

---

## Association Rule Mining Pada Data Geokimia Pemboran

Gusti Ngurah Mega Nata  
STMIK STIKOM Bali  
mega.fuzzy@gmail.com

### Abstrak

Mengetahui korelasi atau asosiasi kemunculan mineral pada data pemboran merupakan hal penting untuk mengetahui potensi mineral yang akan di tambang. Metode konvensional untuk pengolahan data sudah tidak sanggup lagi untuk menemukan aturan asosiasi pada data pemboran. Teknik data mining sudah sangat diperlukan untuk mengetahui aturan asosiasi pada data mineral yang disebut juga data geokimia. *Association Rule Mining* (ARM) merupakan *task* dari *data mining* yang berfungsi khusus untuk mencari aturan asosiasi antar *items* dalam *frequent itemsets* pada dataset yang besar. Tetapi, data pemboran belum memiliki *itemsets* yang jelas dan data yang akan menjadi item disimpan dalam bentuk numerik. Nilai item yang bertipe numerik akan membuat jumlah kemunculan item menjadi sangat banyak sehingga dimensi *itemset* menjadi sangat rumit. Fokus permasalahan pada penelitian ini yaitu bagaimana melakukan ARM pada data geokimia pemboran yang bernilai numerik. Untuk implementasi ARM pada data pemboran membutuhkan pemodelan data untuk *items* dan *itemsets*. Penanganan nilai numerik menggunakan cara diskritisasi dengan logika *fuzzy* dan *percentile* untuk membagi nilai numerik menjadi 5 *range*. Proses ARM menggunakan algoritma *FP-Growth*. Hasil pengujian menunjukkan bahwa diskritisasi data geokimia menggunakan *fuzzy logic* dan *percentile* dapat mengurangi dimensi kemunculan item pada ARM.

**Kata Kunci :** data pemboran, data geokimia, *association rule mining*, *FP-Growth*, *Fuzzy Logic*,

### Abstract

*Determine the correlation or association of mineral occurrence in drilling data is essential to determine the mineral potential will be in mine. Conventional methods for processing the data were not able to find the association rules in the drilling data. Data mining techniques is very necessary to know the rules of the association in the data minerals also called geochemical data. Association Rule Mining (ARM) is the task of the special functions of data mining to mine the association rules between items in frequent itemsets in large datasets. However, drilling data has not yet itemsets and the data stored in numeric type. The value of numeric items will make the number of times an item to be very much. thus, the dimensions itemset becomes very complicated. This paper focus on the issue is how do ARM on drilling geochemical data of type numeric. For the implementation of ARM in the drilling data requires data modeling for items and itemsets. Handling numeric values using discretization method with fuzzy logic and percentile. ARM processes using FP-Growth algorithm. The test results show that the discretization geochemical data using fuzzy logic and percentile can reduce the dimensions of the emergence of the items on ARM.*

*Keywords :* drilling data, geochemistry, *association rule mining*, *FP-Growth*, *Fuzzy Logic*,

### 1. Pendahuluan

*Association rule mining* (ARM) merupakan salah satu *task data mining* yang bertujuan mencari aturan asosiasi antar *items* dalam *frequent itemsets* pada *dataset* yang besar [5]. Pada awalnya, metode ini diimplementasikan untuk mencari hubungan asosisasi antar *item* di dalam swalayan dengan cara menganalisis keranjang belanja dari pelanggan, yang dikenal dengan *Market Basket Analysis* [1]. Pada penelitian berikutnya metode ARM digunakan tidak hanya pada data transaksi di swalayan, melainkan sudah digunakan pada data perpustakaan [9], dan data berupa document atau *text mining*. Tetapi, belum ada penelitian khusus yang membahas penggunaan ARM pada data numerik khususnya data geokimia. Data geokimia memiliki karakteristik yang berbeda dengan data keranjang belanja, dokumen atau data

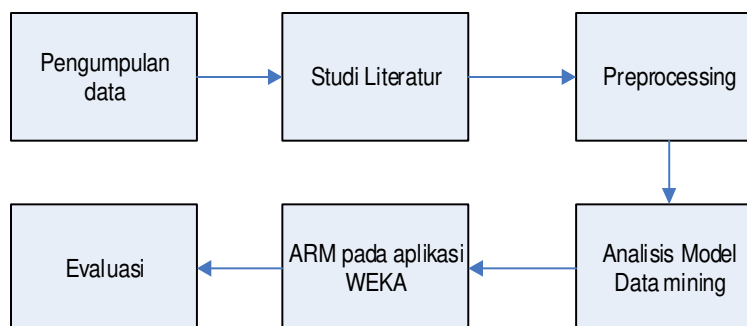
---

transaksi lainnya. Data geokimia merupakan data hasil analisis suatu kadar mineral bahan galian yang bersumber dari alam. *Values* dari kadar mineral pada data geokimia dalam bentuk numerik, selain itu antar mineral memiliki asosiasi dan korelasi dengan mineral pendampingnya [2].

Dalam industry eksplorasi pertambangan mengetahui asosiasi antar mineral logam merupakan hal penting untuk mendukung pengambilan keputusan dan evaluasi potensi mineral bahan galian [3]. ARM mencari aturan asosiasi antar *items* dalam *frequent itemsets* pada dataset yang besar. Tetapi, data pemboran belum memiliki *itemsets* yang jelas. Untuk implementasi ARM pada data pemboran membutuhkan pemodelan data untuk *items* dan *itemsets*. Pada penelitian [10] telah dilakukan representasi data pemboran menjadi *itemsets* namun, data pemboran yang diolah hanya data geologi yang bertipe diskrit. Maka, fokus permasalahan pada penelitian ini yaitu bagaimana melakukan ARM pada data geokimia pemboran yang bernilai numerik. Penanganan nilai numerik menggunakan cara diskritisasi dengan logika *fuzzy* dan *percentile*. Proses ARM menggunakan algoritma FP-Growth. Sebelum mencari asosiasi antar Au dengan mineral lainnya, data pemboran akan dibagi pada setiap interval kedalaman untuk membentuk *itemset* (kumpulan mineral logam dalam satu zona kedalaman) dan nilai numerik dari kandungan mineral logam akan dirubah menjadi bentuk diskrit (*discretisation*) [8].

## 2. Metoda Analisis

Metode pengembangan data mining yang digunakan yaitu CRISP-DM. Model CRISP-DM memiliki urutan mulai dari *business understanding*, *Data understanding*, *preparation / prepoocessing*, pemodelan *data mining* dan terakhir evaluasi. Dalam penelitian ini pemahaman kebutuhan bisnis tidak dimasukkan dalam tahapan penelitian. Berikut adalah tahapan penelitian yang dilakukan:

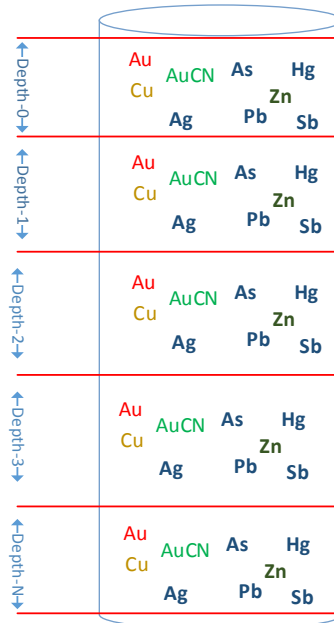


Gambar 1 Alur Analisis Penelitian

1. Pengumpulan data : Data didapat dari database industri pertambangan. Selain pengumpulan dataset, juga dilakukan wawancara dengan orang geologis untuk memahami karakteristik dan informasi yang disimpan dari data pemboran.
2. Studi literature : Pencarian teknik dan algoritma yang dapat digunakan untuk mencari asosiasi pada data pemboran.
3. Preprocessing : Tahapan ini mengolah data pemboran agar dapat dianalisis oleh algoritma *association rule mining*.
4. Analisis Model Data Mining : Tahapan ini memilih algoritma dan menganalisis cara kerja dari Algoritma *data mining* yang dipilih untuk menganalisis data yang dimasukkan
5. ARM pada aplikasi WEKA : Pemrosesan pencarian asosiasi antar data geokimia menggunakan aplikasi WEKA.
6. Evaluasi : Menilai hasil *association rule mining* dari berbagai variable penilaian.

### 2.1 Pembentukan data geokimia menjadi *item* dan *itemsets*

Pembentukan data pemboran menjadi *item* dan *itemsets* adalah langkah *preparation* data sebelum proses *association rule mining*. Pada gambar 2 diilustrasikan sebuah data pemboran secara vertikal. Seperti yang diilustrasikan pada gambar 2, data pemboran dibagi-bagi dalam beberapa zona kedalaman, dan setiap zona kedalaman terdapat beberapa mineral.

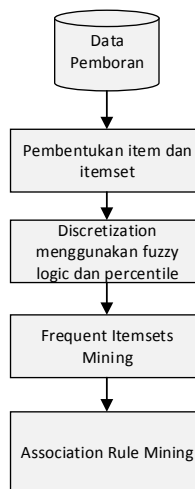


Gambar 2 Pemodelan data geokimia pemboran menjadi itemsets

Setiap zona kedalaman yang telah dibagi – bagi pada data geokimia pemboran dijadikan data *itemsets*. Sedangkan, potensi kandungan setiap mineral pada setiap zona kedalaman dibentuk menjadi item di zona tersebut. Jadi, defth-0 sampai defth-N dibentuk manjadi itemsets. Sedangkan item dibentuk dari mineral Au, AuCN, Cu, Ag, As, Hg, Zn, Pb, dan Sb.

2.2 Proses *Association Rule Mining* Pada Data Geokimia Pemboran

Proses *association rule mining* pada data geokimia pemboran terdiri dari beberapa step dimulai dari pembentukan item dan itemsets dari data pemboran sampai *association rule mining*. Berikut adalah urutan proses *association rule mining* pada data geokimia pemboran.



Gambar 3 Proses *Association Rule Mining* pada data geokimia pemboran

Pada gambar 3 pada step *discretization* menggunakan *fuzzy logic* dan *percentile* bertujuan untuk merubah data geokimia yang bertipe numerik menjadi katagori yang disebut juga fuzzyfikasi. Pada proses fuzzyfikasi data potensi mineral dibagi menjadi 5 *range*. Pembagian *range* dari potensi mineral menggunakan persentase dari teknik *percentile*. Proses *frequent itemsets mining* dan *association rule mining* dilakukan menggunakan algoritma *FP-growth*.

### 3. Hasil dan Analisis

Penjabaran hasil pengujian dilakukan mulai dari pembentukan item dan itemsets dari data pemboran (sub 3.1), kemudian Hasil pembentukan matrik sebagai input software WEKA (sub 3.2) dan terakhir hasil frequent itemsets mining dan association rule mining (sub 3.3).

#### 3.1 Hasil Pembentukan itemsets dan item

Hasil pembentukan data pemboran menjadi Item dan itemsets dijabarkan pada gambar 4 berikut :

	kodelitemset	FROM_	TO_	ITEM
1	DGD742D0	0	6.6	Au=0.4
2	DGD742D0	0	6.6	AuCN=0.6
3	DGD742D0	0	6.6	Ag=0.8
4	DGD742D0	0	6.6	Cu=0.4
5	DGD742D0	0	6.6	Pb=1
6	DGD742D0	0	6.6	Zn=1
7	DGD742D0	0	6.6	As=0.8
8	DGD742D0	0	6.6	Sb=1
9	DGD742D1	6.6	10.6	Au=0.2
10	DGD742D1	6.6	10.6	AuCN=0.6
11	DGD742D1	6.6	10.6	Ag=0.8
12	DGD742D1	6.6	10.6	Cu=0.6
13	DGD742D1	6.6	10.6	Pb=1
14	DGD742D1	6.6	10.6	Zn=1
15	DGD742D1	6.6	10.6	As=0.8
16	DGD742D1	6.6	10.6	Sb=0.8

Gambar 4 Hasil Pembentukan *Item* dan *Itemsets*

Pada gambar 4 menjelaskan pembentukan itemsets berdasarkan range zona kedalaman. Itemsets DGD742D0 pada gambar 4 adalah itemsets yang dibentuk dari data pemboran pada zona kedalaman 0 s/d 6.6 meter. Items dibentuk dari nilai numerik mineral logam yang *difuzzifikasi* kedalam beberapa variabel fuzzy. Hasil pembentukan *itemsets* dan *item* kemudian akan menjadi masukan dari algoritma *frequent itemsets mining* dan *association rule mining*.

#### 3.2 Hasil Pembentukan Matrik Itmesets

Pada penelitian ini untuk melakukan *frequent itemsets mining* dan *association rule mining* menggunakan software WEKA. Inputan software WEKA untuk proses algoritma *frequent itemsets mining* dan *association rule mining* harus menggunakan bentuk matrik dari itemsets dan items. Berikut adalah hasil pembentukan matrik dari itemsets dan items:

kode	Au=0.2	Au=0.4	Au=0.8	Au=1	AuCN=0.2	AuCN=0.4	AuCN=0.8	AuCN=1	Ag=0.2	Ag=0.4	Ag=0.8	Ag=1
1	DGD742D0	0	1	0	0	0	0	0	0	0	1	0
2	DGD742D1	1	0	0	0	0	0	0	0	0	1	0
3	DGD742D2	1	0	0	0	0	0	0	0	0	1	0
4	DGD742D3	0	1	0	0	0	0	0	0	0	1	0
5	DGD742D4	0	0	0	1	0	0	1	0	0	1	0
6	DGD742D5	0	0	0	1	0	0	1	0	0	0	1
7	DGD742D6	0	0	0	1	0	0	1	0	0	0	1
8	DGD742D7	0	0	0	1	0	0	1	0	0	0	1
9	DGD742D8	0	0	0	1	0	0	1	0	0	0	1
10	DGD742D9	0	0	0	1	0	0	1	0	0	0	1
11	DGD742D10	0	0	0	1	0	0	1	0	0	0	1
12	DGD742D11	0	0	0	1	0	0	1	0	0	0	1
13	DGD742D12	0	0	0	1	0	0	1	0	0	0	1
14	DGD742D13	0	0	0	1	0	0	1	0	0	0	1
15	DGD742D14	0	0	0	1	0	0	1	0	0	0	1
16	DGD742D15	0	0	0	1	0	0	1	0	0	0	1
17	DGD742D16	0	0	0	1	0	0	1	0	0	0	1

Gambar 5 Hasil Pembentukan Matrik

Pada gambar 5 ditunjukkan matrik itemsets dan items sebagai masukan dari algoritma *association rule mining* pada software WEKA. Pada gambar 5 setiap row pada tabel menunjukkan sebuah itemset dan kolom selain kolom kode adalah items. Suatu item dikatakan bagian dari itemset pada matrik 5 jika value dari cell antara kode dengan item bernilai 1. Untuk memperjelas isi itemsets dari row ke-1 yaitu DGD742D0 {Au=0.4, Ag=0.8,...}.

Matrik yang telah dibentuk kemudian disesuaikan formatnya dengan kebutuhan software WEKA. Contoh format data untuk software weka ditunjukkan seperti gambar 6

```

1 @relation eksplorasi
2
3 @attribute 'Au=0.2' {0,1}
4 @attribute 'Au=0.4' {0,1}
5 @attribute 'Au=0.8' {0,1}
6 @attribute 'Au=1' {0,1}
7 @attribute 'AuCN=0.2' {0,1}
8 @attribute 'AuCN=0.4' {0,1}
9 @attribute 'AuCN=0.8' {0,1}
10 @attribute 'AuCN=1' {0,1}
11 @attribute 'Ag=0.2' {0,1}
12 @attribute 'Ag=0.4' {0,1}
13 @attribute 'Ag=0.8' {0,1}
14 @attribute 'Ag=1' {0,1}
15 @attribute 'Cu=0.2' {0,1}
16 @attribute 'Cu=0.4' {0,1}
17 @attribute 'Cu=0.8' {0,1}
18 @attribute 'Cu=1' {0,1}
19 @attribute 'Pb=0.2' {0,1}
20 @attribute 'Pb=0.4' {0,1}
21 @attribute 'Pb=0.8' {0,1}
22 @attribute 'Pb=1' {0,1}
23 @attribute 'Zn=0.2' {0,1}
24 @attribute 'Zn=0.4' {0,1}
25 @attribute 'Zn=0.8' {0,1}
26 @attribute 'Zn=1' {0,1}
27 @attribute 'As=0.2' {0,1}
28 @attribute 'As=0.4' {0,1}
29 @attribute 'As=0.8' {0,1}
30 @attribute 'As=1' {0,1}
31 @attribute 'Sb=0.2' {0,1}
32 @attribute 'Sb=0.4' {0,1}
33 @attribute 'Sb=0.8' {0,1}
34 @attribute 'Sb=1' {0,1}
35
36 @data
37 0,1,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,0,0,0,1,0,0,0,1,0,0,1,0,0,0,0,1
38 1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,0,1,0,0,0,1,0
39 1,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,1,0,0,0,1,0,0,1,0,0,0,1,0
40 0,1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,0,1,0,0,0,0,1
41 0,0,0,1,0,0,0,1,0,0,1,0,0,0,0,0,1,0,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1
42 0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0
43 0,0,0,1,0,0,0,1,0,0,0,1,0,0,1,0,0,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1
44 0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0
45 0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,0,1,0,0,0,1
46 0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0
47

```

Gambar 6 Format data untuk inputan WEKA

### 3.3 Pengujian

Pengujian dilakukan pada lingkungan software WEKA, menggunakan algoritma *FP-Growth*. Minimum support yang digunakan untuk menentukan *frequent pattern* yaitu 0.1 (10%). *Frequent pattern* yang digunakan yaitu *frequent itemsets*. *Frequent itemsets* kemudian digunakan untuk mendapatkan aturan asosiasi / *association rule mining*. Berikut adalah *frequent pattern* dan *association rule* yang didapat dari hasil *association rule mining* menggunakan algoritma *FP-Growth*.

1. [Cu=1=1]: 565 ==> [Pb=1=1]: 565 <conf: (1)> lift: (1.01) lev: (0.01) conv: (5.8)
2. [Sb=1=1, Cu=1=1]: 493 ==> [Pb=1=1]: 493 <conf: (1)> lift: (1.01) lev: (0.01) conv: (5.06)
3. [Ag=1=1, Cu=1=1]: 527 ==> [Pb=1=1]: 527 <conf: (1)> lift: (1.01) lev: (0.01) conv: (5.41)
4. [Zn=1=1]: 652 ==> [Pb=1=1]: 650 <conf: (1)> lift: (1.01) lev: (0) conv: (2.23)
5. [Sb=1=1, Zn=1=1]: 559 ==> [Pb=1=1]: 557 <conf: (1)> lift: (1.01) lev: (0) conv: (1.91)
6. [Ag=1=1, Zn=1=1]: 552 ==> [Pb=1=1]: 550 <conf: (1)> lift: (1.01) lev: (0) conv: (1.89)
7. [As=1=1, Zn=1=1]: 523 ==> [Pb=1=1]: 521 <conf: (1)> lift: (1.01) lev: (0) conv: (1.79)
8. [Sb=1=1, As=1=1, Zn=1=1]: 505 ==> [Pb=1=1]: 503 <conf: (1)> lift: (1.01) lev: (0) conv: (1.73)
9. [Ag=1=1, As=1=1, Zn=1=1]: 499 ==> [Pb=1=1]: 497 <conf: (1)> lift: (1.01) lev: (0) conv: (1.71)
10. [Sb=1=1, Ag=1=1]: 737 ==> [As=1=1]: 730 <conf: (0.99)> lift: (1.24) lev: (0.14) conv: (18.35)

Gambar 7 Hasil Association rule mining

Pada gambar 7 ditunjukkan hasil *Association rule mining* pada software WAKA. Pada gambar 7 menceritakan nilai *support* dari *frequent itemsets* dan nilai *confident* untuk aturan asosiasi antar item. Aturan ke-1 yaitu [Cu=1=1] 565 → [Pb=1=1] : 565, memiliki arti bahwa mineral Cu dengan potensi maksimal (Cu=1) dan mineral Pb dengan potensi maksimal (Pb=1) sering muncul bersamaan sebanyak 565 pada dataset, dan tanda anak panah (→) menunjukkan asosiasi kemunculan Cu=1 maka akan muncul Pb=1 dengan jumlah kejadian (*confident*) sebanyak 565 kali.

### 3.4 Evaluasi Hasil

Berdasarkan hasil pengujian *frequent pattern mining* diketahui hanya nilai potensi mineral yang tinggi (bernilai 1) saja yang masuk dalam aturan *frequent itemsets*. Sedangkan untuk hasil *association rule mining* diketahui banyak aturan asosiasi kuat dengan nilai confident sangat tinggi, bahkan ada tiga aturan asosiasi yang sangat kuat dengan nilai confident 100%. Aturan asosiasi yang kuat yaitu rule 1 s/d rule 3. Berdasarkan hasil pengujian membuktikan teknik *frequent pattern mining* dan *association rule mining* dapat membantu menemukan pola kemunculan mineral pada data pemoran yang bertipe numerik dengan cara merubah menjadi bentuk diskrit.

## 4. Kesimpulan

*Association rule mining* pada data bertipe numerik / *continuous* membutuhkan *preprocess* diskritisasi pada setiap item. Penggunaan logika fuzzy dan percentile dapat merubah nilai numerik / *continuous* dari data geokimia pemoran menjadi diskrit (*diskritization*) sehingga dapat mengurangi kompleksitas items pada saat *frequent itemsets mining* dan *association rule mining*. Hasil *frequent itemsets mining* dan *association rule mining* menggunakan algoritma FP-Growth dapat menunjukkan tingkat asosiasi kemunculan antar mineral.

## DAFTAR PUSTAKA

- [1] Agrawal Rakesh, Imielinski, Swami. (1993). *Mining Association Rules Between Sets Of Items In Large Database*. SIGMOD pp 207-216. ACM
- [2] Barnett, C.T and Williams, P.M, (2012), *A Radical Approach to Exploration: Let the Data Speak for Themselves!*, SEG Newsletter, July 2012, pp 12-17
- [3] Marjoribanks, R., (2010), *Geological Method in Mineral Exploration and Mining*, Second Edition, Springer, Australia.
- [4] Setyadi, H., Anggayana, K., (2012), *Database Management and Quality Assurance is The Key of Success in Exploration*, International symposium Earth Science and Technology 2012, pp. 61-66.
- [5] Han Jiwei, Kamber, Pei., (2012), *Data Mining concepts and techniques third edition*. Morgan Kaufmann publishers
- [6] Hall Mark, Frank, Hollmes, Pfahringer. (2009). *The WEKA data mining software: an update*. SIGKDD Explorations, pp 10-18. ACM
- [7] IAN H. witten, Eibe Frank, Mark A. Hall., (2011), *Data Mining practical machine learning tools and techniques third edition*. Morgan Kaufmann publishers
- [8] Xing Xiaoxue, Guan Xiuli, Shang Weiwei. (2014). Continuous Attribute Discretization Algorithm of Rough Set Based On K-Means. Workshop on advanced research and technology in industry applications (WARTIA), IEEE
- [9] Wandu Nugroho, Rully, Handrawan, Muklason. (2012). Pengembangan Sistem Rekomendasi Penelusuran Buku Dengan Pengalihan Asosiasi Rule Menggunakan Algoritma Apriori (Studi Kasus Badan Perpustakaan dan Kearsipan Provinsi Jawa Timur). Jurnal teknik ITS Vol. 1, ISSN: 2301-9271
- [10] Mega Nata. *Knowledge discovery* dari data pemoran untuk mendukung prediksi potensi kandungan emas. Tesis Informatika. Bandung: Institute Teknologi Bandung; 2015