DECISION FOR BUYING AND SELLING STOCK
WITH DECISION TREE ALGORITHM

Nur Aini Rakhmawati, Erma Suryani
Department of Information System, Faculty of Information Technology,
Sepuluh Nopember Institute of Technology
Jl. Raya ITS
Email: iin@its-sby.edu, erma@its-sby.edu

ABSTRACT
Stock is high risk investment. To conduct decision in buying and selling stock is not easy matter because many factors influencing increase and degradation of stock price. A decision tree is built for supporting of taking the decision. This decision tree is built from data training that contains a set of correct moment fact data to buy and sell stock. Attribute of data of training is factors influencing increase and degradation of price stock. The tree building is used C4.5 algorithm. After constructed decision tree, not all attributes have influence decision for buying and selling stock.

Keyword: Stock, Decision Tree, C4.5

1 INTRODUCTION
Conducting a planning of investment particularly again long term investment is not easy matter. One of investment type promising high capital gain is stock investment. In stock exchange, stock price mechanism is fully determined by market condition like banking rate of interest and inflation. Decision of stock sales is decision needing accurate calculation pursuant to long-range historical data. Obliging to sell stock when its price is going up or sell stock when its price is going down. Decision making pursuant to eye emotion can damage investment which have been done.

2 STOCK
Stock is sign ownership of organization or someone in a company. Stock is divided to become two that is common stock and preferred stock.

2.1 Common Stock
Common Stock is stock that placing its owner most bottom on division of rights and dividend of estate properties of company if the company will be liquidation. The other characteristic of common stock is dividend is given during company obtain profit

2.2 Preferred Stock
Preferred stock is stock owning alliance characteristic among common stock and obligation, because its can yield earnings remain to (like obligation interest). The similarity between preferred stock and obligation is
1. There is claim of profit and asset before all
2. Dividend is remaining to during a period of stock.
3. We have rights redeem and can be exchanged with common stock.

3 DECISION TREE AND C4.5
Decision tree is one of classification method in Data Mining. Its uses tree as model for classification process. Its approach to classification is to divide the search apace into rectangular region. If a database D has r1, r2… rn records and A1 A2… An attributes and there are a set of classes C1, C2… Cm, the decision tree has properties:
1. Internal Node is labeled with an attribute A and has two or more outgoing edge.
2. Leaf Node is labeled with a class C and only has one incoming edge. It has not outgoing edge.
C4.5 is a decision tree generating algorithm, based on the ID3 algorithm. It contains several improvements such as:
1. Choosing an appropriate attribute selection measure.
2. Handling training data with missing attribute values.
3. Handling attributes with differing costs.
4. Handling continuous attributes
C4.5 algorithm has several following steps:
1. If records are all of the same class, C then return N as Leaf node labeled with the class C.
2. If attribute-list is empty then return N as leaf
3. If records are belonging to a various class, D must be split into subset item. It uses the
concept of information gain or entropy reduction for selecting the optimal split. Let S is subset of D and F (C_i,S) stand for the number of records in S that belong to class C_i and NS denotes the number of records in the subset S. Then Entropy of S:

\[ E(S) = -\sum_{i=1}^{k} (F(C_i,S)/|S|) \log_2 (F(C_i,S)/|S|) \]

After it splits, Entropy of attribute A must be found:

\[ E(A) = \sum_{j=1}^{n} (|A_j|/|D|).E(A_j) \]

Information Gain is defined:

\[ Gain = E(A) - E(S) \]

Select a test attribute that has highest information gain.

4. Pruning tree is replacing a sub tree with a leaf node. This is need for simplify decision tree. C4.5 conducts pruning after a tree is constructed. A sub tree will be replaced, if the Node predicted error value is greater than the Tree predicted error value.

4 FACTOR FOR STOCK EXCHANGE

Political factor have strong influence at economic world, especially after a period of reform. One of the important economic indicators is Stock Alliance Price Index (IHSG).

Determination of sales decision can be determined by two analyses that are fundamental and technical analysis.

4.1 FUNDAMENTAL ANALYSIS

Stock market Perpetrator basically consist of two groups, the first group is speculators and the second is investors. Investors have long-range directive, because they expecting to get benefit from two earnings source, that are stock increase of price and dividend. Because dividend only release quickest for half year once, hence investor have to keep the stock at least six-months, so it can be said that investors take fundamental analysis. There are many things that influencing stock price:

- Emiten performance
- Interest rate
- Political atmosphere

4.2 TECHNICAL ANALYSIS

Speculator more consider the technical aspect, so their decision making of sales only paying attention growth of movement of price and commerce volume (Stock Alliance Price Index). There are three component of price:

- Opening price
- The highest price
- The lowest price

Price also has some patterns of movement. They can be grouped into six groups. They are:

- Tendency (trend)
- Volatilities
- Momentum
- Cyclic
- Strength of market
- Supporter and resistance indicator

Experience of the expert of technical analysis will determine the indicator which is assumed compatible in certain situation. Technical analysis also needs complicatedly analysis.

4.3 MOVEMENT OF PRICE

Even investor rely on fundamental aspect, so they concern to long-range price graph (more than one year), they still have to concern with short-range graph. It can be happened that the fundamental factors which have been calculated before, in reality cannot accommodate the unforeseen growth short-range as happened in BEJ nowadays. Graph movement of Stock Alliance Price Index for year 1991 to year 2000 can be seen at figure 1.

From the graph of Stock Alliance Price Index, it can be seen that in the reality have happened two low dales that are on 11 November 1991 and 21 September 1998. Connectively both this dale, we can discover the supporter line depicting ugly stock endurance in a period of 12 year.

In the year 2000 we can see that Stock Alliance Price Index is going down incisively and when tops (there are four) which have been formed to be contacted one another, we discover a resistance line that indicates the strength of market is depressing downwards.

The resistance line will continue if it let, and in November 2000, Stock Alliance Price Index can penetrate supporter line at value 269 as forerunner that situation of economics have come to worse than a period of 12 years before.
4.4 DECISION OF SALES STOCK

Existence of stock which have overvalued or undervalued is very important for effect broker, financial analyst and investment manager to get more profit. If stock prices have the including overvalued, it indicate that we should sell the stock and if undervalued, the decision is making indication buy.

Therefore very necessary for effect broker, financial analyst and investment manager to have a model that can be used to take decision sell - buy stock.

In determining decision when correct moment to sell and buy stock, it requires to be considered the Tendency (Trend) and Market Condition. Trend and market condition determine decision type to be taken, so that first matter which ought to be done is to evaluate to process decision making of investment.

Two important decision type in stock market investment is decision when have to buy and sell stock. Long-range Historical data (trend term long) will assist in decision making of stock sales. Decision to sell stock is taken if stock price fall to 10% from ceiling price during specified period. While decision of purchasing of stock is taken at the time of price go up equal to 15% from the lowest price after stock price have selected downhill tendency during a certain periods (two years periods).

5 DATA

For this research, the most appropriate data set is data synthetic because during this time, there was no set of data that gather stock and the factor that influence it. This data is created by studying literature what commonly happens, when the stock must be sold or bought. There are 60 records data and 5 attributes that are used in this research:

5.1 Company Performance

Emiten Performance is related with Company Performance from over all like Return On Equity and Return On Asset. The bigger the value is, the better the performance. Its type data is nominal: [BAIK, BURUK]. From dataset, there were 32 data with BAIK value and 28 with BURUK value.

5.2 Indonesian Bank Interest Rate

Interest Rate is related with Indonesian Bank Interest Rate. If the interest is high, investor will sell the stock and invest the fund in bank. Its type data is numeric. From dataset, there were 32 distinct values data. Its unique value is 25%.

This attribute has continuous numeric values. 13.25 is defined as threshold, because it is median from dataset. From threshold, a binary test outcomes Interest Rate ≤ 13.25 and Interest Rate > 13.25.

5.3 Social Political atmosphere

Politics Social Condition is related with Social Political atmosphere in country. If social condition of politics is conducive, commerce in stock market will achieve high volume, stock price will high. Its type data is nominal: [BAIK, BURUK]. From dataset, there were 38 data with BAIK value and 22 with BURUK value.

5.4 Stock Alliance Price Index

Index Price is related with Alliance of Price Index. The bigger the value is, the better the condition of stock market. Its type data is numeric. From dataset, there were 25 distinct values data. Its unique value is 15%.

This attribute has continuous numeric values. 60 is defined as threshold, because it is median from dataset. From threshold, a binary test outcomes Price Index ≤ 60 and Price Index > 60.

5.5 Stock Price

This attribute is related with Stock Price. If stock prices have the including overvalued, it indicate that we should sell the stock and if undervalued, the decision is making indication for buying. Its type data is numeric. From dataset, there were 17 distinct values data. Its unique value is 12%.

This attribute has continuous numeric values. -2 is defined as threshold, because it is median from dataset. From threshold, a binary test outcomes Stock Price ≤ -2 and Stock Price > -2.

6 BUILDING TREE

Each attributes must be calculated each information gain.

Table 1. Information Gain

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Information Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company Performance</td>
<td>0.4701</td>
</tr>
<tr>
<td>Indonesia Bank Interest Rate</td>
<td>0</td>
</tr>
<tr>
<td>Social Politic Atmosphere</td>
<td>0.0162</td>
</tr>
<tr>
<td>Stock Alliance Price Index</td>
<td>0</td>
</tr>
<tr>
<td>Stock Price</td>
<td>0.2794</td>
</tr>
</tbody>
</table>

From the first iteration the highest information gain is company performance (Kinerja). It is selected as test attribute. For remain attribute is iterated again recursively until all attribute will be Internal Node.

After the tree is built, the sub tree replace with leaf node by pruning. The predicted error for tree and node is used confident factor 25%.

![Figure 2 Decision Tree](image-url)
Decision tree that have been built can be transformed to decision rules.

if kinerja = Buruk
  then Decision = Jual (28)
if kinerja = Baik and Harga <= -2
  then Decision = Beli (15)
if kinerja = Baik and Harga > -2
  and IHSG <= 60
  then Decision = Jual (9)
if kinerja = Baik and Harga > -2
  and IHSG > 60
  then Decision = Beli (8/1)

7 CONCLUSION

From 5 attributes that influence the moving of Stock Alliance Price Index, only 3 attributes that influence person for selling and buying stock. There are company performance, stock price and Stock Alliance Price Index.

REFERENCE

[1]. Dunham, Margaret H. , “Data Mining introductory and advance Topics”, Pearson Education. USA : 2003
[2]. Han, Jiawei. “Data Mining: Concept and Techniques”. Morgan Kaufman Publisher. USA : 2001