 Surakarta yang beralamat di Jalan LU Adi Sucipto No. 40 Surakarta. Waktu penelitian dan pengambilan data pada bulan Juni 2013 sampai Desember 2013.

Target/subjek penelitian adalah siswa tingkat X SMK Negeri 4 Surakarta Tahun Pelajaran 2013/2014 sejumlah 416 siswa.

Prosedur penelitian menggunakan tahapan-tahapan KDD (*Knowledge Data Discovery*). Tahapan-tahapannya adalah: (1) *Data Selection*, yaitu tahapan ini dilakukan untuk memilih data yang sesuai dengan variabel yang dibutuhkan dalam penelitian; (2) *Pre-processing/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 duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi); (3) *Transformation, coding* adalah proses transformasi 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.

Penelitian ini memperoleh data dari angket dan dokumentasi. Teknik dokumentasi digunakan untuk mengambil data status sosial ekonomi orang tua, kedisiplinan, dan prestasi masa lalu, sedangkan data motivasi siswa diperoleh melalui angket.

### Teknik Analisis Data

Penelitian ini menggunakan teknik *Decision Tree*, CHAID dan regresi ganda untuk melakukan prediksi prestasi belajar siswa SMK Negeri 4 Surakarta berdasarkan status

ekonomi orang tua, motivasi, kedisiplinan dan prestasi masa lalu.

### 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 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 popu-

lasi sejumlah 416 siswa akan digunakan untuk membuat model prediksi *Decision Tree*. Model yang telah dibuat kemudian akan dihitung tingkat akurasi prediksinya.

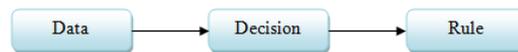
Ada banyak metodologi *Data Mining*, salah satu yang populer adalah Pohon Keputusan (*Decision Tree*). Pohon keputusan merupakan salah satu metode klasifikasi yang sangat menarik yang melibatkan konstruksi pohon keputusan yang terdiri dari node keputusan yang di hubungkan dengan cabang-cabang dari simpul 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 (Larose, 2005, p.107).

Menurut Maimon (2005, p.8), 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*.

Klasifikasi data mining adalah penempatan objek-objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Klasifikasi banyak digunakan untuk memprediksi 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 mengklasifikasikan 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*).



Gambar 2. Alur *decision tree*

### CHAID

Tujuan dari metode ini adalah untuk memisahkan data secara berurutan dengan pembagian biner menjadi beberapa subgrup. Pada tiap tahap, pembagian sebuah grup menjadi dua bagian didefinisikan oleh salah satu variabel prediktor, sebuah himpunan bagian dari kategori-kategorinya mendefinisikan salah satu bagian, dan sisa kategori lainnya mendefinisikan bagian yang lain. Pada AID, prediktornya memiliki dua tipe utama, yaitu monotonik dan bebas. (Kass, 1980, p.241).

Alur proses analisis data dengan CHAID adalah memeriksa tiap variabel independen menggunakan uji *chi-square*, menentukan variabel independen mana yang paling signifikan, membagi data menggunakan kategori variabel independen tersebut dengan peringkat yang paling signifikan, mengulangi langkah ke-4 untuk semua subgrup sampai teridentifikasi semua pembagian yang secara statistik telah signifikan.

### Regresi

Regresi linier adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat (dependen) dengan satu atau lebih variabel bebas (independen). Apabila banyaknya variabel bebas hanya ada satu, disebut sebagai regresi linier sederhana, sedangkan apabila terdapat lebih dari 1 variabel bebas, disebut sebagai regresi linier berganda. Hal ini sesuai dengan definisi tentang regresi linier menurut Xin Yan dan Xiao Gang Su (2009:2) sebagai berikut:

Tabel 1. Detailed accuracy by class

Total Number of Instances  
Detailed Accuracy By Class

TP Rate Class	FP Rate	Precision	Recall	F-Measure	ROC	Area
0.933	0.004	0.99	0.933	0.961	0.984	C
0.979	0.012	0.969	0.979	0.974	0.991	B
1	0.043	0.917	1	0.957	0.984	D
0.864	0.003	0.95	0.864	0.905	0.981	A
0.727	0	1	0.727	0.842	0.997	E
W.Avg.0.957	0.018	0.959	0.957	0.956	0.986	

Regression analysis is the method to discover the relationship between one or more response variables (also called dependent variables, explained variables, predicted variables, or regressands, usually denoted by  $y$ ) and the predictors (also called independent variables, explanatory variables, control variables, or regressors, usually denoted by  $x_1; x_2; \dots; x_p$ ).

Menurut Deny Kurniawan (2008, p.1) analisis regresi setidaknya-tidaknya memiliki 3 kegunaan, yaitu untuk tujuan deskripsi dari fenomena data atau kasus yang sedang diteliti, untuk tujuan kontrol, serta untuk tujuan prediksi. Regresi mampu mendeskripsikan fenomena data melalui terbentuknya suatu model hubungan yang bersifatnya numerik. Regresi juga dapat digunakan untuk melakukan pengendalian (kontrol) terhadap suatu kasus atau hal-hal yang sedang diamati melalui penggunaan model regresi yang diperoleh. Selain itu, model regresi juga dapat dimanfaatkan untuk melakukan prediksi untuk variabel terikat. Namun yang perlu diingat, prediksi di dalam konsep regresi hanya boleh dilakukan di dalam rentang data dari variabel-variabel bebas yang digunakan untuk membentuk model regresi tersebut.

Alur proses dalam regresi ganda adalah menghitung mean, menghitung standard deviasi, menghitung kuadrat varian, menghitung korelasi product Moment, menghitung nilai beta, menghitung nilai  $b_1, b_2$ , dan  $a$ , membuat persamaan garis regresi, menghitung sumbangan relatif dari kedua prediktor ( $R^2$ ), menghitung nilai signifikansi ( $F$ ) dan menghitung signifikansi tiap-tiap prediktor ( $t$ ).

## HASIL PENELITIAN 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. Salah satunya adalah menentukan atribut-atribut untuk variabel Sosial Ekonomi Orang Tua yang diambil dari data operasional yaitu Data Pribadi Siswa yang ada di BP/BK sekolah; (2) *Pre-processing/Cleaning*, proses *cleaning* tersebut dilakukan terhadap keseluruhan data yang diteliti yang berjumlah 416 siswa. Setelah dilakukan proses *cleaning* data sejumlah 416, dihasilkan data bersih sebanyak 346 record data yang digunakan untuk proses analisis berikutnya; (3) *Transformation*, tahap ini menghasilkan satu *recordset* data yang siap untuk analisis data; (4) Analisis data.

Analisis data yang pertama adalah dengan menggunakan teknik *decision tree* algoritma J48. Software yang digunakan adalah WEKA 3.6.9. Hasil yang diperoleh adalah tingkat akurasi prediksi sebesar 95,7%, seperti tampak pada tabel 1.

Analisis yang kedua menggunakan CHAID (*Chi Squared Automatic Interaction Detection*). Analisis ini menggunakan software SPSS 16.0. Hasil yang diperoleh adalah tingkat akurasi prediksi sebesar 82,1%, seperti tampak pada tabel 2.

Tabel 2. CHAID Classification

Observed	Predicted					
	A	B	C	D	E	F
A	0	22	0	0	0	0%
B	0	83	14	0	0	85.6%
C	0	0	90	15	0	85.7%
D	0	0	0	111	0	100%
E	0	0	0	11	0	0%
Overall Percentage	0%	30.3%	30.1%	39.6%	0%	82.1%

Growing Method: CHAID

Dependent Variable: nilai\_uts

Analisis yang ketiga menggunakan regresi ganda. Analisis ini menggunakan software SPSS 16.0. Hasil yang diperoleh adalah tingkat signifikansi seluruh variabel independen terhadap variabel dependen sebesar 90,6%.

#### Perbandingan Hasil Akurasi Prediksi antara Metode J48, CHAID dan Regresi Ganda

Perbandingan akurasi klasifikasi antara metode J48, CHAID dan regresi ganda untuk *class* nilai UTS sebagai target adalah sebagai berikut:

Tingkat akurasi atau signifikansi variabel prediktor terhadap prestasi belajar siswa dengan menggunakan metode J48 sebesar 95,7%. Tingkat akurasi atau signifikansi variabel prediktor terhadap prestasi belajar siswa dengan menggunakan metode CHAID sebesar 82,1%. Tingkat akurasi atau signifikansi variabel prediktor terhadap prestasi belajar siswa dengan menggunakan regresi ganda sebesar 90,1%. Berdasarkan hasil tersebut terlihat bahwa penggunaan metode J48 lebih baik dari metode CHAID dan regresi ganda. Hal terlihat seperti pada tabel di bawah ini.

Tabel 3. Perbandingan Hasil Prediksi J48, CHAID dan Regresi Ganda

Observed	J48	CHAID	Regresi
Rata-rata	95.7%	82.1%	90,1

## SIMPULAN DAN SARAN

### Simpulan

Berdasarkan analisis data menggunakan *decision tree* data mining untuk memprediksi

prestasi belajar siswa berdasarkan status sosial ekonomi orang tua, motivasi, kedisiplinan siswa dan prestasi masa lalu 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 dibuktikan dengan adanya variabel motivasi yang menjadi *root node* dalam *decision tree* yang terbentuk. Variabel prestasi masa lalu merupakan variabel kedua yang penting dalam keberhasilan siswa menempuh studinya. 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. Rata-rata keberhasilan algoritma J48 dalam melakukan klasifikasi data mencapai akurasi di 95,7%. Hal ini menunjukkan bahwa algoritma ini memiliki performa yang handal dalam melakukan klasifikasi; (2) Hasil akurasi klasifikasi menggunakan metode CHAID dengan data yang sama memperoleh hasil rata-rata prediksi 82,1. Hal ini menunjukkan bahwa berdasarkan hasil klasifikasi menggunakan metode CHAID tersebut memiliki tingkat akurasi yang tergolong baik, yaitu lebih dari 80%; (3) Berdasarkan tabel di atas angka R Square adalah 0,906 untuk predictor Motivasi dan Prestasi\_ml. Hasil tersebut adalah hasil kuadrat korelasi. *Standar Error of the Estimate* adalah 0,30477, perhatikan pada analisis deskriptif statistik bahwa standar deviasi nilai uts adalah 0,99246 yang jauh lebih besar dari dari standar *error*, oleh karena lebih besar daripada standar deviasi nilai uts maka

model regresi ini bagus dalam bertindak sebagai predictor nilai uts.

Berdasarkan ketiga hasil analisis di atas dapat diketahui bahwa akurasi prediksi prestasi belajar menggunakan metode *J48* adalah sebesar 95,7%, sedangkan metode CHAID mempunyai akurasi prediksi sebesar 82,1% dan analisis regresi ganda mempunyai tingkat signifikansi 90,6%. Berdasarkan data tersebut dapat disimpulkan bahwa penggunaan metode data mining dengan algoritma *decision tree* (*J48*) memiliki tingkat akurasi yang lebih baik dari 2 (dua) metode lain yaitu CHAID dan regresi ganda.

### Saran

Berdasarkan simpulan di atas, penulis menyarankan beberapa hal, yaitu: (1) Keterbatasan data yang cukup beserta informasi pendukungnya yang dimiliki oleh sekolah. Ketersediaan data yang dimiliki sangat mempengaruhi tingkat akurasi data dalam melakukan prediksi. Salah satu diantaranya adalah data tentang kedisiplinan yang tidak hanya menggunakan rekap absensi siswa; (2) Perlunya untuk menyimpan data yang disimpan untuk jangka waktu yang lama, sehingga menjamin ketersediaan data tersebut jika dibutuhkan sewaktu-waktu. 3) Diperlukan kebijakan yang jelas dan tepat sasaran untuk mengatasi siswa yang berpotensi kesulitan dalam hal prestasi belajarnya. Misalnya dengan melakukan pendampingan siswa tersebut.

### DAFTAR PUSTAKA

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