

MENERAPKAN PENDEKATAN DATA MINING PENELITIAN TANAH DI INDONESIA

Applying Data Mining Approach On Soil Research In Indonesia

Yiyi Sulaeman dan Rizatus Shofiyati

Balai Besar Penelitian dan Pengembangan Sumbidaya Lahan Pertanian

ABSTRACT

Results of soil data from soil research activity may be useful for other purposes if they are stored in digitalized database and analyzed using the appropriate approach and tool. The data mining approach provides a superior and speedy means in problem solving, presentation of the results, data assumption and research goals. Such an approach has not been fully used in national soil research.

This paper aims to discuss the potential contribution of data mining and the challenges to apply it in national soil research, as well as to propose a framework in its application. Data mining provides techniques for discovering data pattern(s) and building models using large soil dataset. Data pattern(s) and models are used to formulate hypothesis and develop tools for soil and land management. The proposed framework being tailored to national condition covers soil digital database development, dataset selection, algorithm application, and presentation and interpretation of the results. Examples of decision tree and rule model are provided. Yet, to fully take advantage of these opportunities, national soil scientists are challenged to find ways on how: (i) to develop good soil digital database; (ii) to understand data mining algorithms, and (iii) to increase their capability in operating available software. Hopefully, studies benefiting from data mining approach may increase in near future.

Keyword : data mining, soil database, data mining framework.

PENDAHULUAN

Setiap tahun, pengumpulan data tanah dilakukan para peneliti tanah di seluruh wilayah Indonesia, sehingga dari tahun ke tahun jumlah data tanah semakin bertambah. Data tersebut diperoleh baik secara langsung dari lapangan maupun dari data sekunder yang ada. Data yang telah terkumpul dan berjumlah sangat banyak ini merupakan aset yang sangat berharga dan merupakan informasi yang sangat berguna jika dikelola dengan benar. Syaratnya, pertama, data harus tersimpan dalam bentuk digital sehingga data dapat diakses dengan mudah. Kedua, peneliti dapat memahami cara menggunakan perangkat (*tool*) untuk menganalisis, mengolah, dan mengelola *database* yang besar. Tanpa terpenuhinya persyaratan tersebut, peneliti akan kaya data tetapi miskin informasi, karena tidak dapat memanfaatkan data yang dimilikinya menjadi informasi yang berguna (Pramudiono, 2003). Hal ini merupakan tantangan bagi peneliti tanah untuk memikirkan cara mengelola dan menganalisis data yang mereka miliki untuk memecahkan permasalahan yang berhubungan dengan tanah saat ini.

Eswaran dan Kimble (2003) mengemukakan bahwa saat ini pengguna mempunyai cara yang terbatas untuk menemukan cara yang benar dalam memberikan solusi, sebab beberapa keputusan dibuat dengan menggunakan data yang minim, tidak memenuhi syarat, atau data yang rendah kebenarannya. *Database* tanah menawarkan data yang cukup dan sesuai untuk pembuatan keputusan tentang pengelolaan tanah dan lahan. Inisiatif pengembangan *database* tanah dapat berasal dari institut atau peneliti tanah secara perseorangan. Sebagai contoh *database* tanah institusi yang sudah dipublikasikan antara lain Indonesian Land Resources Information System (Balai Penelitian Tanah, 2005), SOTER digital database (FAO, 1995), ISRIC's Soil Information System (Temple *et al.*, 1996), and WISE Database (Batjes, 1995). Sedangkan contoh *database* yang dibuat secara perseorangan adalah Topsoil Dijital Database (Sulaeman and Nursyamsi, 2005).

Selanjutnya untuk mengembangkan *database*, peneliti tanah juga harus meningkatkan pengetahuan dalam memilih teknik untuk menganalisis ratusan bahkan ribuan dataset. Teknik visualisasi data dan statistik yang ada membuat komputer bekerja secara pasif dan membiarkan peneliti menentukan model yang diinginkan, atau permintaan korelasi yang diperlukan. Sementara itu, teknik pengelolaan data dengan menggunakan *database* membuat komputer bekerja secara aktif, dimana peneliti dapat melakukan satu perintah "membangun sebuah metode untuk memprediksi tujuan menggunakan beberapa faktor sebagai *input*".

Data mining (DM) mengacu pada proses penggalian pengetahuan baru (antara lain korelasi baru, pola, model, dan tren) tentang penyimpanan data di tempat penyimpanan, baik pengemasan data dalam jumlah besar maupun penggunaan teknologi *pattern recognition* (StatSoft, 1999; Luan, 2002; Pramudiono, 2003). Hal tersebut dapat menyelesaikan masalah dengan analisis data yang sudah tersedia di dalam *database* (Lavrac and Grobelnik, 2003).

DM telah digunakan di bidang lain seperti survei media, data kecelakaan lalu lintas, data pengenalan produk (Lavrac and Grobelnik, 2003), pendidikan (Luan, 2002), pengkajian obat-obatan dan efek pengobatan (SPSS Inc, 1999a). Akan tetapi DM belum digunakan secara penuh pada ilmu tanah khususnya pada penelitian tanah di Indonesia, dimana data yang berjumlah besar belum dikelola secara baik. Tulisan ini menjelaskan kontribusi yang potensial dari DM untuk penelitian tanah. Framework dan tantangan dalam pengaplikasian DM pada penelitian tanah nasional juga dibahas dalam tulisan ini.

BAGAIMANA “DATA MINING” MEMBANTU PENELITI TANAH

Meskipun statistik memiliki kemampuan sebagai alat untuk menganalisis data, DM diperlukan untuk beberapa alasan yang berkaitan dengan kecepatan pemecahan permasalahan, menyajikan hasil, asumsi data, dan tujuan (Friedman, 1997 seperti yang dikutip oleh Lavrac and Grobelnik, 2003). Banyak metode DM yang dapat menangani dataset besar secara efektif, sementara metode statistik tidak dapat melakukannya. Selain itu, DM menyajikan hasil analisis secara baik, umumnya mudah dimengerti dengan menggunakan bentuk lain yang lebih formal (seperti *decision trees*, dan aturan). Hal ini dicoba untuk memenuhi struktur yang terurut dari data tabular yang terstruktur sebaik data yang tidak terstruktur. Lebih dari itu, DM menciptakan hipotesis yang mudah dimengerti, sementara statistik adalah hipotesis penguji. Yang lebih penting, DM dapat membantu peneliti tanah dalam memformulasikan hipotesis yang mudah dimengerti untuk penelitian mereka.

Klasifikasi, prediksi, penggalian dan penemuan pola sifat tanah merupakan permasalahan dasar di dalam penelitian tanah. Hal-hal tersebut menjadi lebih penting di negara sedang berkembang dimana anggaran menjadi masalah utama untuk penelitian tanah. Sehubungan dengan permasalahan tersebut, DM menawarkan teknik (Tabel 1) untuk mengembangkan model dan menggali pola data dari data yang tersedia. Model dan pola data tersebut tidak hanya dapat digunakan untuk memformulasikan hipotesis yang mudah dimengerti, tetapi juga untuk mengembangkan perangkat untuk pengelolaan lahan dan tanah, seperti *expert system* dan *decision support system* (Gambar 1).

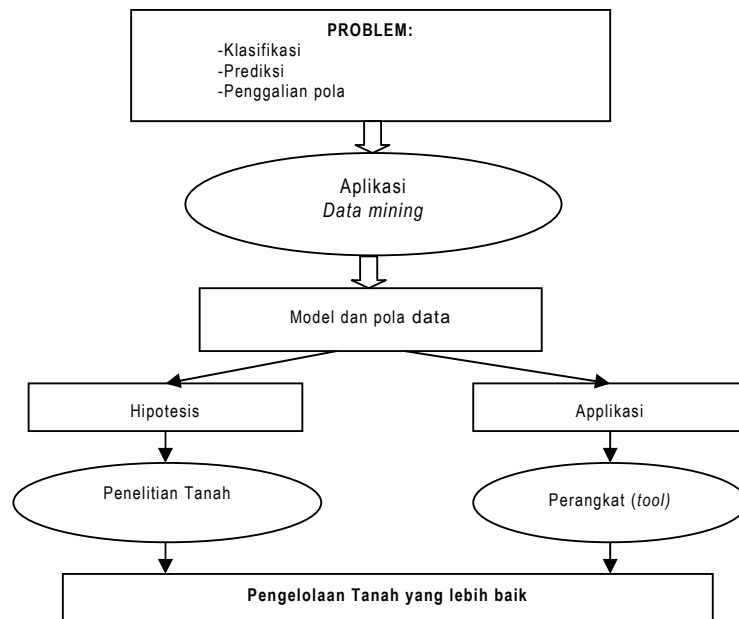
Dengan memanfaatkan DM, peneliti tanah dapat mengembangkan model yang terkait dengan tanah. Proses ini mencakup fase pembelajaran dimana bagian dari data digunakan untuk mengembangkan model perkiraan dan fase pengujian dimana model yang dikembangkan diuji dengan bagian data lainnya untuk mengetahui keakuratan model. Jika keakuratannya cukup tinggi, model dapat digunakan untuk meramalkan kelas dari data yang tidak dikenal. Teknik yang paling populer untuk tujuan ini adalah *regression tree* (SPSS Inc, 1990b, 2001) dan *induction rule* (Lavrac and Grobelnik, 2003).

Tabel 1. Teknik *Data Mining* dan Penggunaannya.

Permasalahan	Tujuan	Teknik yang sesuai
<i>Klasifikasi dan prediksi</i>		
Klasifikasi	Membangun model klasifikasi (kadang disebut pengelompokan) yang dapat menandai label kelas dengan benar untuk objek yang tidak berlabel dan tidak dikenal.	<ul style="list-style-type: none"> • <i>Discriminant analysis</i> • <i>Rule induction methods</i> • <i>Decision tree learning</i> • <i>Neural nets</i> • <i>K Nearest Neighbor</i> • <i>Case-based reasoning</i> • <i>Genetic algorithms</i>
Prediksi	Menemukan nilai numerik dari atribut target untuk objek-objek yang tidak dikenal.	<ul style="list-style-type: none"> • <i>Regression analysis</i> • <i>Regression trees</i> • <i>Neural nets</i> • <i>K Nearest Neighbor</i> • <i>Box-Jenkins methods</i> • <i>Genetic algorithms</i>
<i>Penggalian pola</i>		
Segmentasi	Memisahkan data ke dalam subgrup atau kelas yang menarik dan memiliki arti	<ul style="list-style-type: none"> • <i>Clustering techniques</i> • <i>Neural nets</i> • <i>Visualization</i>
Konsep deskripsi	Memperoleh pendekatan konsep atau kelas	<ul style="list-style-type: none"> • <i>Rule induction methods</i> • <i>Conceptual clustering</i>
Analisis ketergantungan	Menemukan suatu model yang mendiskripsikan ketergantungan yang signifikan (asosiasi) antara item data dan kejadian	<ul style="list-style-type: none"> • <i>Correlation analysis</i> • <i>Regression analysis</i> • <i>Association rules</i> • <i>Bayesian networks</i> • <i>Inductive Logic Programming</i> • <i>Teknik visualisasi</i>

Peneliti tanah juga dapat menggali dan menemukan pola data tanah, seperti cluster, asosiasi, dan korelasi. Teknik DM dapat memisahkan data dari dataset yang besar ke dalam kelompok-kelompok yang lebih kecil. Selanjutnya, teknik tersebut dapat digunakan untuk menentukan tingkat evaluasi diantara dua variabel atau dua grup variabel. Teknik yang paling populer untuk tujuan ini adalah pengelompokan tanpa menggunakan kelas data yang spesifik dan pengenalan asosiasi untuk penggalian aturan asosiasi dari kombinasi beberapa item (Pramudiono, 2003).

Model yang dihasilkan tersebut (seperti aturan, persamaan matematika, dan *decision tree*) dapat digunakan untuk memformulasikan hipotesis untuk penelitian tanah lebih lanjut atau mengembangkan perangkat untuk pengelolaan lahan dan tanah. Aturan "*if...then*" dan persamaan matematika, sebagai contoh, seringkali digunakan dalam pengembangan *expert system*, sementara itu *decision tree* dapat digunakan untuk mengembangkan *decision support system*. Integrasi antara DM dan *decision support* serta bagaimana menggunakan DM untuk membuat keputusan yang lebih baik didiskusikan dalam Lavrac and Bohanic (2003).



Gambar 1. Data Mining pada Ilmu Tanah.

USULAN FRAME WORK

Tidak seperti halnya pada bidang bisnis dan pemasaran, *framework standard* untuk aplikasi DM pada ilmu tanah belum tersedia (Chapman *et al.*, 2000). Oleh karenanya, *framework* integrasi dari aplikasi DM dan *database* tanah perlu dibangun. Usulan *framework* tersebut disajikan pada Gambar 2. *Framework* ini diharapkan akan lebih bermanfaat di Indonesia, dimana *database* tanah dan DMnya belum berkembang dan dimanfaatkan sepenuhnya.

Pengembangan *Database* Dijital Tanah

Database dijital tanah mengacu pada kumpulan data tanah yang terorganisasi dengan baik dimana data mudah untuk disimpan dan digunakan (seperti diperbaharui, dicari, dianalisis, dan dihapus). Digital dalam hal ini diindikasikan bahwa *database* berbasis komputer. Untuk mengembangkan sistem pengelolaan *database* (Data Base Management System/DBMS), data tanah dan program komputer (seperti MS Access, Sybase, SIG) harus tersedia.

Struktur *database* harus dirancang berdasarkan *framework* sistem informasi tanah (Burough, 1991) atau tanah dan topografi (FAO, 1995). Sistem ini merupakan integrasi data tanah spasial dan non spasial. Data tanah spasial adalah peta tanah, peta kesesuaian lahan, dan lain-lain. Sedangkan data non spasial tanah atau data atribut termasuk deskripsi profil tanah, sifat tanah, dan lain-lain. Selama pengembangan *database*, peta dijital dibuat dengan menyertakan plot observasi lapang di dalamnya. Untuk masing-masing observasi, karakteristik site dan pedon dimasukkan, kemudian untuk masing-masing karakteristik horison pedon didokumentasikan.

Penelitian yang hanya memperhatikan data non-spasial bisa mengikuti *framework database* konvensional (Nugroho, 2004). Dalam hal ini, peneliti hanya membuat tabel yang terdiri dari tanah dan sifat site observasi. Sulaeman dan Nuryamsi (2005), Temple *et al.* (1996), Batjes (1995), menggunakan pendekatan ini untuk mengembangkan *database* sifat tanah.

Input data tanah dari berbagai sumber dapat dilakukan dengan berbagai teknik. Pedon dan karakteristik horison dalam buku, jurnal, dan laporan, bisa dimasukkan ke dalam DBMS atau dipindai dan dipindah ke dalam format DBMS. Data tanah dari internet dalam format HTML atau PDF dapat diformat secara langsung ke dalam DBMS. Transfer elektronik umum dilakukan di dalam pengelolaan *database* dan diperlukan waktu yang singkat. Sebagai contoh, format Excel dapat ditransfer secara langsung ke dalam Access. Modul I-mark (FAO, 2003) yang tersedia secara gratis, menawarkan petunjuk bagaimana mengolah dokumen elektronik tersebut.

Dalam prakteknya, pengembangan *database* menggunakan sedikit pengetahuan lanjutan tetapi waktu yang diperlukan cukup banyak. Pemasukan dan pengeditan peta analog menjadi peta digital memerlukan banyak waktu. Eastman (2001) melaporkan bahwa pengembangan *database* seringkali mengkonsumsi 90% waktu dari sebuah proyek atau kegiatan. Oleh karena itu, operator harus mempersiapkan struktur dan disain secara baik sehingga sumberdaya dapat digunakan secara efisien.

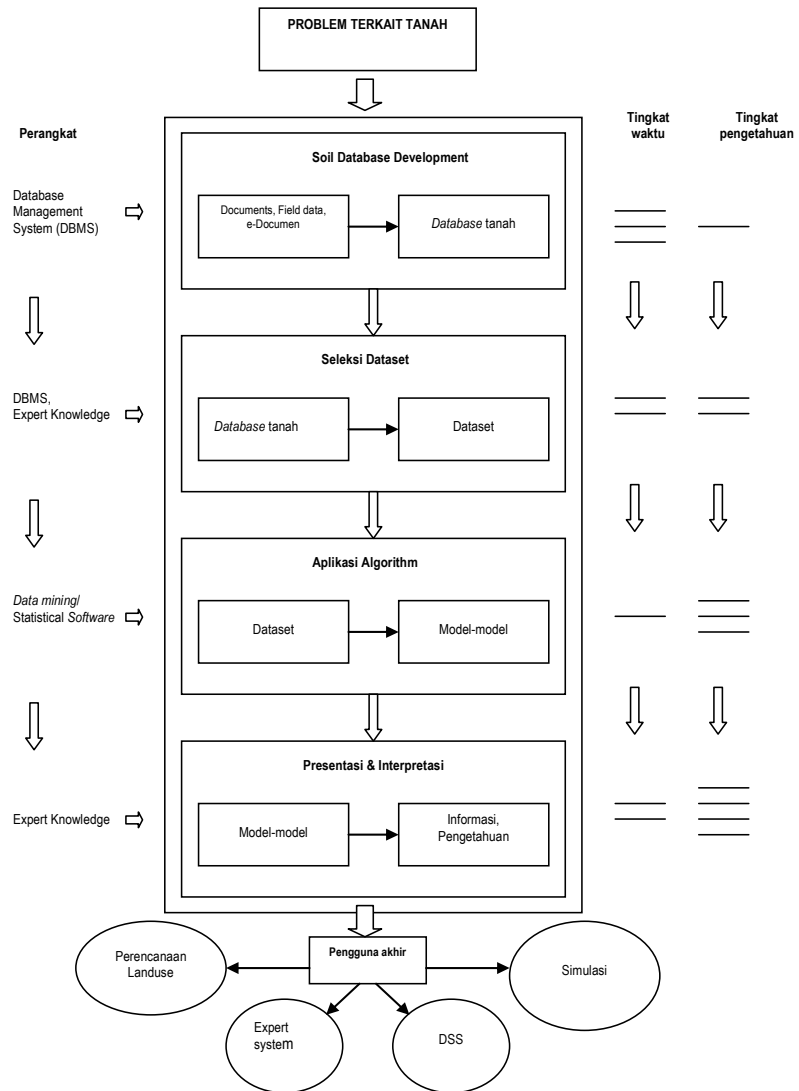
Unit komputer Pusat Penelitian Tanah dan Agroklimat - Puslittanak (sekarang Instalasi Basisdata Balai Besar Penelitian dan Pengembangan Sumberdaya Lahan Pertanian - BBSDLP) telah mengembangkan Sistem Informasi Sumberdaya Lahan (Balai Penelitian Tanah, 2005). Sistem yang disebut *tbase* ini adalah untuk mengembangkan data tanah tabular yang termasuk di dalamnya data morfologi tanah dan fisika kimia tanah. Bagian lain dari sistem tersebut adalah *tMap* yang berfungsi untuk mengelola data spasial. Selanjutnya, para peneliti juga mengembangkan *database* tanah non-spasial khususnya fungsi spesifik sifat-sifat tanah. Selain itu, *Topsoil Fertility Digital Database* yang dikemas dalam *software Phosphat Kalium Dessision Support System (PKDSS)* juga telah dikembangkan (Sulaeman and Nursyamsi, 2005). Pada saat ini, *database* tersebut telah menyimpan sifat-sifat kesuburan tanah sawah di Pulau Jawa.

Seleksi Dataset (Lembar Data)

Pada saat *database* digital diperkenalkan, dataset dapat diekstrak darinya. Macam sifat tanah dan jenis data yang terdapat di dalam dataset tergantung pada permasalahan dan tujuan studi. Sebagai contoh, jika permasalahan untuk memprediksi kandungan air tanah pada areal tertentu dan tujuannya adalah untuk membangun persamaan matematika menggunakan kelas tekstur, maka data kandungan air dan kelas tekstur harus ada pada dataset. Tetapi, jika permasalahannya adalah untuk memprediksi distribusi spasial kandungan air dalam wilayah tertentu dan tujuannya adalah untuk mengembangkan *decision tree*, maka peta tekstur dan peta kandungan air harus tersedia di dalam dataset. Untuk permasalahan tersebut, beberapa tujuan dapat dirancang dan beberapa dataset dapat diperoleh hanya dari satu dataset.

Setelah dataset diperkenalkan, *data cleaning* harus dilakukan untuk menghilangkan data yang tidak konsisten. Dataset bisa berisi nilai yang ekstrim (terlalu tinggi atau rendah) atau tidak logik. Nilai tersebut menyebabkan kesalahan (*error*) dalam penghitungan atau *input* data. Data tersebut lebih baik dikeluarkan dari dataset. Selanjutnya, data bisa dihitung dari prosedur berbeda. Sebagai contoh, karbon organik dapat dihitung menggunakan prosedur Kurmies atau Walkey dan Black. Agar konsisten, satu prosedur harus dipilih. Lebih dari itu, dataset dapat

menggunakan unit yang berbeda. Sebagai contoh, Kapasitas Tukar Kation (KTK) dapat menggunakan satuan me/100g atau cmol/kg. Agar konsisten, satuan harus dipilih. Bagaimanapun, kesalahan dalam interpretasi dapat terjadi jika *data cleaning* diabaikan.



Gambar 2. Framework untuk Aplikasi Data Mining pada Ilmu Tanah.

Aplikasi algoritma

DM dapat diaplikasikan setelah dataset diimpor ke dalam *software*. Selanjutnya, dataset dibagi ke dalam satu set perintah untuk membangun model atau menggali pola data dan set pengujian untuk memvalidasi hasil. Di dalam aplikasi algoritma, peneliti dapat menggunakan satu teknik atau lebih (Tabel 1.) tergantung dari permasalahan dan tujuan. Setelah itu validasi hasil dilakukan untuk menentukan tingkat kepercayaan. Peneliti harus menentukan modifikasi interval kepercayaan terhadap domain permasalahan. Jika data berasal dari bentuk eksperimen, maka 95% interval kepercayaan menjadi terlalu tinggi dan 75% merupakan angka yang bagus (Wilding and Dress, 1983).

Peneliti tanah seringkali harus mempertimbangkan beberapa perangkat lunak, sebab satu perangkat lunak memiliki teknik atau fasilitas yang tidak tersedia pada perangkat lunak lainnya. Sebagai contoh, S-PLUS dan STATISTICA memiliki teknik *regression tree* tetapi tidak tersedia analisis *neural network*, sementara MATHLAB memiliki fasilitas untuk analisis *neural network*.

Penyajian dan Interpretasi

Penyajian terdiri dari visualisasi model dalam bentuk aturan 'if... then', decision tree, persamaan matematik, atau *neural network*. Model aturan terdiri dari aturan if-then dengan format, 'if' menerangkan kondisi yang akan ditentukan sebagai sebuah kelas, dengan sebuah nilai atribut pada kondisi tertentu dan sebuah label pada kelas tersebut. Suatu set aturan umumnya diinterpretasi dengan cara *top-down* sebagai aturan 'if-then-else' (suatu daftar keputusan). Aturan pertama yang mempunyai kondisi yang sesuai dengan nilai atribut yang menerangkan obyek menandai kelasnya untuk suatu obyek (Lavrac and Grobelnik, 2003). Model *decision tree* terdiri dari nomor *node* dan *arc* (garis). Pada umumnya, node diberi label oleh sebuah nama atribut, dan *arc* (garis) oleh nilai tertentu dari atribut tersebut. Atribut bisa berupa sifat tanah atau variabel lingkungan sekitarnya (seperti kemiringan). Node paling atas disebut akar pohon dan node paling dasar disebut daun. Masing-masing daun diberi label oleh kelas (Lavrac and Grobelnik, 2003). *Neural network* terdiri dari satu set unit input, output dan unit tersembunyi, yang menghubungkan input dan output. Unit tersembunyi memindahkan informasi yang berguna dari unit input dan menggunakannya untuk memperkirakan unit output (Minasyni et al., 1999).

Model interpretasi, aplikasi praktis, dan asumsi dijelaskan lebih baik dalam penyajian. Pengguna akhir (seperti model dan pola data) untuk tujuan mereka seperti simulasi, pembuatan *decision support system*

(DSS), pembuatan *expert system*, perencanaan penggunaan lahan atau sejenisnya.

Contoh-contoh

Pada contoh berikut, aplikasi *regression tree* dan teknik instruksi aturan didiskusikan pada sub bab ini. Contoh pertama, *regression tree* telah digunakan untuk mengembangkan model *decision tree* untuk memprediksi kedalaman horison A. Sementara itu sebagai contoh kedua, teknik instruksi aturan digunakan untuk mengembangkan faktor koreksi dalam dinamika hara tanah. Pada kenyataannya, semua hasilnya merupakan sebuah hipotesis yang harus diuji di lapangan.

Satu hasil pemodelan *regression tree* adalah *decision tree*. Gambar 3 memperlihatkan sebuah contoh model *decision tree* untuk memperkirakan kedalaman horison A menggunakan kemiringan dan elevasi di Provinsi Lampung. Root (0,153 m) adalah rata-rata terbaik dari kedalaman horison A dalam dataset. Selanjutnya, *tree* mengindikasikan bahwa kelerengkan merupakan faktor yang paling penting dalam memperkirakan kedalaman Horison A dalam suatu wilayah. *Tree* ini juga menyarankan kelerengkan 2,75% merupakan tingkat yang kritis untuk membedakan daerah datar dari lahan yang berlereng.

Teknik instruksi aturan digunakan untuk menentukan faktor koreksi dalam PKDSS, sebuah model untuk menghitung rekomendasi pemupukan berdasarkan uji tanah (Nursyamsi dan Sulaeman, 2001; Nursyamsi *et al.*, 2003). Semua hasil studi kesuburan tanah dikompilasi dan dianalisis untuk memperoleh aturan-aturan. Aturan yang disajikan berikut adalah untuk menentukan faktor koreksi untuk pemupukan unsur K dan contoh aturan dengan menggunakan *if.. then* :

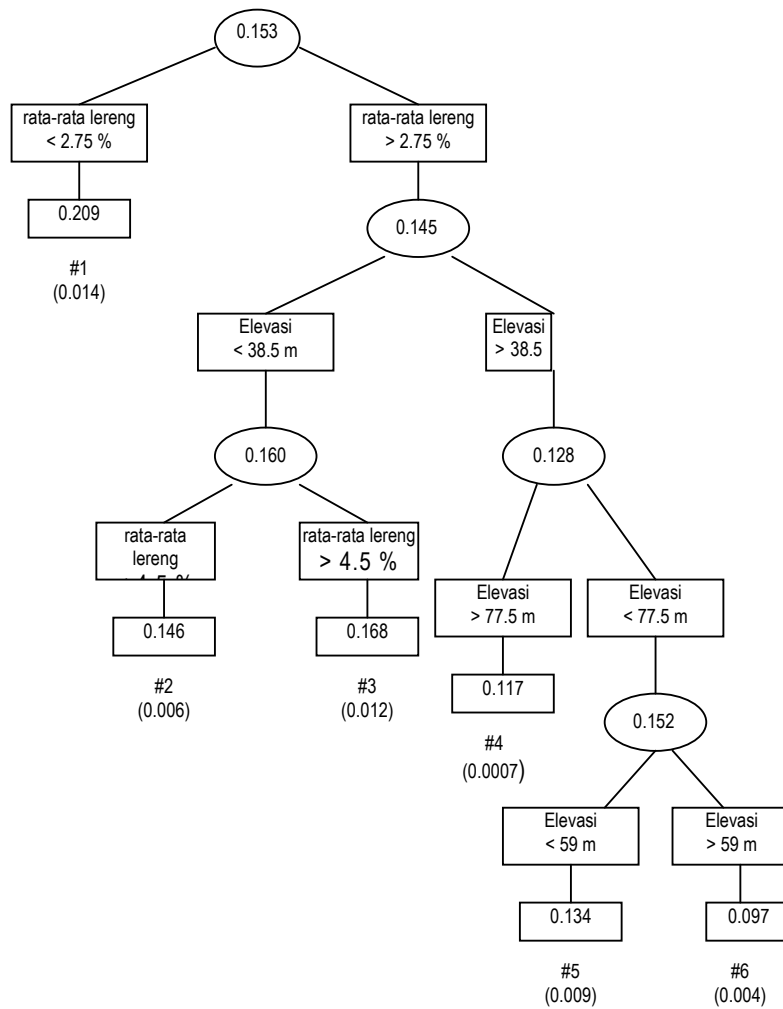
IF *soil texture is sandy* AND *organic carbon is less than 2 %* THEN

IF *clay CEC is less than 10 cmol/kg* THEN *the correction factor value is 1.320*

ELSE IF *clay CEC ranges 10 to 20 cmol/kg* THEN *the correction factor value is 1.900*

ELSE IF *clay CEC ranges 20 to 60 cmol/kg* THEN *the correction factor value is 1.320*

ELSE *the correction factor value is 1.540*



Gambar 3. *Regression tree* untuk memperkirakan kedalaman horizon A (dalam meter) pada tanah tuff di Provinsi Lampung

Tampilan model PKDSS dikembangkan oleh DM dan telah diuji di lapangan (Setyorini *et al.*, 2003) dengan *framework* lain untuk menentukan rekomendasi pemupukan. Teknik tersebut adalah teknik kurva umum, teknik Mitcherlich, Peta P dan K, dan teknik IRR. PKDSS

menunjukkan tampilan terbaik dalam hal produksi dan indeks rasio BC (Tabel 2).

Tabel 2. Perbandingan Tampilan PKDSS dan Teknik Lainnya Untuk Rekomendasi Pemupukan Pada Inceptisols di Pati, Jawa Tengah.

Perlakuan	Produksi (kg gabah/ha)	Indeks rasio BC
Kontrol	1,814	1,00
Kebiasaan petani	2,232	1,40
Rekomendasi kurva umum	2,700	2,68
Kebiasaan petani	2,232	1,40
Rekomendasi Mitcherlich	2,450	1,73
PKDSS	2,802	2,77
Peta P dan K	2,320	1,60
Rekomendasi IRR1	2,250	1,42

Sumber: Widowati et al. (2004)

TANTANGAN PADA PENGAPLIKASIAN DATA MINING

Teknik DM menawarkan algoritma untuk mengembangkan model dan menggali dan menemukan pola data. Dilengkapi dengan penggunaan teknik ini, peneliti tanah dapat 1) mengembangkan hipotesis tentang permasalahan dan menguji mereka, membuat model peramalan, 2) menemukan *cluster-cluster* dalam *database*, 3) menggali dan menemukan asosiasi antara aktivitas, 4) mereview deviasi, 5) membuat keputusan yang lebih baik dengan mengontrol baik data dan kepakaran, dan 6) meningkatkan kepercayaan untuk membuat keputusan yang lebih cerdas berdasarkan sejarah data. Dengan hasil DM, peneliti tanah dapat mengubah apa yang dikerjakan melalui pembelajaran apa yang sedang terjadi, dan akan menawarkan ke pengguna dengan informasi tentang apa yang telah terjadi, sehingga memberdayakan pengambil keputusan untuk mengubah sesuatu yang akan terjadi (SPSS Inc, 1999a).

Kesuksesan DM memerlukan pengetahuan subyek dan beberapa kemampuan analisis. Pramudiono (2003) menekankan bahwa penting untuk memahami tidak hanya tentang teknik DM akan tetapi juga tentang permasalahan apa yang ingin dipecahkan. DM memiliki algoritma untuk menemukan suatu pola. Hanya kepakaran manusia dalam hal pengetahuan arti dan konteks data, dapat menentukan bagaimana data dapat dianalisis secara beralasan dan bagaimana menginterpretasi dan mengevaluasi beberapa hasil. Peneliti tanah

selanjutnya bertanggung jawab untuk menguji secara hati-hati beberapa model atau memperkirakan dan secara kritis menentukan tindakan pada rekomendasinya.

Untuk memperoleh keuntungan pendekatan DM ini secara penuh, peneliti tanah nasional ditantang menemukan cara bagaimana memanfaatkan *database*, bagaimana mengerti teknik DM, dan bagaimana meningkatkan kemampuan mengoperasikan *software*. Pertama, peneliti tanah harus mengelola data tanah, mengoptimalkan pemanfaatan data, dan memilah informasi yang berguna sehingga data mereka bermanfaat untuk masyarakat. Untuk mengelola data yang lebih baik, peneliti harus membangun suatu *database*, dianjurkan berformat digital. Beberapa *software* pengelolaan *database* (DBMS) seperti MS Access, MS FoxPro, dan Sybase menawarkan perangkat dan panduan bagaimana mengelola data menggunakan sistem.

Dengan adanya data berjumlah besar, pembangunan *database* akan memerlukan waktu lama dan tenaga. Sehubungan dengan itu, peneliti harus menawarkan perencanaan yang baik dalam proses *input* dan penyimpanan data. Bagaimanapun, hal ini lebih mudah untuk mengembangkan *database* melalui suatu transfer elektronik dari file elektronik dan format yang berbeda daripada pengetikan secara manual. Dengan demikian, pengembangan *database* tanah merupakan suatu perencanaan dan pekerjaan sistematis dan strategis.

Kedua, peneliti tanah ditantang untuk menemukan cara bagaimana mengerti algoritma DM itu sendiri. Mereka - seringkali sebagai pemilik data - bebas mempelajari teknologi analisis dan bahkan paling tidak mempelajari mesin. Bagaimanapun, mereka membutuhkan suatu tambahan keahlian "teknologi" untuk mengoperasikan teknologi analisis. Kondisi seperti itu lebih diharapkan sebab: 1) biaya keahlian tersebut tinggi; 2) saat pemodelan telah lengkap, pemilik data tergantung pada sumberdaya dari luar, dan masih mahal; 3) Peneliti tersebut menghilangkan partisipasi langsung dalam analisis, pembatasan penggunaan pengetahuan subyek yang dimilikinya untuk proses feedback yang menghabiskan waktu; dan 4) DM benar-benar lebih merupakan suatu seni daripada suatu ilmu. Hal tersebut tidak efisien jika para profesional tanah harus tetap mengatakan apa yang harus dilakukan para ahli teknologi, dan lantas menunggu beberapa menit atau bahkan berjam-jam atau berhari-hari untuk mendapatkan hasil.

Oleh karena itu, peneliti tanah harus mengerti teknik DM untuk menginterpretasi hasil dan untuk menjelaskan mengapa hasil yang relevan dapat terjadi. Paling tidak, mereka harus mengetahui algoritma dibelakang proses perhitungan dengan mempelajari buku statistik dan matematika termasuk penelusuran dari internet. Saat ini, beberapa perangkat lunak statistik, seperti S-Plus 2000, STATISTICA,

MATLAB, menawarkan penjelasan algoritma yang digunakan di dalam CD penginstal atau vendor website.

Peneliti tanah juga ditantang untuk meningkatkan kemampuannya dalam pengoperasian *software*. Hal tersebut penting karena algoritma *data mining* terlalu kompleks untuk diaplikasikan secara manual. Program komputer akan mempercepat proses perhitungan. Cara lebih baik adalah *learning by doing*. *Software* umumnya menawarkan buku petunjuk yang dirancang untuk pengguna. Menghadiri *training* tentang *software* pengoperasian statistik merupakan alternatif lain.

KESIMPULAN

DM 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 DM pada penelitian tanah nasional diusulkan meliputi pengembangan *database* tanah, seleksi dataset, aplikasi algoritma, dan penyajian dan interpretasi hasil.

Untuk memanfaatkan kesempatan secara penuh hal-hal yang dapat dilakukan oleh teknik *data mining*, peneliti tanah ditantang untuk membuat *database* tanah yang baik, mengerti algoritma *data mining*, dan meningkatkan kemampuannya dalam mengoperasikan *software* statistik.

DAFTAR PUSTAKA

- Balai Penelitian Tanah. 2005. *Database* Tanah. Leaflet. Badan Litbang Pertanian, Departemen Pertanian. Bogor.
- Batjes, N.H. 1995. A global data set of soil pH properties. Technical Paper 27. ISRIC, Wageningen.
- Burrough, P.A. 1991. Soil information System. In D.J. Maguire, M.F. Goodchild, and D. W. Whind (Editors) Geographic Information System: Principles and Applications. Longman Sci and Tech., New York. Pp. 153-169.
- Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth. 2000. CRISP-DM 1.0 Step-by-step *data mining* guide. CRISPWP-0800. (www.spss.com).
- Eastman, J.R. 2001. Guide to GIS and Image Processing Volume I. Clark University. USA.

- Eswaran, H. and J. Kimble. 2003. Land quality assesment and monitoring: the next challenge for soil science. *Pedosphere*, 13(1):1-10.
- FAO. 1995. Global and national soil soils and terrain dijital databases. World Soil Resource report # 74 Rev.1. Rome, Italy.
- FAO. 2003. Information management resource kit v.1.0: Management of electronic document module. Rome, Italy.
- Lavrac, N and M. Grobelnik. 2003. *Data mining*. In D. Mladenic, N. Lavrac, M. Bohanec, and S. Moyle (Editors) *Data mining and Decision Support: Integration and Collaboration*. Kluwer Academic Publisher, Boston.
- Luan, J. 2002. *Data mining* applications in higher education. DMHEWP-0702. (www.spss.com).
- Minasny, B., A.B. McBratney and K.L. Bristow. 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. *Geoderma*, 93:225-253.
- Nugroho, Adi. 2004. Konsep Pengembangan Sistem *Database*. Informatika. Bandung. 508p.
- Nursyamsi, D and Y. Sulaeman. 2001. Phosphorus and Potassium Decision Support System (PKDSS). Makalah disampaikan dalam seminar rutin Puslitbang Tanah dan Agroklimat. 15 January 2001 (Tidak dipublikasikan).
- Nursyamsi, D., L.R. Widowati, D. Setyorini, Y. Sulaeman, and A. Sofyan. 2003. P and K recommendation (P&KR) for food crops in Indonesia. Paper presented on the International workshop on applying information technology for site-specific agriculture in small farms of the tropics, Bangkok, August 4-10th 2003 (Unpublished).
- Pramudiono, I. 2003. Pengantar *Data mining*: Menambang Permata Pengetahuan di Gunung Data. (www.ilmukomputer.com).
- Setyorini, D., L.R. Widowati, dan J. S. Adiningsih. 2003. Validasi model rekomendasi pemupukan tanaman pangan. Laporan akhir. Balai Penelitian Tanah. (Tidak dipublikasikan)
- SPSS Inc. 1999a. *Data mining*: an introduction. WPDMINTR0-0699. (www.spss.com).
- SPSS Inc. 1999b. AnswerTree Algorithm Summary. ATALGWP-0599. (www.spss.com).

- SPSS Inc. 2001. The SPSS C&RT Component: A decision tree component enabling more effective classification and prediction of target variables. CRTWP-0101. (www.spss.com).
- StatSoft. 1999. STATISTICA for window [computer program manual]. Tulsa.OK
- Sulaeman, Y dan D. Nursyamsi. 2001. Perangkat lunak PKDSS: suatu pengantar. Puslitbang Tanah dan Agroklimat. Bahan workshop Pembinaan Penelitian Kalibrasi Uji Tanah Hara P dan K Lahan Kering. (Tidak dipublikasikan).
- Sulaeman, Y. dan D. Nursyamsi. 2005. PKDSS v.2.0. Balittanah, Bogor.
- Tempel, P, N.H. Batjes, and V.W.P. van Engelen. 1996. IGDP-DIS Soil dataset for pedotransfer function development. Working paper and reprint 96/05. International Soil Reference Center (ISRIC), Wageningen.
- Widowati, L.R., D. Setyorini, dan A. Darmawan. 2004. Validasi model rekomendasi pemupukan lahan sawah pada tanah Inceptisol bertekstur kasar di Pati. *In* K. Subagyo (Editor in cheft) Prosiding Seminar Nasional Inovasi Teknologi Sumberdaya Tanah dan Iklim, Bogor 14-15 September 2004. Puslitbangtanak. Bogor. Pp. 431-448.
- Wilding, L.P. and L.R. Dress. 1983. Spatial variability and pedology. In L.P. Wilding, N.E.Smeck, and G.F. Hall (Editors) Pedogenesis and Soil Taxonomy: I. Concepts and Interactions. Elseiver, Amsterdam. Pp 83-116.