

# Real Time Business Analytics for Buying or Selling Transaction on Commodity Warehouse Receipt System

**Taufik Djatna<sup>1</sup>, Wellem A. Teniwut<sup>2</sup>, Nina Hairiyah<sup>3</sup>, Marimin<sup>1</sup>**

<sup>1</sup>Department of Agroindustrial Technology, Faculty of Agricultural Technology, Bogor Agricultural

<sup>2</sup>Fisheries Agribusiness Study Program, Tual State Fisheries Polytechnic, Langgur, Indonesia

<sup>3</sup>Agro-Industrial Technology Study Program, Tanah Laut State Polytechnic, South Kalimantan, Indonesia

Email: [taufikdjatna@ipb.ac.id](mailto:taufikdjatna@ipb.ac.id)

**Abstract.** The requirement for smooth information such as buying and selling is essential for commodity warehouse receipt system such as dried seaweed and their stakeholders to transact for an operational transaction. Transactions of buying or selling a commodity warehouse receipt system are a risky process due to the fluctuations in dynamic commodity prices. An integrated system to determine the condition of the real time was needed to make a decision-making transaction by the owner or prospective buyer. The primary motivation of this study is to propose computational methods to trace market tendency for either buying or selling processes. The empirical results reveal that feature selection gain ratio and  $k$ -NN outperforms other forecasting models, implying that the proposed approach is a promising alternative to the stock market tendency of warehouse receipt document exploration with accurate level rate is 95.03 %.

## 1. Introduction

Warehouse receipt system is a financial system for storing agricultural commodity and values to cope with a fluctuation in prices. Commodities stored in the warehouse will be secured by a document warehouse receipt issued by the warehouse management and has been officially registered with the National Registration Centre. Warehouse receipt document can be traded (buying and selling) on the stock market by the holder of the warehouse receipt document. Currently, open access to the internet and mobile telecommunication enabled each party in warehouse receipt system to transact in a farm of buying or selling. Parties may consist of farmers, traders, financial institution or else. They need an integrated tool such as business intelligence system anywhere and anytime with the real time data access. A warehouse receipt is a document of proof of ownership of goods stored in a registered warehouse specifically published by the warehouse manager. Warehouse Receipt can be transferred enough with an endorsement. Receipt Warehouse with it being "Negotiable." Warehouse here means it can be various, depending on the stored commodities, ranging from dried seaweed, chocolate, coffee, rice, to crude palm oil (CPO). This warehouse receipt can be used as collateral for bank loans. Since the warehouse receipt is a proof of ownership, the warehouse receipt may be traded, traded, exchanged or used as collateral for the loan, or may be utilized for the delivery of goods in derivative transactions such as futures contracts.



But unfortunately, the use of warehouse receipts is still insufficient because most countries are not willing to accept the concept of proof of ownership of the moving goods. Usually, the evidence of ownership exists only for immovable goods.

The scope of Business Intelligent (BI) should include making the best use of information for strategic, tactical, and operational needs [1]. The purposes in building BI strategy is to help business with long-term planning, help middle management with tactical reporting, and help operations with day-to-day decision-making to run the business efficiently. BI is all about providing people with the information they need to do their jobs more efficiently. A wide range of BI services needs to be given to meet a wide variety of requirements. A scope of BI strategy should be determined by the business drivers and business goals. The scope should always account for the changing business requirements to keep the BI strategy aligned with business (Figure 1).

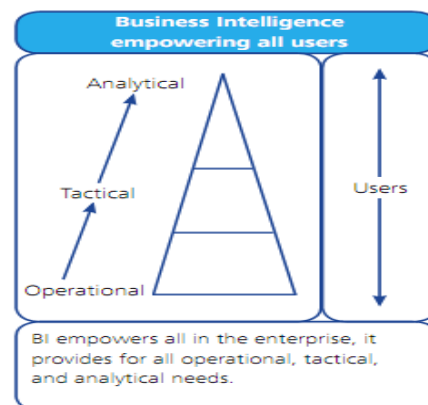
BI framework brings together data governance, data architecture, technical architecture, data integration, data quality, end-user information delivery, data security, etc. to empower the BI initiatives. A framework should set standards that BI participants must adhere to. The framework should connect significant components that are part of overall BI vision. Establishing BI competency center or center of excellence as part of the BI framework will help in integrating BI best practices with the ongoing BI work and the BI environment of the enterprise.



**Figure 1.** BI aligning with enterprise strategy to deliver value [1]

The overarching methodology that refers to the skills and technologies to explore past business performance to make better decisions is called Business Analytics. Business Analytics uses data, statistical and quantitative analysis, predictive modeling, and optimization to make businesses work better [2]. Business analytics are needed to make the best decision on the sale or purchase transaction tendencies in real time by the owner or prospective buyer. The position of business analytics on enabling intelligent business all in the enterprise is located in the upper position (Figure 2).

BI strategy should aim to support the complete breadth of decision-making ability in the enterprise. Strategic decisions deal with the long-term planning, performed by top management, focus mostly on demographic, and industry trends. These address broad issues to achieve general objectives. Strategic decisions are where BI traditionally has been implemented. Businesses today want more than just strategic insight from their BI implementations.



**Figure 2.** BI enabling all in the enterprise [1]

There is an expectation from BI to enable better execution of the numerous tactical and operational decisions that enterprise makes every day. For example, a product manager has to decide the discount schedule or pricing decision for a product every day relies on the market price on condition. Since the market has high volatility and noisy environment, tactical and operational decisions are the drivers for day-to-day management of the business at different levels. These decisions had a smaller business impact when it measured in silos as compared to strategic decisions. When the information gathered and integrated, multiple tactical and operational decisions will have higher and significantly help driving business better. BI strategy should embody the approach to enable better decision-making at all levels of the enterprise.

The stock market tendency has a big challenge task due to its high volatility price environment. There have many studies using artificial neural networks (ANNs) in this area. The early days of these studies focused on the application of ANNs to stock market prediction, such as [3-5]. Recent research tends to hybridize several artificial intelligence (AI) techniques [6]. Some researchers tend to include novel factors in the learning process. Kohara et al. [6] Incorporated prior knowledge to improve the performance of stock market prediction. Kim and Han [7] proposed a genetic algorithms approach to feature discretization and the determination of connection weights for ANN to predict the stock price index. They suggested that their approach reduced the dimensionality of the feature space and enhanced the prediction performance.

The main limitations of ANN above can be overcome, a novel intelligent learning algorithm, feature selection with gain ratio and  $k$ -Nearest Neighbour ( $k$ -NN) approach are proposed in this study. We focus on  $k$ -NN techniques because of their ease of interpretability to determine the trend of sales or purchases in stock market warehouse receipt documents that depend on many parameters. The primary objective of this study is to propose computational methods to trace market tendency both for buying or selling processes warehouse receipt document using feature selection gain ratio and  $k$ -NN technique.

## 2. Model Building Process

### 2.1. Feature Selection with Gain Ratio

It is well known that implementing feature selection improves the accuracy of a classifier. The degree of improvement will depend on many factors: the type of classifier, the effectiveness of the feature selection and the quality of the features. In the case of the simple  $k$ -NN classifier, the feature selection deletes noisy features and reduces the feature-space dimension. Moreover, for an ensemble of classifiers, the feature selection can promote diversity among the ensemble members and can improve their local specialization. The potential for an ensemble to be more accurate than its constituent members depends on the diversity among its members [11].

The information gain measure favors attributes with many values over those with few values [12]. Gain ratio (GR) overcomes this problem by introducing an extra term taking into account how the feature splits the data:

$$GR(S, A) = \frac{IG(S, A)}{SI(S, A)}$$

$$SI(S, A) = \sum_{i=1}^d \frac{s_i}{s} \cdot \log_2 \frac{s_i}{s} \quad (1)$$

SI is d subsets of examples resulting from partitioning S by the d-valued feature A. Since the SI term can be zero in some special cases, we define:  $GR(S, A) = IG(S, A)$  if  $SI(S, A) = 0$  for feature A. For the most part, this improvement over IG proves significant in the evaluation presented here.

## 2.2. *k*-NN Based Classifiers

Real-time business analytic for buying or selling transaction on a commodity warehouse receipt system is using a technique of classification in data mining that is *k*-NN. *K*-NN classifiers are instance-based algorithms by taking a conceptually straight forward approach to approximate real or discrete valued target functions. The learning process includes storing the presented data. All instances correspond to points in an n-dimensional space and the nearest neighbors of a given query that already defined regarding the standard Euclidean distance [8]. The probability of a query q belongs to a class c that can be calculated as follows:

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (2)$$

K is the set of nearest neighbors, kc the class of k and  $d(k, q)$  the Euclidean distance of k from q.

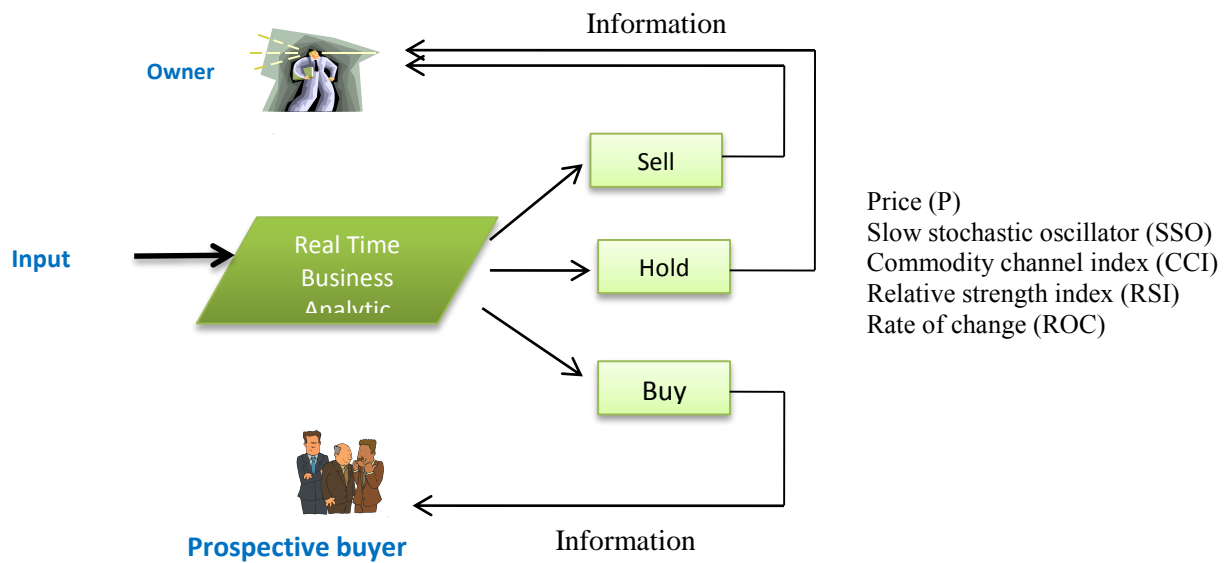
Despite their simplicity, *k*-NN classifiers suffer a serious drawback. The distance between items is calculated based on all the attributes. This implies any features that are in fact irrelevant for classification have the same impact at relevant features. This noise-sensitivity leads to miss-classification problems and degradation in the system accuracy. Such behavior is well known in the literature and is usually referred as a curse of dimensionality [11]. In experimental research section, we will show that this problem affects heavily the classification accuracy of the system. In fact, not only noisy features that affect the classifier accuracy but correlated features may also cause problems.

However, *k*-NN classifiers that used in conjunction with effective feature subset selection techniques are readily interpretable and can provide important insight into a weak theory domain. Black-box classifiers (e.g., neural nets) do not offer the same insight. To overcome the problem of the high dimensional feature space, it is possible to use different strategies. In the next section, we present a set of ensemble methods that we have applied to simplify the decision surface of the *k*-NN deals with. This simplification is obtained in the ensemble members by reducing the number of classes that used to train the *k*-NN or by reducing the feature space dimensionality.

In this study, prediction of decision classification is made to determine the best time to sell, buy or hold the warehouse receipt document by owners or prospective buyers in real time based on transaction history data.

## 2.3. The Model Architecture

The architecture of real-time business analytic for buying or selling transaction in this research illustrated in Figure 3. From the vast amount of information available in an input, real-time business analytic that using feature selection gain ratio and *k*-NN will prediction the best condition for sell, hold or buy transact for the owner or prospective buyer.



**Figure 3.** The architecture of real-time business analytic for buying or selling transaction warehouse receipt document

### 3. Empirical Study

#### 3.1. Research Data

The research data that used in this study is generated from upper and lower data range. Since we attempt to mine the stock index movement tendency, technical indicators are used as input variables. This study selects 18 technical indicators to make up the initial attributes, as being determined by the review of domain expert and prior research [13-15]. The description of initially selected attributes is presented in Table 1.

**Table 1.** Initially selected feature indicators and their formulas

| Feature indicators                 | Formula   |
|------------------------------------|---|
| Price (P)                          | $x_t, (t = 1, 2, \dots, n)$                           |
| Stochastic oscillator (SO)         | $\left( \frac{x_t - x_i(m)}{x_A(m) - x_i(m)} \right)$ |
| Moving stochastic oscillator (MSO) | $\frac{1}{m} \sum_{i=l-m+1}^t (SO_{t-i})$             |
| Slow stochastic oscillator (SSO)   | $\frac{1}{m} \sum_{i=l-m+1}^t (MSO_{t-i})$            |
| Rate of change (ROC)               | $\frac{x_t}{x_{t-m}}$                                 |
| Momentum (M)                       | $x_t - x_{t-m}$                                       |
| Moving average (MA)                | $\frac{1}{m} \sum_{i=l-m+1}^t (X_i)$                  |

Moving variance (MV)

$$\frac{1}{m} \sum_{i=l-m+1}^t (X_i - X_l)$$

Moving variance ratio (MVR)

$$\frac{MVt^2}{MV_{t-m}}$$

Exponential moving average (EMA)

$$a \times x_t + (1 - a) \times x_{t-m}$$

Moving average convergence &amp; divergence (MACD)

$$\sum_{i=l-m+1}^t EMA_{20}(i) - \sum_{i=l-m+1}^t EMA_{40}(i)$$

Accumulation/ distribution oscillator (ADO)

$$((x_l(m) - x_t)/(x_h(m) - x_l(m)))$$

Disparity 5 (D5)

$$\frac{x_t}{MA_5}$$

Disparity10 (D10)

$$\frac{x_t}{MA_{10}}$$

Price oscillator (OSCP)

$$\frac{(MA_5 - MA_{10})}{MA_5}$$

Commodity channel index (CCI)

$$\frac{(M_t - SM_t)}{0.015D_t} \text{ where } M_t = x_h(t) + x(t) + x_l(t), SM_t$$

$$= \sum_{i=l-m+1}^t \frac{M_i}{m}, D_t$$

$$= \sum_{i=l-m+1}^t |M_i - SM_t|/m$$

Relative strength index (RSI)

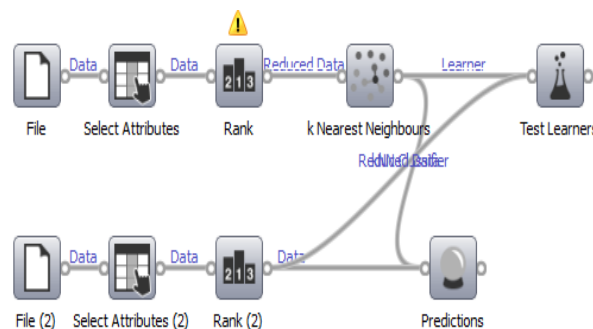
$$100 - \frac{100}{1 + RS} \text{ where } RS = \frac{\sum_{i=l-m+1}^t (x(i) - x(i-1))}{\sum_{i=l-m+1}^t (x(i) - x(i-1))}$$

Linear regression line (LRL)

$$\frac{m \times \sum_{i=l-m+1}^t i \times x(i) - \sum_{i=l-m+1}^t i \times \sum_{i=l-m+1}^t x(i)}{m \times \sum_{i=l-m+1}^t i^2 - (\sum_{i=l-m+1}^t i)^2}$$

### 3.2. Experiment Result

Figure 4 displays the knowledge flow of computational methods to obtain a prediction market trend towards buying or selling a warehouse receipt documents. This knowledge flow and data analysis is using Orange Software.



**Figure 4.** The knowledge flow

From 18 parameters in Table 1, after the sort by using a feature selection gain ratio method obtained five most influential parameters on the prediction of selling and buying warehouse receipt document.

The parameters we consider are price (P), a slow stochastic oscillator (SSO), commodity channel index (CCI), relative strength index (RSI) and rate of change (ROC) as in Table 2.

**Table 2.** Result of feature selection gain ratio

| Attribute                        | Gain ratio |
|----------------------------------|------------|
| Price (P)                        | 0.001      |
| Slow stochastic oscillator (SSO) | 0.016      |
| Commodity channel index (CCI)    | 0.005      |
| Relative strength index (RSI)    | 0.006      |
| Rate of change (ROC)             | 0.009      |

Datasets are used to classify the  $k$ -NN uses five parameters generated in the feature selection stage with the gain ratio

**Table 3.** Result of prediction time to buying or selling using  $k$ -NN

| MSE    | RMSE   | MAE   | RRSE   | R2      |
|--------|--------|-------|--------|---------|
| 4.9617 | 2.2174 | 2.083 | 1.4284 | -1.0403 |

Each of the models described in the last section is estimated and validated by in sample data. The model estimation selection process is then followed by an empirical evaluation based on the out-of-sample data. At this stage, the relative performance of the models is measured by MSE, RMSE, MAE, RRSE and R2. Table 3 reports the experimental results. The results show that the prediction means square error value of 4.9617 which means that the resulting level of accuracy rate is 95.03%.

#### 4. Conclusions

This study uses a feature selection method of computing the ratio of profits and  $k$ -NN to track trends in the market either to buying or selling transaction process. Regarding empirical results, we find that the whole forecasting model can be used as an alternative solution to buy or sell a stock market trend of warehouse receipt documents with accurate rate is 95.03%.

#### References

- [1] Pant, P 2009 Business intelligence (BI) How to build successful BI strategy. Deloitte Consulting LLP 1-17.
- [2] Kamijo, K., Tanigawa, T 1990 Stock Price Pattern Recognition: A Recurrent Neural Network Approach. In: *Proceedings of the International Joint Conference on Neural Networks*, San Diego, CA 215-221.
- [3] Yoon, Y., Swales, G 1991 Predicting Stock Price Performance: A Neural Network Approach. In: *Proceedings of the 24th Annual Hawaii International Conference on System Sciences*. Hawaii 156-162.
- [4] Trippi, R.R., DeSieno, D 1992 Trading Equity Index Futures with a Neural Network. *Journal of Portfolio Management* 19 309-317.
- [5] Witten, I.H., Frank, E 2000 Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann Publishers, San Francisco, CA.
- [6] Kohara, K., Ishikawa, T., Fukuhara, Y., Nakamura 1997 Stock Price Prediction Using Prior Knowledge and Neural Networks. *International Journal of Intelligent Systems in Accounting, Finance and Management* 6 11-22.
- [7] Tsaih, R., Hsu, Y., Lai, C.C 1998 Forecasting S&P 500 Index Futures with a Hybrid AI system. *Decision Support Systems* 23 161-174.
- [8] Quah, T.S., Srinivasan, B 1999 Improving Returns on Stock Investment through Neural Network Selection. *Expert Systems with Applications* 17 295-301.



- [9] Kim, K., Han, I 2000 Genetic Algorithm Approach to Feature Discretization in Artificial Neural Networks for the Prediction of Stock Price Index. *Expert Systems with Applications* 19 (2000) 125-132.
- [10] Zhang, G., Patuwo, B.E., Hu, M.Y 1998 Forecasting with Artificial Neural Networks: the State of the Art. *International Journal of Forecasting* 14: 35-62.
- [11] T M. Mitchell 1997 Machine Learning, McGraw-Hill.
- [12] Kupinski, A.M., Giger, M.L 1999 Feature selection with limited datasets. *Medical Physics* 26: 2176-2182.
- [13] Hann, T.H., Steurer, E 1996 Much Ado about Nothing.Exchange Rates Forecasting: Neural Networks vs. Linear Models Using Monthly and Weekly Data. *Neurocomputing* 10: 323-339
- [14] Yu, X.H.: Can Backpropagation Error Surface not have Local Minima.
- [15] G. Zenobi, P. Cunningham 2001 Using Diversity in Preparing Ensemble of Classifiers Based on Different Subsets to Minimize Generalization Error, *12<sup>th</sup> European Conference on Machine Learning (ECML 2001)*, L. De Raedt & P. Flach (Ed.), 576-587, Springer Verlag, 2001.
- [16] Orange Software. University of Ljubljana. Website : [orange.biolab.si](http://orange.biolab.si).