

PENDEKATAN ATURAN ASOSIASI UNTUK ANALISIS PERGERAKAN SAHAM

Azhari¹, Anshori²

^{1,2} Intelligent System Research Group, Computer Sciences, FMIPA UGM
arisn@ugm.ac.id

Abstract

Financial analysis of listed companies is essential in stock investment management for maximizing investment return with controlled risk. Usually the analysis involves large amount of data and the underlying patterns not easy to identify. With the increasing power of computer, data mining becomes possible in the domain of stock investment.

The focus of this research is to extract association rules for stock ratio against certain variables. A data mining application for association rules on Jakarta Stock Exchange stock movement archives has been built and to be accessed through the WEB. Preprocessing has been done to convert the data on financial report to financial ratios to allow for direct comparison. Apriori, an algorithm for association rules mining is implemented for efficiency comparison. Apriori algorithm used support and confidence in searching the rules. Support is the amount of transactions which contain an itemset. Confidence is used in order to measure how often an itemset in Y appear in transactions that contain an itemset X. The prototype has been tested for same data stock exchanges. And the results showed that the prototype can construct same association patterns and display it in a bar chart.

Keywords: *Data mining, knowledge discovery, association rule, stock exchange*

1. Pendahuluan

Dunia investasi saham yang penuh persaingan membuat para pelakunya harus selalu memikirkan strategi-strategi jitu yang dapat menjamin kelangsungan investasi mereka. Pergerakan saham dipengaruhi banyak faktor yang kemudian menjadi pertimbangan perusahaan dalam berinvestasi.

Kemampuan teknologi informasi untuk mengumpulkan dan menyimpan berbagai tipe data jauh meninggalkan kemampuan untuk menganalisis, meringkas dan mengekstraksi pengetahuan dari data. Metodologi tradisional untuk menganalisis data yang ada, tidak dapat menangani data dalam jumlah besar, kondisi yang sering disebut "*rich of data but poor of information*" karena data yang terkumpul tidak dapat digunakan untuk aplikasi yang berguna.

Peneliti melihat peluang untuk melahirkan sebuah teknologi baru yang menjawab kebutuhan ini, yaitu *data mining*. *Data mining* merupakan suatu proses yang menggunakan teknik statistik, visualisasi, kecerdasan buatan, dan *machine-learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database (Han dkk., 2001). Salah satu teknik *data mining* yang sering digunakan adalah aturan asosiasi. Dimana dengan aturan asosiasi dapat ditemukan hubungan asosiatif antara data dan menyimpulkan aturan dari data-data yang ada tersebut. Disinilah proses *data mining* menggunakan aturan asosiasi dapat digunakan.

Sebagai contoh, seorang investor bertanya "*apakah nilai return perusahaan berhubungan dengan volume permintaan ataupun volume penawaran?*". Aturan asosiasi berfungsi menemukan pola-pola, asosiasi dan hubungan antara data tersebut dan dapat menemukan fakta yang sebelumnya tidak diketahui atau bahkan tidak diperhatikan oleh sebuah perusahaan.

2 Knowledge Discovery dan Data Mining

2.1 Knowledge Discovery didalam Basisdata

Fayyad (1996) dan Dunham (2003) menggunakan istilah *Knowledge Discovery in Databases* (KDD) untuk menunjuk pada keseluruhan proses pencarian pengetahuan dalam kumpulan data jumlah besar. Mereka menggolongkan *data mining* sebagai salah satu langkah dalam proses KDD dikarenakan oleh penerapannya terhadap suatu algoritma spesifik dalam mencari pola-pola (model) dalam data.

Dunham (2003), Witten dan Frank (2005), dan Kantardzic (2003) dalam aplikasi penyaringan pengetahuan pada sebuah data mining proses KDD dapat terdiri dari langkah-langkah berikut (lihat Gambar 1):

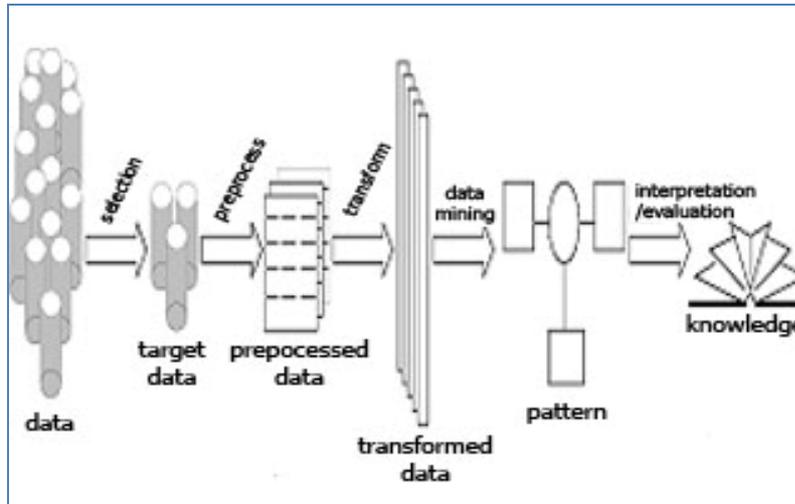
1. *Data Selection*

Proses pemilihan dan identifikasi data yang akan dipertimbangkan dalam proses

2. *Data Preprocessing*

Proses dimana kesalahan dan kehilangan data harus diatasi. Hal ini melibatkan koreksi atau menggunakan nilai-nilai terprediksi.

3. *Data Transformation*
Proses dimana data dibentuk menjadi format yang dapat diolah aplikasi
4. *Data mining*
Proses dimana algoritma diterapkan terhadap data yang telah ditransformasikan untuk menghasilkan output yang dikehendaki.
5. *Interpretation / Evaluation*
Agar hasil dari proses berguna, maka harus dipresentasikan dalam sebuah cara yang bermanfaat bagi pengguna.



Gambar 1. Proses Knowledge Discovery

2.2 Data Mining

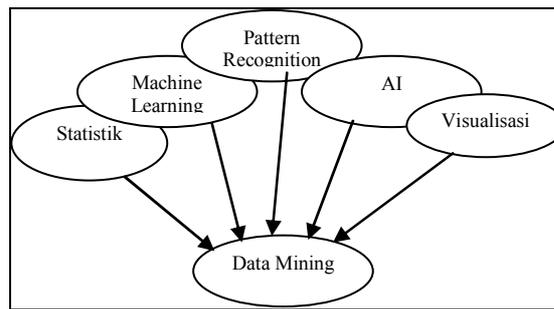
Data mining adalah eksplorasi dan analisis terhadap sejumlah data dengan tujuan untuk menemukan pola dan aturan yang sangat penting (Dunham, 2003; Truban, dkk. 1988).

Bidang Ilmu Data Mining

Menurut Han, dkk. (2001), *data mining* merupakan penerapan bidang ilmu interdisipliner, statistik, database, *machine learning*, *pattern recognition*, kecerdasan buatan, dan visualisasi. Bidang-bidang tersebut memiliki peranan yang saling berhubungan, sehingga sangat sulit untuk memberikan batasan yang jelas diantara masing-masing disiplin ilmu tersebut dengan *data mining*, ditunjukkan pada Gambar 2.

Teknik-teknik Data Mining

1. *Klasifikasi (Classification)*
Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.
2. *Regresi (Regression)*
Model dengan tipe regresi ini menggunakan serangkaian nilai-nilai atribut yang ada untuk memprediksi suatu nilai kontinu.
3. *Aturan Asosiasi (Association Rule)*
Aturan asosiasi adalah teknik *data mining* untuk menemukan aturan asosiatif antara kombinasi *item*. Penting atau tidaknya suatu aturan asosiasi dapat diketahui dengan dua parameter, *support* yaitu persentase kombinasi *item* tersebut dalam database dan *confidence* yaitu kuatnya hubungan antara *item* dalam *association rule*.
Algoritma untuk aturan asosiasi yang paling populer dikenal dengan nama *algoritma apriori* dengan paradigma *generate and test*, yaitu pembuatan kandidat kombinasi *item* yang mungkin berdasarkan aturan tertentu, kemudian diuji apakah kombinasi *item* tersebut memenuhi syarat *support* minimum. Kombinasi *item* yang memenuhi syarat tersebut disebut *frequent item set* atau *large item set*, yang kemudian dipakai untuk membuat aturan-aturan yang memenuhi syarat *confidence* minimum. Algoritma dari aturan asosiasi yang lain adalah algoritma *predictive apriori* dan *tertius*.
4. *Clustering*
Berbeda dengan klasifikasi, regresi, dan aturan asosiasi dimana kelas data telah ditentukan sebelumnya. *Clustering* melakukan pengelompokan data tanpa berdasarkan kelas data tertentu. Bahkan *clustering* dapat digunakan untuk memberikan label pada kelas data yang belum diketahui tersebut.



Gambar 2. Bidang Ilmu yang Mempengaruhi Data Mining

2.3 Association Rules Mining

Association rules mining (aturan mining asosiasi) berfungsi untuk menemukan asosiasi antar variabel, korelasi atau suatu struktur diantara *item* atau objek-objek didalam database transaksi, database relasional, maupun pada penyimpanan informasi lainnya (Agrawal dan Srikant, 1994; Tan dkk., 2006)

Sebuah aturan asosiasi adalah implikasi dari $X \Rightarrow Y$, dimana $X \subset I$, $Y \subset I$, dan $X \cap Y = \phi$. Aturan $X \Rightarrow Y$ berada di dalam himpunan transaksi D dengan kepercayaan (*confidence*) c , jika $c\%$ dari transaksi dalam D yang ada X terdapat juga Y. Aturan $X \Rightarrow Y$ memiliki dukungan (*support*) s di dalam set transaksi D, jika $s\%$ transaksi dalam D terdapat $X \cup Y$ (1).

Support itemset (1) adalah jumlah transaksi T yang didalamnya terdapat (1). *Confidence* mengukur seberapa besar ketergantungan suatu *item* dengan item yang lainnya. *Frequent itemset* adalah *itemset* yang memiliki nilai support lebih atau sama besar dengan nilai minimum *support* yang ditentukan (2).

Dimisalkan suatu aturan R : $X \Rightarrow Y$, maka:

$$\text{Support}(X \Rightarrow Y) = \text{support}(\{X\} \cup \{Y\}) \quad (1)$$

$$\text{Confidence}(X \Rightarrow Y) = \frac{\text{support}(\{X\} \cup \{Y\})}{\text{support}(\{X\})} \quad (2)$$

Sebuah aturan (*rule*) biasanya terdiri dari dua bagian, yaitu kondisi (*condition*) dan hasil (*result*). Dan biasanya disajikan dalam pernyataan sebagai berikut :

$$\text{Jika kondisi maka hasil (If condition then result)} \quad (3)$$

2.4 Algoritma Apriori

Algoritma *apriori* merupakan algoritma pertama dan sering digunakan untuk menemukan aturan asosiasi di dalam aplikasi *data mining* dengan teknik aturan asosiasi (Witten dan Frank, 2005; Kuntardzic, 2003).

Tujuan dari *algoritma apriori* adalah untuk menemukan aturan (*rule*) yang memenuhi minimum *support* yang telah ditetapkan sebelumnya dan memenuhi nilai *confidence* yang disyaratkan.

Tugas dari algoritma *apriori* dalam menemukan aturan asosiasi dapat dibagi menjadi dua fase, yang pertama untuk menemukan *itemsets* dan kemudian membentuk aturan (*rule*) dari *itemsets* yang ditemukan. Potongan *pseudo code* dari algoritma *apriori* ditunjukkan pada Gambar 4.

```
Ck: Candidate itemset of size k
Lk : frequent itemset of size k

L1 = {frequent items};
for (k = 1; Lk !=∅; k++)
do begin
    Ck+1 = candidates generated from Lk;
    for each transaction t in database do
        increment the count of all candidates in Ck+1 that are contained
        in t
    Lk+1 = candidates in Ck+1 with min_support
end
return ∪k Lk;
```

Gambar 4. Algoritma Apriori

2.5 Analisis Investasi

Pengujian efisiensi pasar umumnya terbukti apabila dipergunakan uji *filter rules*. Korelasi antar variabel (indikator teknis) yang terdapat pada data historis diharapkan memiliki pola – pola aturan (*rules*) tertentu yang belum tentu diketahui para analisis (Suad, 1990). Uji *filter rules* yang bersifat deskriptif ini berarti kita mencari *trading rules* tertentu, dan kemudian kita bandingkan antar variabel atas *rules* tersebut. *Trading rules* adalah salah satu bentuk pengaplikasian aturan asosiasi Pola antar variabel yang ditemukan dapat digunakan sebagai acuan dalam penerapan *buy-and-hold strategies*.

3 Desain dan Implementasi

3.1 Arsitektur Sistem

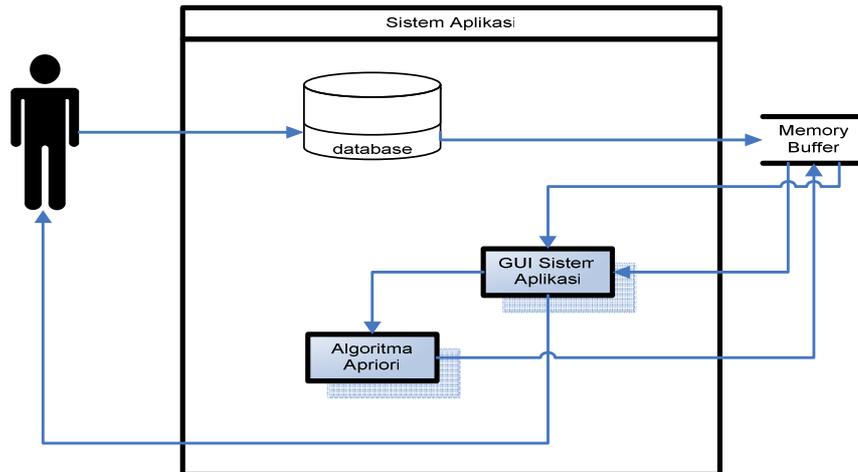
Arsitektur dari aplikasi menggunakan database sebagai input dataset yang akan diproses oleh sistem. Hasil pemilihan parameter pada database tersebut kemudian disimpan ke dalam memory buffer sistem. Proses utama dari aplikasi data mining aturan asosiasi menggunakan klas asosiasi, yang didalamnya terdapat algoritma *apriori* untuk memproses input, dan mencari aturan yang dapat disimpulkan dari dataset yang telah diinputkan. Kemudian hasil ditampilkan melewati antarmuka sistem aplikasi. Perincian dari arsitektur aplikasi ini dapat dilihat pada Gambar 5.

3.2 Implementasi

Algoritma apriori sebagai proses utama dari association rules mining dimulai dengan proses pembentukan kandidat *itemset*, dan akan dilanjutkan dengan *pruning* yang dimaksudkan untuk menemukan *frequent itemset*.

Pada proses pembentukan kandidat *itemset* yang pertama ini cukup sederhana, yaitu semua itemset yang terdapat didalam data transaksi dapat menjadi kandidat *itemset*. Kemudian kandidat *itemset* yang tidak memenuhi nilai minimum *support* akan dihapus.

Setelah proses penggabungan *itemset* dilakukan, kemudian lakukan proses *pruning itemset*. Proses *pruning itemset* dilakukan dengan mengecek kandidat *itemset* (C_k) dengan L_{k-1} . Apabila pada C_k terdapat *itemset* yang tidak terdapat di L_{k-1} , maka kandidat tersebut tidak akan dibentuk menjadi C_k . Pada Gambar 6 ditunjukkan potongan program dari implementasi proses merge itemset, sedangkan pada Gambar 7 ditunjukkan potongan program dari implementasi pruning itemset.



Gambar 5. Arsitektur Sistem

```
function Get($atr,$x,$atrVal,$atrJns,&$i,$atrR,$n) {
  for ($j=1;$j<$x;$j++) {      $temp = array_pop($atrR); }
  for ($j=1;$j<$n;$j++) {      $temp = array_shift($atrR); }

  foreach ($atrR as $row => $value) {
    $n++;
    foreach ($atrR[$row] as $row2 => $value2) {
      if (count($this->atrVal[$i])==0) {$this->atrVal[$i] = array();}
      if (count($this->atrJns[$i])==0) {$this->atrJns[$i] = array();}
      if (count($this->tempVal[0])==0) {$this->tempVal[0] = array();}
      if (count($this->tempJns[0])==0) {$this->tempJns[0] = array();}
      $this->atrVal[$i] = $this->tempVal[0];
      $this->atrJns[$i] = $this->tempJns[0];
      array_push($this->atrVal[$i],$row);
      array_push($this->atrJns[$i],$value2);
      $x--;
      if ($x > 0) {
        array_push($this->tempVal[0],$row);
        array_push($this->tempJns[0],$value2);
        $this->Get($atr,$x,$this->atrVal,$this->atrJns,$i,$atr,$n);
        $temp = array_pop($this->tempVal[0]);
        $temp = array_pop($this->tempJns[0]);
      }
    }
    else {
      $i++;
      $this->tempN++;
    }
  }
  $x++;
}
}
```

Gambar 6. Merge Itemset

```
function SetItemSet($jnsAtrAktif, $dataAktif, $repNum, $supportBawah,
$supportAtas, &$v)
{
.
.
.
foreach ($tempNum as $row => $value) {
$tempNum[$row] /= $n;
if (($tempNum[$row] >= $supportBawah) && ($tempNum[$row] <=
$supportAtas)) {
array_push($temp, $row);
}
else {
$flag = true;
}
}
}
```

Gambar 7. Pruning Itemset

Implementasi dari user interface secara sederhana dengan dua window utama, yaitu menu dan window dari proses utama ditunjukkan pada Gambar 8.



Gambar 8. The Main Screen of Application

4 Hasil dan Pembahasan

Pada pengujian sistem aplikasi ini akan dicari aturan-aturan dari keseluruhan data transaksi saham yang terdapat pada database dengan pemilihan sektor industri barang konsumsi, periode tanggal dari 16 agustus 2004 hingga 16 oktober 2004. Dengan mencari korelasi akan dipengaruhi oleh 4 jenis variabel saham, sub industri, hari kejadian transaksi, return harian dan beta mentah. User menginputkan parameter yang dibutuhkan untuk menjalankan proses data mining, yaitu:

- BatasBawah Support : 0.2
- BatasAtas Support : 1
- Confidence : 0.7
- Jumlah Aturan : 10

Dari *frequent itemset* diatas, kemudian dibentuk aturan asosiasi yang memenuhi minimum *confidence* 0.7. Didapatlah 5 aturan asosiasi yang memenuhi minimum *confidence*, ditunjukkan pada Tabel I.

Tabel I. Sample assosiasi rule

No	Aturan	Conf
1	(SUBINDNAMA) farmasi, (NHARI) Selasa, (RETH) ~ (3) maka (BETAMENHST) - (3)	1
2	(SUBINDNAMA) farmasi, (NHARI) Selasa, (BETAMENHST) - (4) maka (RETH) ~ (3)	0,75
3	(SUBINDNAMA) farmasi, (NHARI) Selasa (4) maka (RETH) ~, (BETAMENHST) - (3)	0,75
4	(NHARI) Selasa, (RETH) ~, (BETAMENHST) - (4) maka (SUBINDNAMA) farmasi (3)	0,75

Dari pengujian yang telah dilakukan, aturan asosiasi yang ditemukan seperti yang ditunjukkan pada Gambar 9. Dari seluruh transaksi saham yang terjadi, yaitu sebanyak 20 transaksi yang sesuai, dapat disimpulkan bahwa ketika sub industri farmasi yang melakukan transaksi pada hari selasa dan mempunyai *return* konstan maka memiliki beta negatif dengan confidence sebesar 1. Pada aturan berikutnya dari pengujian satu, terdapat aturan yang menggambarkan bahwa ketika sub industri farmasi yang melakukan transaksi pada hari selasa maka akan memiliki *return* konstan dan beta negatif dengan confidence 0,75. Dari sini terlihat bahwa saham sub industri farmasi pada hari selasa dapat dikatakan *fair* karena tingkat resiko yang dihadapi rendah untuk mendapatkan *return* konstan.

Aturan asosiasi atau informasi yang telah ditemukan selama proses data mining, kemudian digunakan oleh pemilik data dalam peningkatan efisiensi perusahaan dan yang lainnya.

Informasi tadi dapat digunakan oleh perusahaan, dengan cara sebagai berikut:

1. Investor dan perusahaan dapat menilai baik tidaknya saham pada periode tertentu.
2. Investor akan mempertimbangkan pembelian saham dengan tingkat resiko yang rendah tetapi memiliki *return* yang tinggi dan nilai peningkatan pada *abnormal return*
3. Investor akan enggan membeli saham dengan tingkat resiko yang tinggi, tetapi menghasilkan *return* yang cenderung turun.
4. Perusahaan dimungkinkan akan meningkatkan volume penjualan sahamnya pada periode tertentu ketika harga permintaan lebih tinggi dari pada harga penawaran
5. Perusahaan dapat mengobrol sahamnya jika volume penawaran lebih tinggi dari pada volume permintaan dan harga penawaran lebih rendah dari pada harga penawaran, serta memiliki *return* yang cenderung turun.

4 Kesimpulan

Sistem ini dapat digunakan untuk mencari pola-pola tersembunyi atau aturan-aturan asosiasi yang sebelumnya tidak diketahui dari sekumpulan pada data pergerakan saham. Prototip dari aplikasi sistem ini telah dicoba untuk sejumlah sampel data pergerakan saham dan mampu menampilkan diagram batang sebagai visualisasi jumlah aturan asosiasi yang ditemukan.

References

- Agrawal R. and Srikant R., "Fast Algorithm for Mining Association Rules", VLDB Conference, 1994.
Dunham M., "Data Mining: Introductory and Advance Topics", Prentice Hall, 2003.
Fayyad, U., "From Knowledge Discovery in Database", AI Magazine, 1996.
Hand D., Mannila H. and Smyth P., "Principles Of Data Mining", The MIT Press, 2001.
Kantardzic M., "Data Mining: Concepts, Models, Methods, and Algorithms", John Wiley and Sons, 2003.
Suad H., "Dasar – Dasar Teori Portofolio dan Analisis Sekuritas", UPP AMP YKPN, 2001.
Tan P., Steinbach M., and Kumar V., "Association Analysis: Basic Concepts and Algorithms", <<http://www-users.cs.umn.edu/~kumar/dmbook/index.php>>, diakses terakhir 22-03-2006.
Turban, E., Jay E. A., and Liang T., "Decision Support Systems and Intelligent Systems", Prentice-Hall, 1998.
Witten H. and Frank E., "Data Mining Practical Machine Learning Tools and Techniques", Morgan Kaufmann, 2005.

----00----