

Komparasi Algoritma C4.5, Naïve Bayes Dan Neural Network Untuk Klasifikasi Tanah

Amirul Mukminin¹, Dwiza Riana²

¹Universitas BSI
e-mail: amirul.amk@bsi.ac.id

²STMIK Nusa Mandiri Jakarta
e-mail: dwiza@nusamandiri.ac.id

Abstrak

Penentuan jenis tanah pada kedalaman tertentu untuk kebutuhan perencanaan pembangunan perumahan dilakukan berdasarkan data Cone Penetration Test. Tujuan penelitian ini untuk mengkomparasi Algoritma C4.5, Naïve Bayes, and Neural Network sehingga ditemukan pemodelan yang terbaik untuk mengklasifikasikan tanah. Hasil dari penelitian ini didapatkan algoritma terbaik yaitu Algoritma C4.5. Algoritma C4.5 dalam klasifikasi dua kelas mencapai akurasi 98,45% dan AUC 0,981. Dalam klasifikasi tiga kelas C4.5 juga mencapai akurasi tertinggi (93,21%), demikian juga pada klasifikasi tujuh kelas (83,40%). Hasil penelitian ini menyimpulkan bahwa Algoritma C 4.5 dapat dijadikan pilihan dalam mengklasifikasi tanah untuk pembangunan perumahan.

Kata Kunci : Data Mining, Klasifikasi Tanah, C4.5, Naïve Bayes, Neural Network

Abstract

Determining the type of soil at a certain depth to the needs of residential development planning is done based on the data Cone Penetration Test. The purpose of this research to compare the data mining algorithm C4.5, Naïve Bayes, and Neural Network to find the best modeling can be used for land classification. The results of this research, the best algorithm is C4.5. Algoritma C45 in binary-class classification accuracy reaches 98% and AUC 0,981. In the three-class classification C4.5 also have scored the highest accuracy (93.21%), as well as on the seven-class classification (83.40%). The results of this research concluded that the algorithm C 4.5 can be selected in classifying soil for residential development.

Keywords: Data mining, C4.5, Naïve Bayes, Neural Network, Soil Classification.

1. Pendahuluan

Klasifikasi tanah untuk pembangunan perumahan dapat dilakukan melalui *index properties* dan *mechanical properties (engineering properties)*. *Index properties* diperoleh melalui percobaan di laboratorium, sedangkan *mechanical properties* dapat diambil langsung di lapangan atau melalui percobaan benda uji (*sample*) di laboratorium (Robertson, 2009). Mengingat Laporan Sondir atau *Cone Penetrometer Test (CPT)* merupakan salah syarat untuk mengurus Ijin Mendirikan Bangunan di Kota Bandung (Bandung, 2011) sehingga analisa hasil dari uji dilapangan diharapkan cepat diselesaikan.

Uji Sondir atau *Cone Penetrometer Test (CPT)* merupakan salah uji

penetrometer statis di lapangan (*in situ testing*) yang umum digunakan. Uji sondir untuk mendapatkan parameter (angka) *mechanical properties* kekuatan tanah. *Output* uji sondir menghasilkan tiga parameter, yaitu: gesekan selimut (kg/cm), daya dukung ujung konus (kg/cm²) dan rasio gesekan (%). Uji sondir memiliki keandalan uji yang dianggap cukup baik dan cepat karena dilakukan langsung di lapangan dengan hasil uji yang langsung dapat diketahui (Robertson, 2009).

Dalam penelitian, tanah diklasifikasikan kedalam jenis *gravel, sand, silt/loam, clay, heavy clay, peat*, dan *organic/peat* (Bhattacharya & Solomatine, 2005). Berdasarkan data penyelidikan lapangan dengan alat *Cone Penetration Test*

(CPT) masih sulit mengenali beberapa jenis tanah yang berbeda-beda tapi mempunyai sifat serupa ke dalam kelompok-kelompok dan subkelompok-subkelompok berdasarkan pemakaiannya. Grafik klasifikasi *Cone Penetration Test (CPT)* tidak dapat diharapkan untuk memberikan prediksi yang akurat dari jenis tanah berdasarkan ukuran bulir tetapi dapat memberikan panduan untuk jenis perilaku tanah (Robertson, 2009).

Salah satu cara yang efektif untuk menganalisis semua sumber informasi dalam rangka untuk menemukan trend dan pola adalah melalui *Business Intelligence* dan teknik *data mining*, untuk membangun model *data-driven* dan kemudian mengambil pengetahuan yang berguna. Menurut (Chapman, 2000), *The Cross-Industry Standard Process for Data Mining (CRISP-DM)* adalah metodologi yang populer untuk meningkatkan keberhasilan proyek *data mining* (Moro, Cortez, & Laureano, 2013). CRISP-DM memungkinkan pembangunan dan implementasi model data mining ke dalam enam fase, dimana beberapa proses iterasi dilakukan untuk menyesuaikan proyek *data mining* terhadap tujuan penelitian untuk mendukung keputusan penelitian. *Data Mining* merupakan alat dan teknik yang digunakan untuk mengekstrak informasi yang berarti dari data yang dikumpulkan, memungkinkan pengguna untuk membuat proses pengambilan keputusan yang lebih baik (Jayasree & Balan, 2013).

Di antara metode *data mining*, *Decision Tree (DT)*, *Naive Bayes (NB)*, dan *Neural Network (NN)* dapat digunakan untuk mengklasifikasikan sekumpulan data dengan mengidentifikasi pola dari input data. Dalam beberapa penelitian, *Decision Tree* memiliki akurasi yang paling tinggi diantara *Naive Bayes* dan *Neural Network* dalam prediksi penyakit jantung (Alfisahrin, 2014). Sementara itu, pendekatan dengan menggunakan *Naive Bayes* memiliki kinerja yang cukup tinggi untuk mengklasifikasikan data (Hamzah, 2012). Dalam penelitian ini, akan dilihat perbandingan dari hasil klasifikasi *Decision Tree*, *Naive Bayes* dan *Neural Network*.

Rumusan masalah yang terkait dengan permasalahan penelitian adalah algoritma mana yang memiliki performa paling baik diantara *Decision Tree (C4.5)*, *Naive Bayes* dan *Neural Network* dalam mengklasifikasikan *Dataset Cone Penetration Test*? Hasil pengolahan data

menggunakan algoritma yang terbaik akan dapat digunakan sebagai salah satu dasar untuk perencanaan pembangunan perumahan.

Klasifikasi Tanah

Sistem klasifikasi tanah adalah suatu sistem pengaturan beberapa jenis tanah yang berbeda-beda tapi mempunyai sifat yang serupa ke dalam kelompok-kelompok dan subkelompok-subkelompok berdasarkan pemakaiannya (Das, 2005). Sistem klasifikasi tanah dibuat dengan tujuan untuk memberikan informasi karakteristik dan sifat-sifat fisis tanah. Karena sifat dan perilaku tanah yang begitu beragam, sistem klasifikasi mengelompokkan tanah ke dalam kategori yang umum dimana tanah memiliki kesamaan sifat fisik. Klasifikasi tanah juga berguna untuk studi yang terperinci mengenai keadaan tanah tersebut serta kebutuhan pengujian untuk menentukan sifat teknis tanah seperti karakteristik pematatan, kekuatan tanah, berat isi dan sebagainya (Bowles, 1989).

Secara umum pendekatan klasifikasi tanah menggunakan perkiraan morfologi tanah dan uji laboratorium untuk menentukan jenis tanah. Pendekatan lain dalam menentukan jenis tanah adalah dengan klasifikasi *numeric* atau metode perkiraan. Dimana setiap lapisan tanah dikelompokkan dengan metode statistik *multivariate* seperti analisa klaster. Hal ini menghasilkan pengelompokan alami tanpa memerlukan penelitian tentang genesis tanah (Bhargavi & Jyothi, 2011).

Sifat dan perilaku tanah merupakan hal yang telah menarik banyak peneliti untuk melakukan pemodelan. Mengembangkan korelasi rekayasa antara berbagai parameter tanah merupakan isu yang banyak dibahas. Menggunakan *neural network* untuk model korelasi antara kepadatan relatif dan resistensi kerucut dari *Cone Penetration Test (CPT)*, untuk kedua pasir terkonsolidasi normal dan over-konsolidasi berhasil pemodelan hubungan non-linear antara resistensi CPT kerucut dan parameter lainnya (M.A. Shahin, 2008).

Data hasil *Cone Penetration Test (CPT)* dapat digunakan untuk menentukan klasifikasi tanah dengan mempertimbangkan nilai tahanan konus (Q_c) dan tahanan selimut (f_s) (M.A. Shahin, 2008).

Cone Penetration Test (CPT)

Cone Penetration Test (CPT) atau lebih sering disebut *sondir* adalah salah satu

survey lapangan yang berguna untuk memperkirakan letak lapisan tanah keras. Tes ini baik dilakukan pada lapisan tanah lempung. Dari tes ini didapatkan nilai perlawanan penetrasi konus. Perlawanan penetrasi konus adalah perlawanan tanah terhadap ujung konus yang dinyatakan dalam gaya per satuan luas. Sedangkan hambatan lekat adalah perlawanan geser tanah terhadap selubung bikonus dalam gaya per satuan panjang. Nilai perlawanan penetrasi konus dan hambatan lekat dapat diketahui dari bacaan pada manometer (BSN, 2008). Selengkapnya parameter yang didapatkan dari uji sondir berupa perlawanan konus (q_c), perlawanan geser (f_s), angka banding geser (R_f), dan geseran total tanah (T_f), yang dapat digunakan untuk interpretasi perlawanan tanah yang merupakan bagian dari desain fondasi (Hardiyatmo, 2010). *Cone Penetration Test atau Sondir* memiliki tiga aplikasi utama:

1. Untuk menentukan profil tanah dan mengidentifikasi tanah.
2. *Interpolasi* kondisi tanah antara lubang bor kontrol.
3. Untuk mengevaluasi parameter rekayasa dari tanah dan untuk menilai daya dukung dan penyelesaian pondasi (Robertson, 2009).

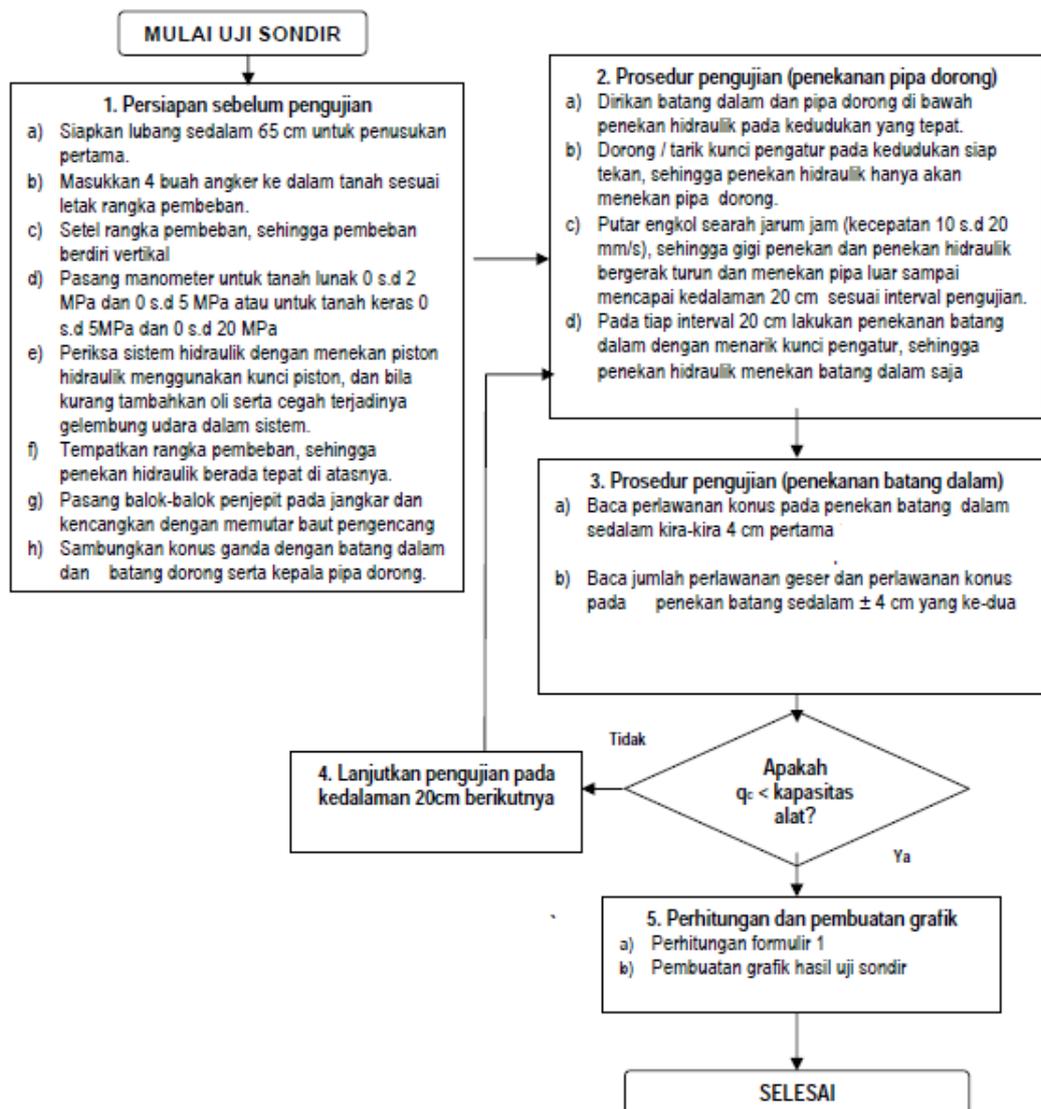
Cone Penetration Test (CPT) dilakukan dengan menggunakan alat CPT atau alat Sondir. Alat-alat *Cone Penetration Test (CPT)* yang digunakan adalah sebagai berikut (BSN, 2008):

1. Konus.

Konus merupakan ujung alat penetrasi yang berbentuk kerucut untuk menahan perlawanan tanah.

2. Selimut (bidang) geser.
Selimut (bidang) geser adalah bagian ujung alat ukur penetrasi ganda, tempat terjadinya perlawanan geser lokal.
3. Batang dalam.
Batang dalam terbuat dari bahan baja dan terletak di dalam pipa dorong, diameter luar batang dalam harus konstan. Batang dalam mempunyai penampang melintang yang dapat menyalurkan perlawanan konus tanpa mengalami tekuk atau kerusakan lain.
4. Pipa dorong.
Pipa terbuat dari bahan baja dengan panjang 1,00 m, pipa harus menerus sampai konus ganda agar penampang pipa tidak tertekuk jika di sondir atau didorong.
5. Mesin pembebanan hidrolik.
Rangka mesin pembeban berfungsi sebagai dudukan sistem penekan hidrolik yang dapat digerakkan naik dan turun, sistem penekan hidrolik terdiri atas engkol pemutar, rantai, roda gigi, gerigi dorong dan penekan hidrolik yang berfungsi untuk mendorong dan menarik batang dalam dan pipa dorong. Pada penekan hidrolik terpasang 2 buah manometer yang digunakan untuk membaca tekanan hidrolik yang terjadi pada waktu penekanan batang dalam, pipa dorong dan konus (tunggal atau ganda).

Melakukan *Cone Penetration Test (CPT)* harus mengikuti aturan langkah-langkah operasionalnya untuk mendapatkan data yang valid. Dibawah ini diagram alir tata cara pengujian *Cone Penetration Test (CPT)* (BSN, 2008) :



Gambar 1. Alur Uji CPT
Sumber : BSN (2008)

Decision Tree (C4.5)

Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan yang merupakan pengembangan dari algoritma ID3, algoritma tersebut digunakan untuk membentuk pohon keputusan (Alfisahrin, 2014). Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer. Dalam klasifikasi pohon keputusan terdiri dari sebuah *node* yang membentuk akar. *Node* akar tidak memiliki *input*. *Node* lain yang bukan sebagai akar tetapi memiliki tepat satu *input* disebut *node internal* atau *test node*, sedangkan *node* lainnya dinamakan daun. Daun mewakili nilai target yang paling

tepat dari salah satu *class* (Maimon & Rokach, 2010).

Pohon keputusan dibuat dengan membagi nilai-nilai atribut menjadi cabang untuk setiap kemungkinan. Cara kerja pohon keputusan yaitu dengan melakukan penelusuran dari akar hingga ke cabang sampai *class* suatu objek ditemukan. *Instance* diklasifikasikan dengan mengarahkan dari akar pohon sampai ke daun sesuai dengan hasil tes melalui *node internal* (Alfisahrin, 2014).

Pada dasarnya konsep dari algoritma C4.5 adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (*rule*). C4.5 adalah algoritma

yang cocok untuk masalah klasifikasi dan *data mining*. C4.5 memetakan nilai atribut menjadi *class* yang dapat diterapkan untuk klasifikasi baru (Wu & Kumar, 2009).

Langkah-langkah membangun pohon keputusan menggunakan algoritma C4.5 menurut Kusri dan Luthfi adalah sebagai berikut (Alfisahrin, 2014):

1. Pilih atribut sebagai akar.

Pemilihan atribut sebagai akar berdasarkan pada nilai *gain* tertinggi dari atribut yang ada. Untuk menghitung nilai *gain* tertinggi digunakan persamaan 1 dibawah ini :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Dimana:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke-i

|S| : Jumlah kasus dalam S

Nilai *Entropy* dapat dihitung dengan persamaan 2 dibawah ini :

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Dimana:

S : Himpunan kasus

n : Jumlah partisi S

Pi : Proporsi dari Si terhadap S

2. Buat cabang untuk tiap-tiap nilai.

3. Bagi kasus dalam cabang.

Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Naive Bayes

Naive Bayes merupakan metode yang tidak memiliki aturan (Alfisahrin, 2014). *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada *data training*. *Naive Bayes* merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam *data mining* (Wu & Kumar, 2009). Algoritma ini juga dikenal dengan nama *Idiot's Bayes*, *Simple Bayes*, dan *Independence Bayes* (Bramer, 2013).

Klasifikasi *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Klasifikasi *Bayesian* didasarkan pada teorema Bayes, diambil dari nama seorang ahli matematika yang juga menteri *Prebysterian* Inggris,

Thomas Bayes (1702-1761) (Bramer, 2013). Menurut Kusri dan Luthfi, klasifikasi bayesian memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* (Alfisahrin, 2014).

Bayes rule digunakan untuk menghitung probabilitas suatu *class*. Algoritma *Naive Bayes* memberikan suatu cara mengkombinasikan peluang terdahulu dengan syarat kemungkinan menjadi sebuah formula yang dapat digunakan untuk menghitung peluang dari tiap kemungkinan yang terjadi (Alfisahrin, 2014).

Bentuk umum dari teorema bayes dapat dilihat pada persamaan 3 berikut:

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

Dimana :

$P(A)$ dan $P(B)$ adalah peluang dari pengamatan A dan B tanpa memperhatikan satu sama lain.

$P(A|B)$ berarti peluang kejadian A bila B terjadi.

$P(B|A)$ peluang kejadian B bila A terjadi.

Naive Bayes adalah penyederhanaan metode bayes. Teorema bayes disederhanakan menjadi persamaan 4 berikut:

$$P(H|X) = P(X|H) P(X)...[4]$$

Bayesian rule diterapkan untuk menghitung *posterior* dan probabilitas dari data sebelumnya. Dalam analisis bayesian, klasifikasi akhir dihasilkan dengan menggabungkan kedua sumber informasi (*prior* dan *posterior*) untuk menghasilkan *probabilitas* menggunakan aturan bayes.

Neural Network

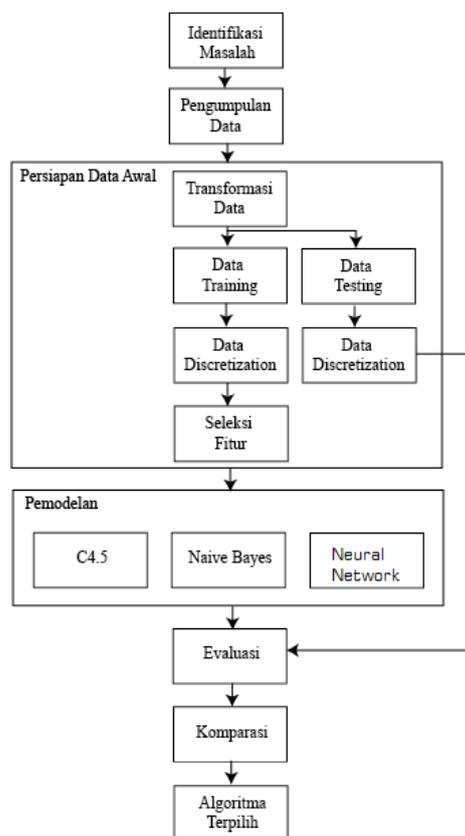
Neural network atau jaringan saraf adalah satu set unit *input/output* yang terhubung, dimana tiap koneksinya memiliki bobot. Selama fase pembelajaran, *neural network* menyesuaikan bobot sehingga dapat memprediksi *class* yang benar dari *tupple* (Han & Kamber, 2006). Informasi atau *input* akan dikirim ke *neuron* dengan bobot kedatangan tertentu, *input* ini akan diproses oleh suatu fungsi perambatan yang akan menjumlahkan nilai-nilai semua bobot yang akan datang. Hasil penjumlahan tersebut akan dibandingkan dengan suatu nilai ambang (*threshold*) tertentu melalui fungsi aktivasi setiap *neuron* (Kusumadewi, 2004).

Algoritma *neural network* yang paling populer adalah *Backpropagation*,

algoritma *backpropagation* melakukan pembelajaran pada jaringan saraf *multi layer feed forward* yang terdiri dari tiga lapisan/*layer*, yaitu: lapisan *input*, lapisan tersembunyi, dan lapisan keluaran (Han & Kamber, 2006). Pada beberapa diagram *neural network* dimungkinkan terdapat lebih dari satu lapisan tersembunyi, meskipun kebanyakan hanya mengandung satu lapisan tersembunyi yang dirasa cukup untuk berbagai tujuan (Larose, 2006).

2. Metode Penelitian

Gambaran mengenai metodologi penelitian klasifikasi jenis tanah ini dapat dilihat pada Gambar 2. di bawah ini :



Gambar 2. Desain Penelitian

Dalam penelitian ini dilakukan beberapa langkah yang dilakukan dalam proses penelitian :

1. Identifikasi masalah

Permasalahan dalam penelitian ini adalah untuk menentukan pola klasifikasi yang baik atau buruk dalam set data yang digunakan. Dengan menggunakan *data mining*, diharapkan dapat dikembangkan sebuah sistem pendukung keputusan yang baik dalam

mengklasifikasikan jenis tanah yang konsisten, cepat dan tepat.

2. Pengumpulan data

Pada tahap ini ditentukan data yang akan diproses. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data kedalam *dataset*, termasuk variabel yang diperlukan dalam proses. Penelitian ini menggunakan data *cone penetration test* dari lapangan yang sudah dilakukan pengolahan oleh ahli geoteknik di Konsultan Geoteknik CV. Tatabumi Indonesia, Bandung. Data yang dikumpulkan adalah data yang dibuat tahun 2015.

3. Pengolahan data awal

Ditahap ini dilakukan pengurutan data, penyeleksian data, data dibersihkan dari *missing value* atau *noise*, membuang atribut yang tidak diperlukan, dan data ditransformasikan ke bentuk yang dapat digunakan sebagai data input ke dalam *Software Rapidminer* sehingga dapat dilakukan persiapan dalam pembuatan model.

4. Metode yang diusulkan

Pada tahap ini data dianalisis, dikelompokkan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model-model yang sesuai dengan jenis data. Pembagian data kedalam data latihan (*training data*) dan data uji (*testing data*) juga diperlukan untuk pembuatan model.

5. Modeling

Pada tahap ini model yang diusulkan akan diuji untuk melihat hasil berupa *rule* yang akan dimanfaatkan dalam pengambilan keputusan. Ekperimen yang dilakukan pada penelitian ini menggunakan teknik klasifikasi data mining yang berbeda, yaitu : *C4.5*, *Naive Bayes*, dan *Neural Network*.

Pemodelan dilakukan dengan menggunakan *Software Rapidminer* yang merupakan salah satu *software* pemodelan terbaik saat ini yang menyediakan lingkungan kerja yang bagus untuk tugas *data mining*.

6. Evaluasi dan validasi

Pada tahap ini dilakukan evaluasi terhadap model yang ditetapkan untuk mengetahui tingkat keakurasian model. Pada tahap ini, akan dilakukan evaluasi

terhadap kualitas dan efektivitas dari masing-masing model yang diterapkan. Proses evaluasi menggunakan metode ROC (*Receiver Operating Characteristic*) (Fawcett, 2005) untuk mencari nilai AUC (*Area Under the ROC Curve*) tertinggi dengan *software Rapidminer*. AUC merupakan teknik untuk memvisualisasikan, mengatur dan memilih pengklasifikasian berdasarkan kinerja model (Akthar & Hahne, 2012).

7. Komparasi

Setelah diketahui hasil evaluasi dari ketiga model pada masing-masing kelompok atribut, dilakukan komparasi untuk melihat algoritma terbaik berdasarkan tingkat akurasi dan kategori pemodelan klasifikasi pada kurva ROC.

Tahap evaluasi dengan menggunakan *software rapidminer* dapat dilakukan dengan cara memasukkan *data testing* seperti yang sudah dilakukan pada proses *modeling* kemudian memasukkan operator *Apply Model* dan *Performance* pada panel operator.

3. Hasil dan Pembahasan

Data yang digunakan untuk *training* dan *testing* adalah data *Cone Penetration Test (CPT)* yang telah diolah oleh ahli geoteknik sebanyak 1115 *record*.

Masing-masing *record* data terdiri dari 8 atribut yang berupa atribut numerik seperti pada Tabel 1 dibawah ini :

Tabel 1. Atribut dalam Klasifikasi Tanah

Atribut	Nilai		Tipe
Location	Lokasi tes		Teks
Depth	Kedalaman konus		Numerik
C	Tekanan konus		Numerik
C+F	Total resisten		Numerik
F	Friksi lokal		Numerik
HL	Friksi		Numerik
JHL	Jumlah Friksi		Numerik
I	Index soil behaviour type		Numerik

Dalam penelitian ini dilakukan eksperimen dengan tiga macam klasifikasi, yaitu :

- Klasifikasi dua kelas, dimana dalam klasifikasi ini jenis tanah hanya diklasifikasikan kedalam tanah yang tergolong berpasir atau tidak berpasir (sandy dan non-sandy).

- Klasifikasi tiga kelas, pada klasifikasi ini kelas dibagi tiga yang merupakan jenis dasar tanah, yaitu pasir, liat, dan gambut (sand, clay, dan peat).
- Klasifikasi tujuh kelas, klasifikasi yang lebih detil yang dapat diidentifikasi pada area yang di observasi, yaitu organik/gambut (organic / peat), gambut (peat), liat keras (heavy clay), liat (clay), lanau (silt/loam), pasir (sand), dan kerikil (gravel).

Klasifikasi dilakukan dengan menggunakan *Software Rapidminer* dengan *rule* berbeda sesuai dengan model yang digunakan (Jungermann, 2009).

3.1. Algoritma C4.5

Pemodelan *Decision Tree* dilakukan dengan nilai gain tertinggi dari setiap atribut. Pembuatan *Decision Tree* dengan menggunakan *Software Rapidminer* dengan pengaturan parameter *default* dan *criterion information_gain*. Hasil dari eksperimen dapat dilihat pada Tabel 2 dibawah ini :

Tabel 2. Hasil Eksperimen Dua Kelas Tanah dengan C4.5

No	Kelas Tanah	Presisi Kelas
1	SANDY	99.17%
2	CLAY	97.93%
	Rata-rata	98.55%

Eksperimen dengan dua kelas menunjukkan **Akurasi 98,49 %**. Sedangkan hasil yang di peroleh dari pengolahan kurva ROC dapat dilihat pada Gambar 5 menghasilkan nilai **AUC sebesar 0,981** yang masuk kategori *Excellent Classification*.



Gambar 3. Kurva ROC Algoritma C4.5 dengan Dua Kelas

Sedangkan hasil eksperimen Algoritma C4,5 dengan tiga kelas dapat dilihat pada Tabel 3 seperti dibawah ini :

Tabel 3. Hasil Eksperimen Tiga Kelas Tanah dengan C4.5

No	Kelas Tanah	Presisi Kelas
1	SANDY	99,17 %
2	CLAY	94,52 %
3	PEAT	81,94 %
	Rata-rata	91,33 %

Pada eksperimen dengan tiga kelas ini menunjukkan besaran **Akurasi 93,21%**. Dilihat dari hasil presisi kelas, kelas tanah SANDY dengan 2 kali eksperimen yaitu dengan dua kelas dan tiga kelas menunjukkan nilai presisi tetap yaitu 99,17%.

Hasil eksperimen Algoritma C4,5 dengan tujuh kelas dapat dilihat pada Tabel 4 seperti dibawah ini :

Tabel 4. Hasil Eksperimen Tujuh Kelas Tanah dengan C4.5

No	Kelas Tanah	Presisi Kelas
1	ORGANIC/PEAT	0
2	PEAT	63,64 %
3	HEAVY CLAY	72,41 %
4	CLAY	69,24 %
5	SILT/LOAM	77,14 %
6	SAND	100 %
7	GRAVEL	0
	Rata-rata	54,63 %

Pada eksperimen dengan tujuh kelas ini menunjukkan besaran **Akurasi 71,32%**.

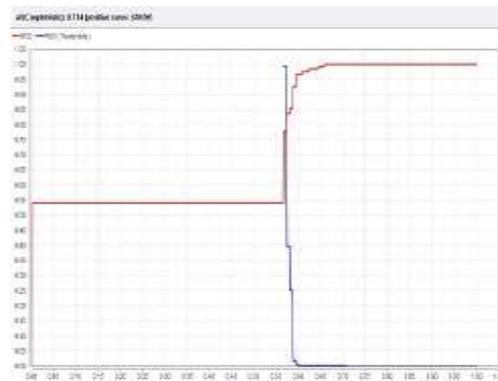
3.2. Algoritma Naïve Bayes

Hasil dari eksperimen dengan Algoritma *Naïve Bayes* dengan dua kelas tanah dapat dilihat pada Tabel 5 dibawah ini:

Tabel 5. Hasil Eksperimen Dua Kelas Tanah dengan Naïve Bayes

No	Kelas Tanah	Presisi Kelas
1	SANDY	97,14 %
2	CLAY	73,49 %
	Rata-rata	85,32 %

Eksperimen dengan dua kelas menunjukkan **Akurasi 80,51 %**. Sedangkan hasil yang di peroleh dari pengolahan kurva ROC dapat dilihat pada Gambar 6 menghasilkan nilai **AUC sebesar 0,734** yang masuk kategori *Fair Classification*.



Gambar 4. Kurva ROC Algoritma Naïve Bayes dengan Dua Kelas

Sedangkan hasil eksperimen Algoritma *Naïve Bayes* dengan tiga kelas dapat dilihat pada Tabel 6 seperti dibawah ini :

Tabel 6. Hasil Eksperimen Tiga Kelas Tanah dengan Naïve Bayes

No	Kelas Tanah	Presisi Kelas
1	SANDY	97,50 %
2	CLAY	67,31 %
3	PEAT	79,17 %
	Rata-rata	81,33 %

Pada eksperimen dengan tiga kelas ini menunjukkan besaran **Akurasi 80,17%**.

Hasil eksperimen dengan tujuh kelas dapat dilihat pada Tabel 7 dibawah ini:

Tabel 7. Hasil Eksperimen Tujuh Kelas Tanah dengan Naïve Bayes

No	Kelas Tanah	Presisi Kelas
1	ORGANIC/PEAT	0
2	PEAT	14,29 %
3	HEAVY CLAY	71,43 %
4	CLAY	48,78 %
5	SILT/LOAM	88,57 %
6	SAND	0
7	GRAVEL	0
	Rata-rata	31,86 %

Pada eksperimen dengan tujuh kelas ini menunjukkan besaran **Akurasi 63,92%**.

3.3. Algoritma Neural Network

Hasil dari eksperimen dengan Algoritma *Neural Network* dengan dua kelas dapat dilihat pada Tabel 8 dibawah :

Tabel 8. Hasil Eksperimen Dua Kelas Tanah dengan *Neural Network*

No	Kelas Tanah	Presisi Kelas
1	SANDY	85.58 %
2	CLAY	79.50 %
	Rata-rata	82.54 %

Eksperimen dengan dua kelas menunjukkan **Akurasi 81,89 %**. Sedangkan hasil yang di peroleh dari pengolahan kurva ROC dapat dilihat pada gambar menghasilkan nilai **AUC sebesar 0,993** yang masuk kategori *Excellent Classification*.



Gambar 5. Kurva ROC Algoritma *Neural Network* dengan 2 Kelas

Sedangkan hasil eksperimen Algoritma *Neural Network* dengan tiga kelas dapat dilihat pada Tabel 9 seperti dibawah ini :

Tabel 9. Hasil Eksperimen Tiga Kelas tanah dengan *Neural Network*

No	Kelas Tanah	Presisi Kelas
1	SANDY	82.35 %
2	CLAY	80.65 %
3	PEAT	85,19 %
	Rata-rata	82.73 %

Pada eksperimen dengan tiga kelas ini menunjukkan besaran **Akurasi 83,40%**.

Hasil eksperimen dengan tujuh kelas dapat dilihat pada Tabel 10 dibawah ini:

Tabel 10. Hasil Eksperimen Tujuh Kelas Tanah dengan *Neural Network*

No	Kelas Tanah	Presisi Kelas
1	ORGANIC/PEAT	0
2	PEAT	65,75 %
3	HEAVY CLAY	65,71 %
4	CLAY	66,67 %
5	SILT/LOAM	82,54 %
6	SAND	37,70 %
7	GRAVEL	0
	Rata-rata	45.48 %

Pada eksperimen dengan tujuh kelas ini menunjukkan nilai **Akurasi 63,40%**.

3.4. Komparasi Hasil Terbaik

Pada tahap ini dilakukan proses komparasi akurasi hasil eksperimen yang telah dilakukan yaitu klasifikasi menggunakan algoritma *C4.5*, *Naïve Bayes*, *Neural Network* dengan hasil pada Tabel 11.

Tabel 11. Hasil Komparasi Algoritma *C4.5*, *Naïve Bayes*, dan *Neural Network*

Kelas	Algoritma		
	<i>C4.5</i>	<i>Naïve Bayes</i>	<i>Neural Network</i>
Dua Kelas	98.46%	80.51%	81.89%
Tiga Kelas	93.21%	80.17%	83.40%
Tujuh Kelas	71.32%	63.92%	63.40%

Dilihat dari Tabel 11 diatas maka model klasifikasi menggunakan Algoritma *C4.5* memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan Algoritma *Naïve Bayes* dan *Neural Network* pada dua kelas, tiga kelas, atau tujuh kelas.

Hasil penelitian ini dapat digunakan sebagai salah satu model untuk mengklasifikasi tanah untuk kebutuhan pembangunan perumahan.

4. Kesimpulan

Penelitian ini membandingkan beberapa algoritma *classifier* seperti *C4.5*, *naïve bayes* dan *neural network* untuk mengklasifikasi *dataset cone penetration test*.

Berdasarkan hasil eksperimen pada penelitian ini dapat ditarik kesimpulan bahwa dari hasil evaluasi dan validasi diketahui Algoritma *C4.5* dengan kelompok dua kelas memiliki akurasi yang paling tinggi. Algoritma *C4.5* dapat dijadikan alternatif model klasifikasi baik pada dua kelas, tiga kelas, maupun tujuh kelas, karena pada setiap kelas nilai akurasi Algoritma *C4.5* lebih tinggi dari model *Naïve Bayes* dan *Neural Network*.

Algoritma *C4.5* dapat digunakan untuk mengklasifikasikan tanah berdasarkan *dataset Cone Penetration Test* yang sulit diklasifikasikan kedalam jenis tanah *organic/peat*, *peat*, *heavy clay*, *clay*, *silt/loam*, *sand*, atau *gravel*. Hal ini tentu memudahkan untuk mengetahui jenis tanah pada kedalaman tertentu sebagai salah satu

pertimbangan dalam perencanaan pembangunan perumahan.

Tingkat keberhasilan penelitian dapat ditingkatkan dengan penambahan data yang diolah dalam penelitian dan mengambil data jenis tanah yang lebih beragam dari berbagai area. Algoritma terbaik dalam penelitian ini dapat di komparasi dengan metode klasifikasi yang lain sehingga didapatkan algoritma yang paling akurat.

Referensi

- Akthar, F., & Hahne, C. (2012, August 24). *Rapidminer 5 "Operator Reference"*. Retrieved from Rapid-I GmbH: www.rapid-i.com.
- Alfisahrin, S. N. (2014). *Komparasi Algoritma C4.5, Naive Bayes dan Neural Network Untuk Memprediksi Penyakit Jantung*. Jakarta: Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
- Bandung, P. (2011). *Penyelenggaraan, Retribusi IMB dan Retribusi Penggantian Cetak Peta*. Bandung: Pemkot Bandung:Perda Kota Bandung Nomor 12 Tahun 2011.
- Bhargavi, P., & Jyothi, D. (2011). Soil Classification Using Data Mining Techniques : A Comparative Study. *International Journal of Engineering Trends and Technology-July to Aug Issue*.
- Bhattacharya, B., & Solomatine, D. (2005). *Machine Learning in Soil Classification*. Montreal, Canada.
- Bowles, E. (1989). *Sifat-sifat Fisis dan Geoteknis Tanah*. Jakarta: PT. Erlangga.
- Bramer, M. (2013). *Principle of Data Mining Second Edition*. London: Springer.
- BSN. (2008). *Cara Uji Penetrasi Lapangan Dengan Alat Sondir*. Bandung: BSN : SNI 2827:2008.
- Chapman, P. e. (2000). *CRISP-DM 1.0 - Step-by-Step Data Mining Guide*. . CRISP-DM Consortium.
- Das, D. (2005). *Fundamentals Of Geotechnical Engineering 2nd*. Thomson USA.
- Fawcett, T. (2005). *An Introduction to ROC Analysis*. Elsevier, 861.
- Hamzah, A. (2012). *Klasifikasi Teks Dengan Naive Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita dan Abstract Akademis*. Yogyakarta: Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III.
- Han, J., & Kamber, M. (2006). *Data Mining Concepts and Technoloques Second Edition*. San Francisco: Diane Cerra.
- Hardiyatmo, H. (2010). *Analisis dan Perancangan Pondasi*. Yogyakarta: Gadjah Mada University.
- Jayasree, & Balan. (2013). A Review On Data Mining In Banking Sector. *American Journal of Applied Sciences 10*.
- Jungermann, F. (2009). Information Extraction with Rapidmanner. *In Proceedings Of The GSCL Symposium "Sprachtechnologie und eHumanities"* (p. 50). Duisburg: Universitat Diusburg, Essen.
- Kusumadewi, S. (2004). *Membangun Jaringan Syaraf Tiruan Menggunakan Matlab & Exel Link*. Yogyakarta: Graha Ilmu.
- Larose, D. (2006). *Data Mining Methods and Models*. New Jersey: John Wiley and Son, Inc.
- M.A. Shahin, J. M. (2008). Future Challenge for Artifical Neural Network Modelling in Geotechnical Engineering. *International Association for Computer Methods and Advancees in Geomechanics (IACMAG)*. Goa-India.
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook Second Edition*. New York: Springer.

Moro, S., Cortez, P., & Laureano, R. S. (2013). *A Data Mining Approach for Bank Telemarketing Using the rminer Package and R Tool*. Lisbon: Business Research Unit Instituto Universitario de Lisboa.

Robertson, P. (2009). *Soil Behaviour Type From The CPT : an update*. California, USA: Gregg Drilling & Testing Inc.

Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. New York: CRC Press.