



Klasifikasi Kualitas Tanah Menggunakan Metode Naive Bayes Classifier

Iis Dewi Ratih¹, S M Retnaningsih², V M Dewi³

^{1,2,3}Departemen Statistika Bisnis, Institut Teknologi Sepuluh Nopember, Indonesia

Email: ¹iisdewiratih@gmail.com, ³vivimentari9@gmail.com

Article Info

Article history:

Received Oct 9, 2022

Revised Nov 20, 2022

Accepted Dec 20, 2022

Keywords:

Klasifikasi

Korosivitas Tanah

Naïve Bayes Classifier

Sistem Perpipaan

ABSTRACT

Korosi merupakan kerusakan atau penurunan mutu dari material akibat bereaksi dengan lingkungan. Hal ini menjadi permasalahan serius dalam kegiatan industri dan infrastruktur. Penggunaan sistem perpipaan dalam kegiatan industri dan infrastruktur yang semakin meningkat, dianggap memiliki tingkat integritas yang tinggi serta lebih efektif dan efisien dibanding sistem transportasi lain. Dalam sistem perpipaan ini kebanyakan jenis bahan pipa yang digunakan adalah jenis logam, disamping keunggulannya logam juga mempunyai banyak kelemahan jika dibandingkan dengan unsur-unsur lain dikarenakan sifat logam yang mudah terkorosi. Pipa logam yang ditanam di bawah permukaan tanah beresiko tinggi terjadi korosi, tanah merupakan penyebab utama pipa logam tersebut mengalami korosi dan dapat menyebabkan bencana alam seperti longsor. Korosivitas tanah menjadi permasalahan penting di PT IPMOMI karena banyaknya pipa logam yang ditanam di bawah tanah yang berfungsi untuk mengalirkan fluida berupa air laut sebagai bahan utama produksi uap. PT IPMOMI ingin menanam pipa logam di bawah tanah, sehingga membutuhkan informasi mengenai keadaan korosivitas tanah pada lokasi tersebut, agar dapat mengatasinya dengan memasang proteksi katodik pada bagian tanah dengan korosivitas tinggi. Salah satu upaya awal untuk menangani hal tersebut adalah dengan melakukan pemetaan zona korosi bawah permukaan. Sebelum melakukan pemetaan zona korosi, maka perlu mengetahui tingkat korosivitas tanah apakah sangat tinggi, tinggi, sedang, ringan, atau sangat ringan, maka dari itu diperlukan metode klasifikasi. Pada penelitian ini, metode klasifikasi yang digunakan adalah metode klasifikasi nonparametrik, yaitu Naive Bayes Classifier (NBC).

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Iis Dewi Ratih

Department of Business Statistics,

Institut Teknologi Sepuluh Nopember,

Surabaya, Indonesia.

Email: iisdewiratih@gmail.com

1. PENDAHULUAN

Listrik menjadi sumber energi yang penting pada era perkembangan teknologi, di mana listrik dapat dihasilkan melalui tenaga air, uap, dan lain-lain. Salah satu perusahaan besar yang mengalirkan listrik untuk Jawa dan Bali adalah PT IPMOMI yang merupakan perusahaan Pembangkit Listrik Tenaga Uap (PLTU) yang berlokasi di Paiton, Probolinggo, Jawa Timur. PT IPMOMI memiliki pipa logam di bawah tanah, fungsinya untuk mengalirkan bahan utama produksi uap berupa fluida air laut. Sistem perpipaan ini dianggap memiliki tingkat integritas yang tinggi serta lebih efektif dan efisien dibandingkan sistem transportasi lain. PT IPMOMI memiliki sistem perpipaan dengan menggunakan jenis bahan pipa logam, di mana logam memiliki sifat yang mudah terkorosi. Kasus korosi ini terjadi pada PT IPMOMI yang mengakibatkan kerusakan dan penurunan mutu pada pipa logam yang disebabkan oleh pH lingkungan dan kelembaban tanah. Oleh karena itu, PT IPMOMI membutuhkan informasi mengenai keadaan korosivitas tanah pada lokasi yang memiliki pipa logam di bawah tanah untuk dilakukan pencegahan dengan memasang proteksi katodik pada bagian tanah dengan korosivitas tinggi. Salah satu upaya awal dari pencegahan tersebut yaitu dengan melakukan pemetaan zona korosivitas tanah. Sebelum melakukan pemetaan zona korosivitas tanah, maka perlu mengetahui tingkat korosivitas tanah tersebut memiliki tingkat korosivitas sangat tinggi, tinggi, sedang, ringan, atau sangat ringan. Sehingga untuk mengetahui tingkat korosivitas tanah tersebut diperlukan metode klasifikasi. Klasifikasi merupakan salah satu metode statistika untuk mengelompokkan suatu data yang memiliki

atribut dan disusun secara sistematis. Metode klasifikasi juga dikembangkan pada data mining yang menggunakan metode klasifikasi nonparametrik, diantaranya Naive Bayes Classifier, K Nearest Neighbours, Decision Tree, dan lain-lain. Penelitian ini merupakan penelitian lanjutan dari penelitian sebelumnya, perlu dilakukan penelitian lanjutan karena sifat korosif tanah dapat menyebabkan perusakan pada pipa logam hingga bencana alam seperti longsor. Berdasarkan uraian diatas, maka pada penelitian ini akan dilakukan pengklasifikasian kualitas tanah pada lokasi dengan pemasangan pipa logam berdasarkan tingkat korosivitas tanah menggunakan algoritma Naive Bayes Classifier (NBC) untuk mendapatkan hasil klasifikasi dengan tingkat akurasi yang lebih tinggi dari penelitian sebelumnya. Sehingga dapat digunakan usulan kepada PT IPMOMI sebagai bahan dasar dalam pemasangan proteksi katodik pada lokasi tanah dengan tingkat korosivitas tinggi dan sangat tinggi.

Algoritma Naive Bayes Classifier (NBC) merupakan sebuah pengklasifikasian probalilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas [1]. Definisi lain mengatakan NBC merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman pada masa sebelumnya [2].

Algoritma Klasifikasi NBC menggunakan data dengan target (*class/label*) yang berupa nilai kategorikal/nominal yang menerapkan teori Bayes dalam klasifikasi. NBC merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Teori keputusan Bayes merupakan pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*), teori Bayesian pada dasarnya adalah kemungkinan kejadian di masa depan yang bisa dihitung dengan menentukan frekuensi pengalaman sebelumnya. Penggunaan algoritma Bayes dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya. Metode Naive Bayes sering disebut dengan algoritma Hypothesis Maximum Apriori Probability (HMAP) yang merupakan penyederhanaan dari metode bayes. Metode ini menyatakan hipotesa dari penghitungan menggunakan probabilitas berdasarkan kondisi prior [13].

Persamaan dari teorema Bayes dapat ditulis sebagai berikut:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (2.1)$$

dengan:

x : Data dengan *class* yang belum diketahui

c : Hipotesis data merupakan suatu *class* spesifik

$P(c|x)$: Probabilitas hipotesis c berdasar kondisi x (*posterios probabilitas*)

$P(c)$: Probabilitas hipotesis c (*prior probabilitas*)

$P(x|c)$: Probabilitas hipotesis x berdasar kondisi pada hipotesis c

$P(x)$: Probabilitas hipotesis x

Proses klasifikasi metode NBC memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Oleh karena itu, metode NBC di atas disesuaikan sebagai berikut :

$$P(c_i | x_1, \dots, x_p) = \frac{P(x_1, \dots, x_p | c_i)P(c_i)}{P(x_1, \dots, x_p | c_1)P(c_1) + \dots + P(x_1, \dots, x_p | c_m)P(c_m)} \quad (2.2)$$

di mana variabel C mempresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $(C|x_1 \dots x_n)$ menggunakan aturan sebagai berikut:

$$P(C|x_1 \dots x_n) = P(C)P(x_1, \dots, x_n|C)$$

$$= P(C)P(x_1|C)P(x_2, \dots, x_n|C, x_1)$$

$$= P(C)P(x_1|C)P(x_2, \dots, x_n|C, x_1)$$

$$= P(C)P(x_1|C)P(x_2|C, x_1)P(x_3|C, x_1, x_2)P(x_4, \dots, x_n|C, x_1, x_2, x_3)$$

$$= P(C)P(x_1|C)P(x_2|C, x_1)P(x_3|C, x_1, x_2)P(x_n, \dots, x_{n-1}|C, x_1, x_2, \dots, x_{n-1}) \quad (3.7)$$

(2.3)

Hasil penjabaran pada persamaan 2.3 menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya,



perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (naïve), bahwa masing-masing petunjuk x_1, x_2, \dots, x_n saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i), \text{ untuk } i \neq j, \quad (2.4)$$

sehingga,

$$P(F_i|C, F_j) = P(F_i|C)$$

Persamaan di atas merupakan model dari teorema NBC yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinu digunakan rumus *Densitas Gauss*.

$$P(X_i = x_i | Y = y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left\{-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right\} \quad (2.5)$$

- P : Peluang
- X_i : Atribut ke $-i$
- x_i : nilai atribut ke $-i$
- Y : Kelas yang dicari
- y : Sub kelas Y yang dicari
- μ_y : mean, menyatakan rata-rata dari seluruh atribut
- σ_y : Standar deviasi, menyatakan varian dari seluruh atribut

Data aktual dan data hasil prediksi dari model klasifikasi disajikan dengan menggunakan tabulasi silang (Confusion Matrix), yang mengandung informasi tentang kelas data aktual direpresentasikan pada baris matriks dan kelas data hasil prediksi pada kolom [3]. Confusion Matrix dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix*

<i>Confusion Matrix</i>		Prediksi	
		Positif	Negatif
Aktual	Positif	TP	FN
	Negatif	FP	TN

Keterangan:

- True Positive* (TP) : *Outcome* positif yang benar diklasifikasikan sebagai kelas positif
- True Negative* (TN) : *Outcome* negatif yang benar diklasifikasikan sebagai kelas negatif
- False Positive* (FP) : *Outcome* negatif yang salah diklasifikasikan sebagai kelas positif
- False Negative* (FN) : *Outcome* positif yang salah diklasifikasikan sebagai kelas negative

Ketepatan klasifikasi dapat dilihat dari akurasi klasifikasi. Akurasi klasifikasi menunjukkan performansi model klasifikasi secara keseluruhan, di mana semakin tinggi akurasi klasifikasi hal ini berarti semakin baik performansi model klasifikasi.

Accuracy Rate (Ketepatan klasifikasi) menunjukkan performansi model klasifikasi secara keseluruhan, di mana semakin tinggi akurasi klasifikasi hal ini berarti semakin baik performansi model klasifikasi.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.6)$$

Error rate dikenal sebagai tingkat kesalahan klasifikasi, dan merupakan fraksi total dari hasil yang diprediksi secara tidak akurat dengan total hasil. *Error Rate* dihitung dengan rumus:

$$Error Rate = 1 - Accuracy Rate \quad (2.7)$$

Importance variable (IV) merupakan signifikansi statistik dari masing-masing variable prediktor dalam data sehubungan dengan pengaruhnya terhadap model yang dihasilkan. IV sebenarnya merupakan setiap peringkat variabel prediktor berdasarkan kontribusi dari masing-masing variabel prediktor terhadap model yang dibuat. Teknik ini membantu pada *data scientist* untuk menyeleksi variabel prediktor tertentu yang tidak memiliki kontribusi yang signifikan terhadap model untuk menghemat waktu dalam proses data dan membuat model.

Korosivitas tanah merupakan masalah multi disiplin yang memiliki dampak terhadap keselamatan lingkungan dan *structural*, korosivitas tanah juga dapat diartikan sebagai kemampuan tanah menyebabkan korosi. Tingkat korosivitas tanah dapat dikelompokkan berdasarkan resistivitas, resistivitas tanah adalah besarnya karakteristik tanah sebagai media elektrolit untuk menghantarkan arus listrik yang menyebabkan terjadinya korosi. Setiap tanah memiliki

tingkat korosivitas yang berbeda, tanah berpasir memiliki korosivitas rendah akibat nilai resistivitasnya yang tinggi sedangkan *clay* memiliki korosivitas yang tinggi akibat nilai resistivitasnya rendah berikut ini adalah tingkat korosivitas tanah berdasarkan nilai resistivitas [4]. Resistivitas tanah diduga dipengaruhi oleh nilai *chargeability* dan tingkat kedalaman tanah. *Chargeability* merupakan pengukuran yang sering dipakai dalam pengukuran induksi polarisasi dengan metode *time domain*, *chargeability* didapatkan dalam satuan milisekon. Nilai *chargeability* lebih teliti dalam menunjukkan adanya mineral-mineral dalam batuan yang berpengaruh pada korosivitas tanah[4]. Sedangkan tingkat kedalaman tanah merupakan ukuran tanah dengan tingkat kedalaman yang diukur dari permukaan tanah dalam satuan meter. Penelitian yang menggunakan tingkat kedalaman tanah yaitu pada metode geolistrik pertama kali dilakukan oleh Conrad Schulumberger pada tahun 1912. Semakin tinggi tingkat kedalaman tanah akan mempengaruhi tingkat korosivitas tanah karena semakin banyak material seperti batuan dan mineral yang bisa mempengaruhi korosivitas.

2. METODE PENELITIAN

Data yang digunakan merupakan data sekunder yaitu data yang diperoleh dari Survey yang dilakukan oleh [4], mahasiswa Departemen Teknik Geofisika Institut Teknologi Sepuluh Nopember tahun 2017 di PT. IPMOMI, Probolinggo. Data penelitian merupakan hasil pengukuran di unit 7 dan 8 PT. IPMOMI pada lokasi 1 dengan panjang lokasi yang bervariasi, kemudian diperoleh nilai *chargeability*, tingkat kedalaman, dan resistivitas.

Variabel penelitian yang digunakan meliputi variabel respon (Y) yaitu tingkat korosivitas tanah, kemudian variabel prediktor (X) yaitu *chargeability* dan tingkat kedalaman berdasarkan hasil pengukuran pada unit 7 dan 8 di PT. IPMOMI, dideskripsikan pada Tabel 2 berikut.

Variabel	Keterangan	Skala	Satuan
Y	Tingkat Korosivitas Tanah	Ordinal	
X ₁	<i>Chargeability</i>	Rasio	Msec
X ₂	Tingkat kedalaman	Nominal	Lokasi

Berikut merupakan penjelasan untuk masing-masing variabel penelitian.

1. Korosivitas tanah merupakan kemampuan tanah menyebabkan korosi. Tingkat korosivitas tanah dapat dikelompokkan berdasarkan resistivitas, resistivitas tanah adalah besarnya karakteristik tanah sebagai media elektrolit untuk menghantarkan arus listrik yang menyebabkan terjadinya korosi. [3]. Tingkat korosivitas berdasarkan nilai resistivitas dideskripsikan pada Tabel 3 berikut.

Resistivitas Tanah (ohm)	Korosivitas
<7	Sangat tinggi
7-20	Tinggi
20-50	Sedang
50-100	Ringan
>100	Sangat Ringan

2. *Chargeability* merupakan pengukuran yang sering dipakai dalam pengukuran induksi polarisasi dengan metode *time domain*, *chargeability* didapatkan dalam satuan milisekon. Nilai *chargeability* lebih teliti dalam menunjukkan adanya mineral-mineral dalam batuan yang berpengaruh pada korosivitas tanah[3].
3. Tingkat kedalaman tanah merupakan ukuran tanah dengan tingkat kedalaman yang diukur dari permukaan tanah dalam satuan meter.

Penelitian ini dilakukan pengukuran pada 1 lokasi di unit 7 dan 8 PT. IPMOMI yang memiliki 6 kategori tingkat kedalaman tanah yaitu 0,17 ; 0,53 ; 0,92 ; 1,36 ; 1,84 ; 2,36 meter.

3. HASIL DAN PEMBAHASAN

Hasil analisis dan pembahasan Klasifikasi Kualitas Tanah Menggunakan Metode Naive Bayes Classifier dijelaskan pada sub bab berikut.

3.1. Klasifikasi Data

Metode analisis yang digunakan dalam penelitian adalah dengan menggunakan *Naive Bayes Classifier* (NBC). Pada penggunaan metode NBC akan dilakukan pemisahan antara training dataset dengan *testing dataset*. Training dataset diambil 75% keseluruhan dari total 46 data sampel. Sedangkan untuk *testing dataset* merupakan sisa dari data yang akan diuji akurasi seperti pada Tabel 4 seperti yang telah dijelaskan pada metode penelitian dalam bab sebelumnya, NBC dilakukan dengan menentukan probabilitas hasil prediksi keanggotaan suatu kelas.

- a) Klasifikasi Data Lokasi 1



Training dataset diambil 75% keseluruhan dari total 46 data sampel. Sedangkan untuk *testing dataset* merupakan sisa dari data yang akan diuji akurasinya seperti pada Tabel 4 berikut.

Tabel 4. Classification Dataset Lokasi 1 (NBC)

Jumlah Data	Training Set	Testing Set
46	34	12

b) Klasifikasi Data Lokasi 2

Training dataset diambil 75% keseluruhan dari total 46 data sampel. Sedangkan untuk *testing dataset* merupakan sisa dari data yang akan diuji akurasinya seperti pada Tabel 5 berikut.

Tabel 5. Classification Dataset Lokasi 2 (NBC)

Jumlah Data	Training Set	Testing Set
46	34	12

c) Klasifikasi Data Lokasi 3

Training dataset diambil 75% keseluruhan dari total 46 data sampel. Sedangkan untuk *testing dataset* merupakan sisa dari data yang akan diuji akurasinya seperti pada Tabel 6 berikut.

Tabel 6. Classification Dataset Lokasi 3 (NBC)

Jumlah Data	Training Set	Testing Set
45	33	12

d) Klasifikasi Data Lokasi 4

Training dataset diambil 75% keseluruhan dari total 46 data sampel. Sedangkan untuk *testing dataset* merupakan sisa dari data yang akan diuji akurasinya seperti pada Tabel 7 berikut.

Tabel 7. Classification Dataset Lokasi 4 (NBC)

Jumlah Data	Training Set	Testing Set
49	36	13

e) Klasifikasi Data Lokasi 5

Training dataset diambil 75% keseluruhan dari total 46 data sampel. Sedangkan untuk *testing dataset* merupakan sisa dari data yang akan diuji akurasinya seperti pada Tabel 8 berikut.

Tabel 8. Classification Dataset Lokasi 5 (NBC)

Jumlah Data	Training Set	Testing Set
45	33	12

3.2. Confusion Matrix

Data aktual dan data hasil prediksi dari model klasifikasi disajikan dengan menggunakan tabulasi silang (*Confusion Matrix*), yang mengandung informasi tentang kelas data aktual direpresentasikan pada baris matriks dan kelas data hasil prediksi pada kolom.

a) *Confusion Matrix* Lokasi 1

Confusion matrix antara prediksi dengan aktual dari *testing dataset* untuk melihat kesesuaian tingkat korosivitas tanah dijelaskan pada Tabel 9 sebagai berikut.

Tabel 9. *Confusion Matriks* NBC Lokasi 1

Prediksi	Aktual				
	Sangat Tinggi	Tinggi	Sedang	Ringan	Sangat Ringan
Sangat Tinggi	1	0	0	0	0
Tinggi	0	2	2	0	0
Sedang	0	1	6	0	0
Ringan	0	0	0	0	0
Sangat Ringan	0	0	0	0	0

Tabel 9 menunjukkan bahwa terdapat empat kategori kualitas tanah di PT. IPMOMI ditinjau dari tingkat korosivitas tanahnya yaitu tingkat korosivitas tanah sangat tinggi, tinggi, sedang ringan, dan sangat ringan. Pada *Confusion Matriks* diatas terdapat total 12 prediksi yang diperoleh dengan menganalisis *testing dataset* untuk melihat kesesuaian tingkat korosivitas tanah di PT. IPMOMI. Dari 12 sampel diatas, klasifier memprediksi tingkat korosivitas tanah “Sangat Tinggi” adalah sebanyak 1 kali, kemudian prediksi tingkat korosivitas tanah “Tinggi” adalah sebanyak 2 kali dan prediksi tingkat korosivitas tanah “Sedang” adalah sebanyak 6 kali, sementara itu klasifier tidak memprediksi tingkat korosivitas tanah “Ringan” dan “Sangat Ringan”. Sedangkan pada data aktual terdapat lokasi dengan tingkat korosivitas kategori “Sangat Tinggi” sebanyak 1, tingkat korosivitas kategori “Tinggi” sebanyak 3, dan lokasi dengan tingkat korosivitas kategori “Sedang” sebanyak 8, serta tidak ada lokasi dengan tingkat korosivitas kategori “Ringan” dan “Sangat Ringan”.

Kinerja dari penggunaan metode NBC berdasarkan tabel *confusion matrix* dapat diukur dengan menghitung nilai *accuracy* sebagai berikut.

$$Accuracy = \frac{TP}{\text{Total Data Set}} = \frac{9}{12} = 0,75$$

Hasil analisis *Naive Bayes Classifier* (NBC) menunjukkan bahwa NBC akan mengklasifikasikan secara benar 9 sampel dari total 12 sampel *testing dataset* yang dapat dilihat dari tabel *confusion matrix* seperti pada Tabel 2.20 dengan nilai akurasi sebesar 0,75 atau 75%.

b) *Confusion Matrix* Lokasi 2

Confusion matrix antara prediksi dengan aktual dari *testing dataset* untuk melihat kesesuaian tingkat korosivitas tanah dijelaskan pada Tabel 10 sebagai berikut.

Tabel 10. *Confusion Matrix*s NBC Lokasi 2

Prediksi	Aktual				
	Sangat Tinggi	Tinggi	Sedang	Ringan	Sangat Ringan
Sangat Tinggi	4	0	0	0	0
Tinggi	0	3	0	0	0
Sedang	0	0	5	0	0
Ringan	0	0	0	0	0
Sangat Ringan	0	0	0	0	0

Tabel 10 menunjukkan bahwa terdapat empat kategori kualitas tanah di PT. IPMOMI ditinjau dari tingkat korosivitas tanahnya yaitu tingkat korosivitas tanah sangat tinggi, tinggi, sedang ringan, dan sangat ringan. Pada *Confusion Matrix*s diatas terdapat total 12 prediksi yang diperoleh dengan menganalisis *testing dataset* untuk melihat kesesuaian tingkat korosivitas tanah di PT. IPMOMI. Dari 12 sampel diatas, klasifier memprediksi tingkat korosivitas tanah “Sangat Tinggi” adalah sebanyak 4 kali, kemudian prediksi tingkat korosivitas tanah “Tinggi” adalah sebanyak 3 kali dan prediksi tingkat korosivitas tanah “Sedang” adalah sebanyak 5 kali, sementara itu klasifier tidak memprediksi tingkat korosivitas tanah “Ringan” dan “Sangat Ringan”.

Kinerja dari penggunaan metode NBC berdasarkan tabel *confusion matrix* dapat diukur dengan menghitung nilai *accuracy* sebagai berikut.

$$Accuracy = \frac{TP}{\text{Total Data Set}} = \frac{12}{12} = 1,00$$

Hasil analisis *Naive Bayes Classifier* (NBC) menunjukkan bahwa NBC akan mengklasifikasikan secara benar dari total 12 sampel *testing dataset* yang dapat dilihat dari tabel *confusion matrix* seperti pada Tabel 2.21 dengan nilai akurasi sebesar 1,00 atau 100%.

c) *Confusion Matrix* Lokasi 3

Confusion matrix antara prediksi dengan aktual dari *testing dataset* untuk melihat kesesuaian tingkat korosivitas tanah dijelaskan pada Tabel 11 sebagai berikut.

Tabel 11. *Confusion Matrix*s NBC Lokasi 3

Prediksi	Aktual				
	Sangat Tinggi	Tinggi	Sedang	Ringan	Sangat Ringan
Sangat Tinggi	3	0	1	0	0
Tinggi	0	1	2	0	0
Sedang	1	0	6	0	0
Ringan	0	0	0	0	0
Sangat Ringan	0	0	0	0	0

Tabel 11 menunjukkan bahwa terdapat empat kategori kualitas tanah di PT. IPMOMI ditinjau dari tingkat korosivitas tanahnya yaitu tingkat korosivitas tanah sangat tinggi, tinggi, sedang ringan, dan sangat ringan. Pada *Confusion Matrix*s diatas terdapat total 12 prediksi yang diperoleh dengan menganalisis *testing dataset* untuk melihat kesesuaian tingkat korosivitas tanah di PT. IPMOMI. Dari 12 sampel diatas, klasifier memprediksi tingkat korosivitas tanah “Sangat Tinggi” adalah sebanyak 3 kali, kemudian prediksi tingkat korosivitas tanah “Tinggi” adalah sebanyak 1 kali dan prediksi tingkat korosivitas tanah “Sedang” adalah sebanyak 6 kali, sementara itu klasifier tidak memprediksi tingkat korosivitas tanah “Ringan” dan “Sangat Ringan”. Sedangkan pada data aktual terdapat lokasi dengan tingkat korosivitas kategori “Sangat Tinggi” sebanyak 4, tingkat korosivitas kategori “Tinggi” sebanyak 1, dan lokasi dengan tingkat korosivitas kategori “Sedang” sebanyak 9, serta tidak ada lokasi dengan tingkat korosivitas kategori “Ringan” dan “Sangat Ringan”.

Kinerja dari penggunaan metode NBC berdasarkan tabel *confusion matrix* dapat diukur dengan menghitung nilai *accuracy* sebagai berikut.

$$Accuracy = \frac{TP}{\text{Total Data Set}} = \frac{10}{12} = 0,833$$



Hasil analisis *Naive Bayes Classifier* (NBC) menunjukkan bahwa NBC akan mengklasifikasikan secara benar 10 sampel dari total 12 sampel *testing dataset* yang dapat dilihat dari tabel *confusion matrix* seperti pada Tabel 2.22 dengan nilai akurasinya sebesar 0,83 atau 83,3%.

d) *Confusion Matrix* Lokasi 4

Confusion matrix antara prediksi dengan aktual dari *testing dataset* untuk melihat kesesuaian tingkat korosivitas tanah dijelaskan sebagai berikut.

Kinerja dari penggunaan metode NBC berdasarkan tabel *confusion matrix* dapat diukur dengan menghitung nilai *accuracy* sebagai berikut.

$$Accuracy = \frac{TP}{\text{Total Data Set}} = \frac{12}{13} = 0,923$$

Hasil analisis *Naive Bayes Classifier* (NBC) menunjukkan bahwa NBC akan mengklasifikasikan secara benar 11 sampel dari total 12 sampel *testing dataset* yang dapat dilihat dari tabel *confusion matrix* seperti pada Tabel 2.25 dengan nilai akurasinya sebesar 0,923 atau 92,3%.

e) *Confusion Matrix* Lokasi 5

Confusion matrix antara prediksi dengan aktual dari *testing dataset* untuk melihat kesesuaian tingkat korosivitas tanah dijelaskan pada Tabel 12 sebagai berikut.

Tabel 12. *Confusion Matriks* NBC Lokasi 5

Prediksi	Aktual				
	Sangat Tinggi	Tinggi	Sedang	Ringan	Sangat Ringan
Sangat Tinggi	3	0	0	0	0
Tinggi	0	2	1	0	0
Sedang	0	0	6	0	0
Ringan	0	0	0	0	0
Sangat Ringan	0	0	0	0	0

Tabel 12 menunjukkan bahwa terdapat empat kategori kualitas tanah di PT. IPMOMI ditinjau dari tingkat korosivitas tanahnya yaitu tingkat korosivitas tanah sangat tinggi, tinggi, sedang ringan, dan sangat ringan. Pada *Confusion Matriks* diatas terdapat total 12 prediksi yang diperoleh dengan menganalisis *testing dataset* untuk melihat kesesuaian tingkat korosivitas tanah di PT. IPMOMI. Dari 12 sampel diatas, klasifier memprediksi tingkat korosivitas tanah “Sangat Tinggi” adalah sebanyak 3 kali, kemudian prediksi tingkat korosivitas tanah “Tinggi” adalah sebanyak 2 kali dan prediksi tingkat korosivitas tanah “Sedang” adalah sebanyak 6 kali, sementara itu klasifier tidak memprediksi tingkat korosivitas tanah “Ringan” dan “Sangat Ringan”. Sedangkan pada data aktual terdapat lokasi dengan tingkat korosivitas kategori “Sangat Tinggi” sebanyak 3, tingkat korosivitas kategori “Tinggi” sebanyak 2, dan lokasi dengan tingkat korosivitas kategori “Sedang” sebanyak 7, serta tidak ada lokasi dengan tingkat korosivitas kategori “Ringan” dan “Sangat Ringan”.

Kinerja dari penggunaan metode NBC berdasarkan tabel *confusion matrix* dapat diukur dengan menghitung nilai *accuracy* sebagai berikut.

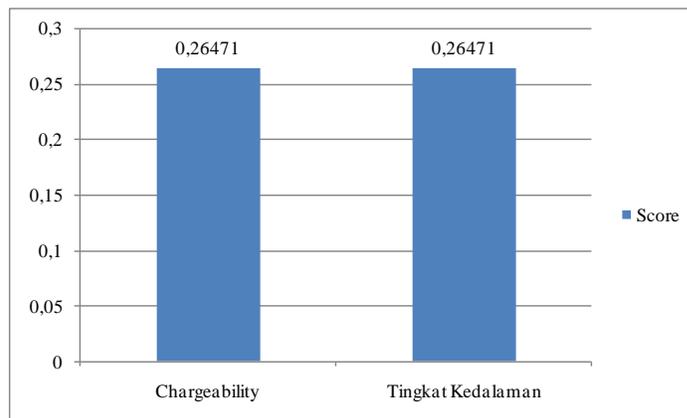
$$Accuracy = \frac{TP}{\text{Total Data Set}} = \frac{11}{12} = 0,917$$

Hasil analisis *Naive Bayes Classifier* (NBC) menunjukkan bahwa NBC akan mengklasifikasikan secara benar 11 sampel dari total 12 sampel *testing dataset* yang dapat dilihat dari tabel *confusion matrix* seperti pada Tabel 2.25 dengan nilai akurasinya sebesar 0,917 atau 91,7%.

3.3. Importance Variable

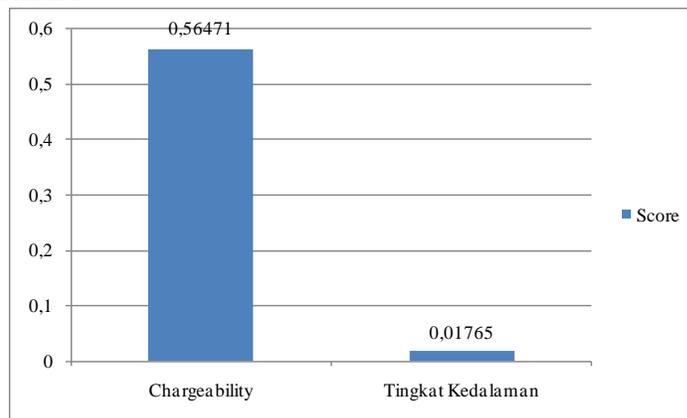
Importance variable dari digunakan untuk menyeleksi variabel prediktor tertentu yang tidak memiliki kontribusi yang signifikan terhadap model untuk menghemat waktu dalam proses data dan membuat model yang dijelaskan sebagai berikut.

a) *Importance variable* Lokasi 1

Gambar 1. Diagram *Importance Variable* Lokasi 1

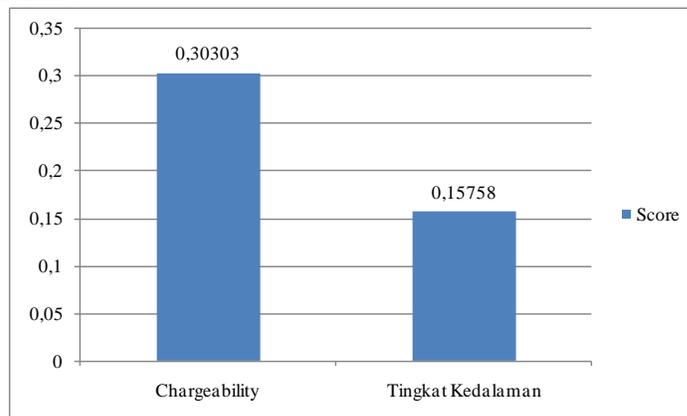
Gambar 1 menunjukkan bahwa *Chargeability* dan *Tingkat Kedalaman* memiliki kontribusi yang sama terhadap model.

b) *Importance variable* Lokasi 2

Gambar 2. Diagram *Importance Variable* Lokasi 2

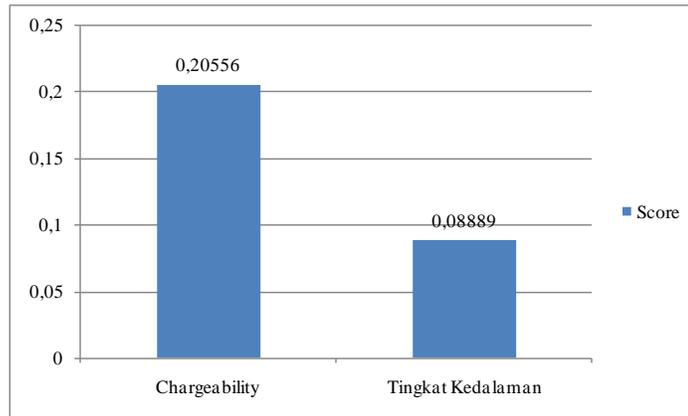
Gambar 2 menunjukkan bahwa *Chargeability* memiliki kontribusi yang jauh lebih besar terhadap model.

c) *Importance variable* Lokasi 3

Gambar 3. Diagram *Importance Variable* Lokasi 3

Gambar 3 menunjukkan bahwa *Chargeability* memiliki kontribusi yang lebih besar terhadap model.

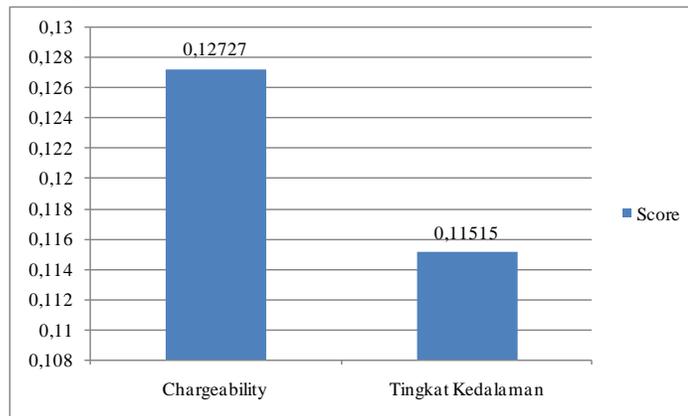
d) *Importance variable* Lokasi 4



Gambar 4. Diagram *Importance Variable* Lokasi 4

Gambar 4 menunjukkan bahwa *Chargeability* memiliki kontribusi yang lebih besar terhadap model.

e) *Importance variable* Lokasi 5



Gambar 5. Diagram *Importance Variable* Lokasi 5

Gambar 5 menunjukkan bahwa *Chargeability* memiliki kontribusi yang lebih besar terhadap model

4. CONCLUSION (10 PT)

Kesimpulan berdasarkan hasil analisis menunjukkan bahwa ketepatan klasifikasi korosivitas tanah di salah satu lokasi PT. IPMOMI menggunakan metode Naive Bayes Classifier (NBC) adalah sebesar 83,3%. Variabel pembentuk model yaitu *Chargeability* memiliki kontribusi yang lebih tinggi dibandingkan tingkat kedalaman dalam pembentuk model klasifikasi.

ACKNOWLEDGEMENTS (10 PT)

Terima kasih kepada Departemen Statistika Bisnis, Institut Teknologi Sepuluh Nopember yang telah memberinial dukungan materiil, tidak lupa kepada Bu Iis selaku dosen pembimbing sehingga jurnal ini dapat terealisasi.

DAFTAR PUSTAKA

- [1] Patil, T. R., Shrekar, M. S. (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification, *International Journal Of Computer Science And Applications* Vol. 6, No.2, Apr 2013
- [2] Bustami (2013), Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi, *Jurnal Informatika*, ISSN : 1978-0524
- [3] Cramer S.D. and Jr. Covino B.S. (2003). *ASM Handbook Volume 13A Corrosion: Fundamentals, Testing, and Protection*, ASM International.
- [4] Dewi, Kiki Kartika (2017). *Pemetaan Zona Korosivitas Tanah Berdasarkan Nilai Chargeability Menggunakan Metode Time Domain Induced Polarization Konfigurasi Dipole-Dipole*. Studi Kasus PT. IPMOMI.

HALAMAN INI SENGAJA DIKOSONGKAN