

PENERAPAN METODE DECISION TREE DAN NAÏVE BAYES UNTUK MENGHITUNG KADAR KARAT EMAS

Bobby Suryo Prakoso¹, Gustap Dimas Sutanto²

^{1,2} STMIK Nusa Mandiri Jakarta
Jl. Kramat Raya No.18, RT.2/RW.7, Jakarta Pusat
E-mail : suryobobby@gmail.com¹, gustap.gds@bsi.ac.id²

ABSTRAK

Sistem analisa perhitungan kadar karat emas oleh toko emas umumnya dilakukan secara manual dengan menggunakan perhitungan dari kalkulator dan tabel. Sehingga diperlukan waktu tambahan untuk menghitung kadar emas tersebut serta hasil yang kurang akurat. Untuk itu digunakan metode *Decision Tree* dan *Naïve Bayes* agar mampu membantu menangani permasalahan tersebut baik dalam hal analisa kadar emas maupun proses perekaman (*record*) data, serta mampu mengoptimalkan perhitungan kadar emas tersebut. Metode *Decision Tree* merupakan sebuah struktur *flowchart* yang mirip seperti struktur pohon, setiap titik pohon merupakan atribut yang telah diuji, setiap cabang merupakan hasil uji dan titik akhir merupakan pembagian *Class* yang dihasilkan oleh algoritma C4.5 dan *Decision Tree*. Serta merupakan dua model yang saling berhubungan, karena untuk membangun sebuah *Decision Tree* dibutuhkan algoritma C4.5. Dan algoritma *Naïve Bayes* adalah teknik prediksi berbasis probabilistik sederhana yang berdasarkan pada penerapan teorema atau aturan bayes dengan asumsi independensi yang kuat pada fitur, artinya adalah sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. Kriteria – kriteria yang digunakan dalam perhitungan kadar emas ini adalah kandungan emas asli, kandungan emas campuran, persentase penandaan, berat emas, dan berat volume.

Kata kunci : Kadar Karat, Emas, Decision Tree, Naïve Bayes

ABSTRACT

The system analysis of calculation of levels of gold in gold store is done manually by using the calculation of the calculator and tables. So the extra time required to calculate the gold levels and results less accurate. For it was used the method of Decision Tree and Naïve Bayes to help deal with these problems both in terms of the analysis of the levels of gold as well as the process of recording (record) data, as well as being able to optimize the calculation of the rate of the gold. The method of Decision Tree is a flowchart that resembles the structure of the tree structure, every point of the tree is an attribute that has been tested, each branch is a test result and the end point is the Division of the Class generated by the algorithm C 4.5 and Decision Tree. And are the two models that are interconnected, because to build a Decision Tree algorithms become C 4.5. And the Naïve Bayes algorithm is a simple probabilistic-based prediction techniques based on the application of the theorem or bayes rules with a strong independence assumption on the features, it is a feature on a data do not exist with regard to or whether other features within the same data. Criteria – the criteria used in the calculation of the level of this gold is the gold content of the original, the gold content of the mixture, the percentage of heavy gold, tagging, and heavy volume.

Keyword : Levels Of Rust, Gold, Decision Tree, Naïve Bayes

1. PENDAHULUAN

Toko emas adalah tempat paling umum seseorang untuk membeli emas, walaupun saat ini banyak sarana untuk membeli emas. Toko emas merupakan tempat untuk melakukan kegiatan jual beli yang memiliki karakteristik bisnis yang spesifik, yakni khusus yang berhubungan dengan emas. Kegiatan bisnis yang dilakukan diantaranya adalah pembelian emas dari *supplier* maupun *customer*, penyimpanan emas, baik berupa emas batangan maupun emas berbentuk perhiasan yang biasa dijumpai, penjualan emas, dan perhitungan kadar karat emas.

Sistem perhitungan kadar emas yang dilakukan saat ini oleh pihak toko umumnya masih diterapkan secara manual atau analisa dilakukan masih menggunakan perhitungan rumusan. Akibatnya, waktu yang dibutuhkan untuk menghitung kadar emas akan lebih panjang dan dari segi hasil perhitungan juga kurang akurat. Maka, tentu diperlukan waktu tambahan guna mengoptimalkan data – data status kadar emas tersebut.

Sehingga dengan perhitungan dengan metode *Decision Tree* dan *Naïve Bayes* dirasa sangat mampu membantu menangani permasalahan tersebut baik dari sisi perekaman (*record*) data maupun dalam hal analisa kadar emas dan diharapkan perhitungan yang dilakukan dapat bekerja lebih baik dengan menggunakan analisa perhitungan nilai – nilai kriterianya.

2. METODOLOGI

Berikut ini adalah tahapan – tahapan dari penelitian ini:

- a. Masukan (*Input*)
Data masukan (*input*) adalah berupa data karakteristik emas yang akan dihitung.
- b. Proses (*Process*)
Pada tahap ini, data emas yang akan dihitung dengan menggunakan algoritma *Decision Tree* dan *Naïve Bayes* serta aplikasi *rapidminer*.
- c. Keluaran (*Output*)
Pada tahap keluaran (*output*) adalah data emas yang sudah dihitung kadar karatnya.

3. LANDASAN TEORI

Emas

Emas merupakan barang tambang yang memiliki unsur kimia dalam tabel periodik yang memiliki simbol Au (bahasa Latin: 'aurum') dan nomor atom 79. Sebuah logam transisi (trivalen dan univalen) yang lembek, mengkilap, kuning, berat. Emas melebur dalam bentuk cair pada suhu sekitar 1000 derajat celcius (Chalimah, 2012).

Mengenal emas, kita terlebih dahulu mengenal istilah "kadar" dalam emas. Kadar merupakan tingkat keaslian emas, atau jumlah kandungan kemurnian emas. Kadar emas dinyatakan dalam karat.

Menurut SNI (Standart Nasional Indonesia) - No : SNI 13-3487-2005 standar karat emas adalah sebagai berikut:

- 24 K = 99,00 - 99,99%
- 23 K = 94,80 - 98,89%
- 22 K = 90,60 - 94,79%
- 21 K = 86,50 - 90,59%
- 20 K = 82,30 - 86,49%
- 19 K = 78,20 - 82,29%
- 18 K = 75,40 - 78,19%

Data Mining

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar (Turban, E. Aronson, & Liang, 2007).

Menurut Gartner Group data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose & Larose, 2014).

Decision Tree

Metode *decision tree* adalah sebuah struktur *flowchart* yang mirip seperti struktur pohon, setiap titik pohon merupakan atribut yang telah diuji, setiap cabang merupakan hasil uji, dan titik akhir merupakan pembagian kelas yang dihasilkan (Han & Kamber, 2001).

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti SQL (*Structured Query Language*) untuk mencari record pada kategori tertentu (Kusrini, 2009).

Algoritma C.45

Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5 (Quinlan, 1996).

Di akhir tahun 1970 hingga di awal tahun 1980-an, J. Ross Quinlan seorang peneliti di bidang mesin pembelajaran mengembangkan sebuah model pohon keputusan yang dinamakan ID3 (Iterative Dichotomiser), walaupun sebenarnya proyek ini telah dibuat sebelumnya oleh E.B. Hunt, J. Marin, dan P.T. Stone. Kemudian Quinlan membuat algoritma dari pengembangan ID3 yang dinamakan C4.5 yang berbasis supervised learning.

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5 (Larose, 2008) yaitu :

1. Mempersiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas – kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropy. Untuk menghitung nilai entropy digunakan rumus :

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

Keterangan :

S : himpunan kasus
A : fitur
n : jumlah partisi S
pi : proporsi dari Si terhadap S

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut – atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam persamaan berikut:

$$Gain(S, A) = entropy(s) - \sum_{i=1}^n |S_i|/S \times Entropy(S_i)$$

Keterangan :

S : himpunan kasus
A : atribut
n : jumlah partisi atribut A
|Si| : jumlah kasus pada partisi ke-i
|S| : jumlah kasus dalam S

3. Ulangi langkah ke 2 dan langkah ke 3 hingga semua *record* terpartisi.
4. Proses partisi pohon keputusan akan berhenti saat :
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut didalam *record* yang dipartisi lagi.
 - c. Tidak ada *record* didalam cabang yang kosong.

Naïve Bayes

Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. *Naïve Bayes* didasarkan pada teorema *Bayes* yang memiliki kemampuan klasifikasi serupa dengan *Decision Tree* dan *Neural Network*. *Naïve Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang besar (Prasetyo, 2012).

Prediksi *Bayes* didasarkan pada formula teorema *Bayes* dengan formula umum sebagai berikut :

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)}$$

Keterangan :

X = Data dengan *class* yang belum diketahui
H = Hipotesis data X merupakan suatu

- class* spesifik
- $P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)
- $P(H)$ = Probabilitas hipotesis H (prior prob.)
- $P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut
- $P(X)$ = Probabilitas dari X

Naïve Bayes merupakan salah satu algoritma dalam teknik *data mining* yang menerapkan teori *Bayes* dalam klasifikasi. Teorema keputusan *Bayes* adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*). *Naïve bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai *output*. Dengan kata lain, diberikan nilai *output*, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Santosa, 2007).

Rapidminer

Rapidminer adalah sebuah lingkungan *machine learning data mining*, *text mining* dan *predictive analytics* (Vercellis, 2011).

Rapidminer menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *Rapidminer* memiliki kurang lebih 500 operator data mining, termasuk operator untuk *input*, *output*, data *preprocessing* dan *visualisasi*. *Rapid Miner* merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri. *Rapidminer* ditulis dengan menggunakan bahasa *java* sehingga dapat bekerja di semua sistem operasi.

Beberapa Fitur dari *Rapidminer*, antara lain:

1. Banyaknya *algoritma data mining*, seperti *decision tree* dan *self-organization map*.
2. Bentuk grafis yang canggih, seperti tumpang tindih diagram *histogram*, *tree chart* dan *3D Scatter plots*.
3. Banyaknya variasi *plugin*, seperti *text plugin* untuk melakukan analisis teks.
4. Menyediakan prosedur data mining dan *machine learning* termasuk: ETL

(*extraction, transformation, loading*), data *preprocessing*, *visualisasi*, *modelling* dan evaluasi.

5. Proses data mining tersusun atas operator-operator yang *nestable*, dideskripsikan dengan *XML*, dan dibuat dengan *GUI*.
6. Mengintegrasikan proyek *data mining Weka* dan *statistika R*.

4. HASIL DAN PEMBAHASAN

Data Sampel

Data yang digunakan untuk perhitungan sebanyak 120 data emas.

Tabel 1. Data Sampel

No	Kandungan Emas	Kandungan Campuran	Pendanaan	Berat Jenis	Berat Volume	Kadar
1	85,6	13,3	856	70	3,1	21
2	78,2	21,6	782	56	5,6	18
3	82,0	18,0	820	53	3,1	19
4	82,0	18,0	820	73	4,6	19
5	86,6	13,4	866	60	3,6	20
6	78,0	22,0	780	57	4,0	18
7	99,0	1,0	990	55	2,9	24
8	99,0	1,0	990	55	2,9	24
9	82,2	17,8	822	70	4,6	19
10	82,3	17,7	823	70	4,4	20
11	78,0	22,0	780	55	3,7	18
12	82,3	17,7	823	50	4,4	20
13	78,0	22,0	780	65	3,6	18
14	90,0	22,0	900	50	3,2	18
15	78,0	22,0	780	66	4,2	18
16	78,0	22,0	780	52	4,3	18
17	78,0	22,0	780	69	3,4	18
18	78,0	22,0	780	50	4,5	18
19	85,7	21,6	857	57	4,0	21
20	85,7	22,0	857	70	4,7	21
21	99,0	1,0	990	73	3,8	24
22	78,0	22,0	780	70	3,3	18
23	86,2	13,8	862	73	4,4	20
24	86,2	13,8	862	57	3,4	20
25	86,2	13,8	862	61	3,7	20
26	86,2	13,8	862	67	3,4	20
27	78,2	9,8	782	50	3,9	21
28	78,0	9,8	780	73	4,6	21
29	90,5	22,0	905	70	4,5	18
30	90,5	22,0	905	70	4,5	18
31	78,0	21,6	780	55	4,0	18
32	78,2	21,6	782	56	5,6	18
33	90,3	9,7	903	61	3,5	21
34	90,3	9,7	903	67	3,9	21
35	90,3	9,7	903	53	3,1	21
36	90,3	9,7	903	73	4,2	21
37	94,9	5,1	949	61	3,3	23
38	78,0	9,8	780	53	3,2	23
39	85,7	21,6	857	57	4,0	21
40	94,8	5,2	948	60	3,3	23

41	94,8	5,2	948	65	3,6	23
42	94,8	5,2	948	55	3,0	23
43	90,0	13,3	900	70	2,3	21
44	85,7	22,0	857	70	4,7	21
45	86,7	21,0	867	50	4,2	18
46	90,0	13,3	900	70	2,3	21
47	78,0	9,8	780	73	4,6	21
48	98,9	1,1	989	60	3,1	23
49	98,9	1,1	989	70	3,7	23
50	98,8	1,1	988	65	3,4	23
51	98,9	1,1	989	50	2,6	23
52	86,7	9,8	867	65	4,2	23
53	90,5	22,0	905	70	4,5	18
54	86,7	21,0	867	50	4,2	18
55	78,2	21,6	782	60	3,9	23
56	82,5	17,5	825	73	4,5	19
57	90,5	9,5	905	60	3,5	21
58	98,8	1,2	988	73	2,8	23
59	90,5	9,5	905	65	3,7	21
60	90,5	9,5	905	55	3,2	21
61	90,5	9,5	905	70	4,0	21
62	90,5	9,5	905	50	2,9	21
63	94,9	5,1	949	57	3,1	23
64	94,9	5,1	949	67	3,7	23
65	78,0	21,6	780	55	4,0	18
66	82,5	17,5	825	61	3,8	20
67	82,5	17,5	825	73	4,6	19
68	82,5	17,5	825	57	3,6	20
69	94,9	5,1	949	53	2,9	23
70	90,8	9,8	908	50	2,6	22
71	90,8	9,8	908	55	2,9	22
72	90,8	9,8	908	60	3,1	22
73	90,8	9,8	908	61	3,3	22
74	82,2	17,8	822	65	4,1	19
75	82,2	17,8	822	60	3,8	19
76	82,2	17,8	822	55	3,4	19
77	82,2	17,8	822	50	3,1	19
78	99,0	1,0	990	50	2,6	24
79	99,0	1,0	990	60	3,1	23
80	99,0	1,0	990	65	3,4	24
81	99,0	1,0	990	53	2,8	24
82	90,0	22,0	900	50	3,2	18
83	78,4	21,6	784	57	3,7	21
84	78,4	21,6	784	61	4,0	19
85	78,2	21,6	782	60	3,9	23
86	78,0	9,8	780	53	3,2	23
87	85,6	13,3	856	70	3,1	21
88	90,5	9,8	905	57	3,1	22
89	85,7	21,0	857	57	3,4	22
90	90,5	9,2	905	67	3,7	22
91	90,5	9,2	905	60	3,1	22
92	94,7	5,3	947	55	2,8	22
93	94,7	5,3	947	70	3,3	22
94	94,7	5,3	947	50	3,6	22
95	78,4	21,6	784	60	3,1	19
96	86,4	13,5	864	70	4,2	20
	86,4	13,6	864	65	3,9	20
98	86,4	13,6	864	60	3,6	20
99	86,4	13,6	864	50	3,0	20
100	86,4	13,6	864	55	3,3	20
101	82,3	17,7	823	65	4,1	20
102	78,4	21,6	784	55	2,9	19
103	78,4	21,8	784	60	3,1	19
104	86,7	9,8	867	65	4,2	23
105	82,3	17,7	823	55	3,5	20
106	82,3	17,7	823	60	3,7	20

107	99,0	1,0	990	57	3,0	24
108	98,9	1,1	989	55	2,9	23
109	78,4	21,8	784	61	3,3	19
110	86,7	13,3	867	53	2,9	21
111	86,7	13,3	867	67	3,7	21
112	82,0	18,0	820	61	3,8	19
113	86,7	13,3	867	60	3,1	21
114	82,0	18,0	820	57	4,2	19
115	86,6	13,6	866	55	2,8	21
116	86,6	13,4	866	57	3,1	21
117	82,5	17,5	825	53	3,3	20
118	86,6	13,6	866	67	3,7	21
119	82,5	17,5	825	67	3,3	20
120	82,0	18,0	820	67	4,2	19

Setelah ditentukan data sampel, selanjutnya dibuat *range* tiap – tiap kriteria agar memudahkan saat perhitungan pada *Microsoft excel*. Berikut adalah keterangan *range* data emas :

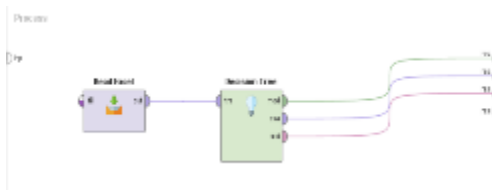
1. Kandungan emas
 - a. $\leq 85,0$
 - b. $85,1 \dots 90,0$
 - c. $> 90,0$
2. Kandungan Campuran
 - a. $\leq 10,0$
 - b. $10,1 \dots 20,0$
 - c. $> 20,0$
3. Penandaan
 - a. ≤ 850
 - b. $851 \dots 900$
 - c. > 900
4. Berat Jenis
 - a. ≤ 60
 - b. $61 \dots 70$
 - c. > 70
5. Berat Volume
 - a. ≤ 3000
 - b. $3001 \dots 4000$
 - c. > 4000
6. Kadar
 - a. Rendah (< 22 Karat)
 - b. Tinggi (≥ 22 Karat)

Perhitungan *Entropy*

Berikut adalah hasil perhitungan *gain* dari data emas yang sudah ditentukan sebelumnya :

- a. Kandungan emas = 0,3128
- b. Kandungan campuran = 0,4600
- c. Penandaan = 0,3128
- d. Berat jenis = 0,0241
- e. Berat volume = 0,1245

Setelah dihitung secara manual menggunakan *Microsoft excel*, selanjutnya dihitung menggunakan *rapidminer* agar didapatkan grafik *Decision Tree*.



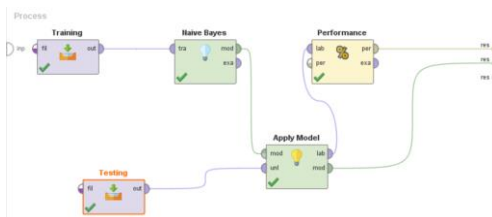
Gambar 1. Penggunaan *Rapidminer* (*Decision Tree*)



Gambar 2. Grafik *Decision Tree*

Perhitungan Naïve Bayes

Setelah dihitung secara manual menggunakan *Microsoft excel*, hasil akurasinya sebesar 88,89% dan selanjutnya dihitung menggunakan *rapidminer* agar didapatkan akurasi dari perhitungan tersebut.



Gambar 3. Penggunaan *Rapidminer* (*Naive bayes*)

accuracy: 100.00%

	true Rendah	true Tinggi	class precision
pred Rendah	0	0	100.00%
pred Tinggi	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 4. Akurasi Perhitungan *Rapidminer* (*Naïve Bayes*)

5. KESIMPULAN

Dari 120 data sampel yang dihitung, dapat diketahui bahwa akurasi perhitungan manual pada algoritma *Naïve Bayes* adalah sebesar 88,89% dan dengan menggunakan *rapidminer* sebesar 100%.

DAFTAR PUSTAKA

- Chalimah, Siti. (2012), Tugas Makalah Kimia Non Pangan Bahan Tambang Emas. *Temanggung*.
- Han, J., & Kamber, M. (2001). *Data mining concepts and techniques* San Francisco Moraga Kaufman.
- Kusrini, E. T. L. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Larose, D. T. (2008). *Data Mining: Methods and Models* by D. T. Larose. *Biometrics*, 64(1), 316–316. https://doi.org/10.1111/j.1541-0420.2008.00962_9.x
- Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons.
- Prasetyo, E. (2012). *Data Mining Konsep dan aplikasi menggunakan matlab*. Yogyakarta: Andi.
- Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*. <https://doi.org/10.1613/jair.279>
- Santosa, B. (2007). *Data mining teknik pemanfaatan data untuk keperluan bisnis*. Yogyakarta: Graha Ilmu, 978(979), 756.
- Turban, E., E. Aronson, J., & Liang, T.-P. (2007). *Decision Support Systems and Business Intelligence*. *Decision Support and Business Intelligence Systems*, 7/E, 1–35. <https://doi.org/10.1017/CBO9781107415324.004>
- Vercellis, C. (2011). *Business intelligence: data mining and optimization for decision making*. John Wiley & Sons.