Analysis of Data Mining for Forecasting Total Goods Delivery with Moving Average Method

(Case Study: Agent "X" Expedition "Z")

M. Azman Maricar[1], Putu Widiadnyana[2], and Wayan Arta Wijaya[3]
[1][2] Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[3] Department of Electrical and Computer Engineering, Udayana University
E-Mail: m.azman.maricar@gmail.com

***Abstract*—In the logistics and distribution of goods, the expedition service is necessary, because the expedition is an important part of a business that has a strong attachment to the distribution. The number of deliveries from an expedition per period is uncertain, sometimes the number increases or decreases. This may result in an imbalance between existing facilities and employees and the number of shipments from customers or company policies. To overcome this, required forecasting techniques that are able to predict total shipments, as well as predict which goods and products are the most widely sent. The moving average method using the last 5 period data is used as a way of forecasting. MAPE (Mean Absolute % Error) is used as a test method, and a result of 34 %, indicates that the method is feasible to use.**

Keywords—Data Mining; Forecasting; Moving Average; MAPE (Mean Absolute % Error); Expedition; Goods; Product.

I. INTRODUCTION (*HEADING 1*)

In the logistics and distribution of goods, the expedition service is necessary, because the expedition is an important part of a business that has a strong attachment to the distribution, with the expedition is considered capable to facilitate the movement of goods efficiently by utilizing the speed and accuracy in the import and export process on Trade of national and international nature. Currently there are many expedition companies, which indirectly force shipping companies to have advantages over other companies [1,2]. The number of deliveries from an expedition per period is uncertain, sometimes the number increases or decreases, caused by the competition. This may result in an imbalance between existing facilities and employees and the number of shipments from customers or company policies.

To overcome the uncertainty of total shipment per period, a forecasting technique is needed, in order to predict the total shipment which certainly affects the company policy in providing the facility, and is expected to overcome the imbalance between the facilities provided by the company and the amount of shipment from the customer. Moving averages are offered as a method for predicting them. Moving averages is one of the methods in forecasting time series by using a data set based on time, the way it works is averaging by utilizing some data of the last period [3].

Tests on forecasting results using moving average method will use MAPE (Mean Absolute % Error) which is one of the methods of accuracy testing by calculating the average %age of error. MAPE gives a clue how much a forecast error is compared to the actual value [4].

II. LITERATURE REVIEW

A. Expedition

In the logistics and distribution of goods, the expedition service is necessary, because the expedition is an important part of a business that has a strong attachment to the distribution, with the expedition is considered capable to facilitate the movement of goods efficiently by utilizing the speed and accuracy in the import and export process on Trade of national and international nature. Currently there are many expedition companies, which indirectly force shipping companies to have advantages over other companies [1,2].

B. Data Mining

The Data Mining is a process of finding relationships, patterns, and trends and has meaning by sorting or observing very large data, previously stored in storage media, using techniques such as statistics and math, to gain new knowledge and benefits [5 , 6,7,8]. Data Mining is known as Knowledge Discovery in Database (KDD) [7,8]. Data Mining is not a new field of science, even through data mining can provide many aspects and techniques for other fields of science [5]. The purpose of data mining is to refine the traditional ways, so that current data mining is able to handle [5]:

1. The amount of data is very large, the quantity of data is very much.
2. Very high data dimensions.
3. Data sets that have diverse properties.

C. Forecasting

Forecasting is the science used to predict a thing or value that has not happened and has a purpose to predict something

that will happen in the future [3,9]. Forecasting is important and applicable to business and industry, government, economics, medicine, social, politics, environment, accounting, and others [9,10]. From that definition, forecasting has a procedure, namely [10,11]:

1. Problem Analyze
2. Collecting Data
3. Analyze the data collected
4. Selecting the appropriate method
5. Testing of selected methods
6. Using the method
7. Monitor performance of forecasting methods.

Forecasting is classified into two general methods namely qualitative and quantitative which have their respective properties. Qualitative has an intuitive nature and does not have data of the past, so can not be solved by mathematical means, because more use of certain opinions, contrary to the quantitative that has Previous data, and can be calculated by mathematical [9]. Quantitative methods are used more often than qualitative, which is better known as time series, whose data form accumulated over a period of time. Time series has four important components namely, Secular Trends, Seasonal Variations, Cycle Variations, and Ireguler Variations [9]. The use of time series, tends to be used to predict the future as well as made using the detailed data sets generated in the past [12,13].

D. Moving Average

Moving Average is a method that uses a collection of actual data that existed before to produce a forecasting calculation in the future [3]. The moving average formula can be seen in formulation (1) [10].

$$F(x)=Ax\text{-}1+Ax\text{-}2+Ax\text{-}3+ \dots Ax\text{-}n / n \quad (1)$$

That is :
Fx: Forecasting in period X
Ax-1, Ax-2, Ax-3 ... Ax-n: Actual data in the period before the period to be calculated
N: number of periods used

E. Testing Accuracy

Not all methods of forecasting can be used for all cases, it necessitates the need for the testing process. Some ways to test forecasting error rates are as follows :

1) Mean Absolute Deviation (MAD): Mean Absolute Deviation is a calculation process used to calculate the absolute error between the true value and the value of the depth [4]. The formulation is shown in the formula (2) [10].

$$MAD=\sum |Actual - Forecast| / n \quad (2)$$

That is :
$\sum$ | Actual - Forecast |: The sum of all actual values minus the value of forecasting (in absolute).
N: number of periods used.

2) Mean Square Error (MSE): Mean Square Error is a calculation process used to calculate the error of rank between the true value and the depth value [4]. The formulation is shown in formula (3) [10].

$$MSE=\sum (Actual - Forecast)^2 / n \quad (3)$$

That is :
$\sum (Actual - Forecast)^2$: The sum of all actual values minus the value of forecasting (in quadratic).
N: number of periods used.

3) Mean Absolute % Error (MAPE): Mean Absolute % Error is a calculation process used to calculate the absolute average error between the true value and the depth value [4]. The formulation is shown in formula (4) [10].

$$MAPE=\sum (|Actual - Forecast|) / Actual / n * 100\% \quad (4)$$

That is :
$\sum$ (|Actual – Forecast|) / Actual: The sum of all actual values minus the forecasting value (in absolute) and divided by its actual value.
N: number of periods used.

MAPE has a standard assessment of the results of accuracy testing performed and gives a hint as to how much error the forecasting process takes with a particular method[14] as shown in table 1 [9].

TABLE I. RANGE OF MAPE

Range	Meaning
<10%	The ability of method is very good
10-20%	The ability of method is good
20-50%	The ability of method is feasible
>50%	The ability of method is bad

III. METHODOLOGY

A. Interview

Interviews were conducted to obtain the required data from Agent X Expedition Z, and carried out directly.

B. Literature Review

Literature review was conducted to obtain data and information related to this research, through various sources.

C. Calculation of Forecasting

The calculation of forecasting of the total delivery agent X expedition Z will use the moving average method.

D. Testing Accuracy of Forecasting

Forecasting accuracy testing using moving average method in the case of total delivery agent X expedition Z, will use MAPE (Mean Absolute % Error) method. The smaller the %age of errors, the better the forecasting method [15].

IV. Result Analysis

Table 1 shows the total shipment per period, table 2 shows the number of items sent per period, while table 3 shows the product used at most per period.

TABLE II. TOTAL SHIPMENT

Period	Total Shipment
1	127
2	111
3	91
4	76
5	154
6	59
7	207
8	200
9	135
10	183
11	219
12	285

TABLE III. TYPE OF GOODS

Period	Cloth	Food	Electronic	Document
1	39	26	5	27
2	33	36	2	40
3	20	18	3	50
4	19	19	7	31
5	26	64	9	55
6	11	23	2	23
7	55	63	12	77
8	65	56	6	23
9	32	63	7	33
10	46	59	11	67
11	66	78	14	61
12	79	97	20	89

TABLE IV. TYPE OF PRODUCT

Period	A	B	C
1	23	34	70
2	24	35	52
3	12	30	49
4	25	22	29
5	29	35	90
6	7	11	41
7	26	64	117
8	33	47	120
9	12	30	93
10	48	61	74
11	53	47	119
12	78	87	120

A. Calculation of Forecasting

This forecasting calculation will use the moving average method with the formula (1) and by using the last 5 period data, is period data 8 to 12.

F(13)=(200+135+183+219+285) / 5

F(13)=1022 / 5

F(13)=204.4

F(13)=204

The calculation is also used to calculate type of goods and type of product, and get results for type of goods for period 13, shown on figure 1:

1. Cloth : 58
2. Food : 71
3. Electronic : 12
4. Document : 65

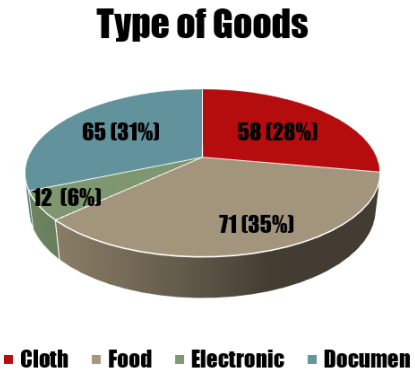

Fig. 1. Result of Forecasting for Type of Goods

While the results obtained for type of product for period 13, shown on figure 2:

A : 45
B : 54
C : 105

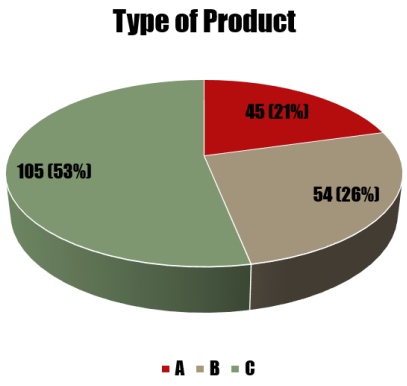

Fig. 2. Result of Forecasting for Type of Product

B. Testing Accuracy

For accuracy testing, sample data will be used for data in period 12. To obtain data forecasting in period 12, will be calculated forecasting using the last 5 period data that is period 7 to 11. The process of calculation is the same as forecasting calculation. From the 13th period using the moving average method. And the forecast for the 12th period is 189. Next, an accuracy test will be performed using Mean Absolute % Error (MAPE) with the formula (4).

MAPE=∑(|285 - 189|) / 285 * 100 %

MAPE=∑|96| / 285 * 100 %

MAPE= 0.34 * 100 %

MAPE= 34 %

V. CONCLUSION

From result of forecasting by using moving average 5 period got result that total delivery period 13: 204 decreased compared to actual value period 12. Type of Goods: cloth, food, electronics, document decreased compared to actual value of period 12. Product: A , B, C, decreased compared to the actual value of period 12. Accuracy of forecasting with sample data of 12 is 34 %. Refers to table 1, the result is declared feasible for used.

REFERENCES

[1] Natasya Yosepha Samapaty. Strategi Pengelolaan dan Pengembangan Usaha Ekspedisi Barang Antar Pulau Pada PT. Bumi Indah Lines di Surabaya (Strategy of Management and Business Expansion of Inter-island Goods Expedition at PT. Bumi Indah Lines in Surabaya). AGORA, 2015; Vol.3 No.2 : p.578-587.

[2] Rachmat Hidayat. Sistem Informasi Ekspedisi Barang dengan Metode E-CRM Untuk Meningkatkan Pelayanan Pelanggan (Information System of Goods Expedition with E-CRM Method To Improve Customer Service). JURNAL SISFOTEK GLOBAL, September 2014; Vol.4 No.2 : p.41-45.

[3] Ni Luh Ayu Kartika Yuniastari and IGP Wirarama Wedashwara Wirawan. Peramalan Permintaan Produk Perak Menggunakan Metode Simple Moving Average Dan Exponential Smoothing (Forecasting Silver Product Demand Using Simple Moving Average And Exponential Smoothing Method). Jurnal Eksplora Informatika, November 2012; Vol.9, No.1 : p.97-106.

[4] Ni Kadek Sukerti. Peramalan Deret Waktu Menggunakan S-Curve Dan Quadratic Trend Model (Forecast Time Series Using S-Curve And Quadratic Trend Models). Konferensi Nasional Sistem dan Informatika, 2015; p.592-597.

[5] David Hartanto Kamagi and Seng Hansun. Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa (Implementation of Data Mining with Algorithm C4.5 to Predict Students Graduation Level). ULTIMATICS, Juni 2014; Vol.6, No.1 : p.15-20.

[6] Irwan Budiman and Artesya Nanda Akhlakulkarimah. Aplikasi Data Mining Menggunakan Multiple Linear Regression Untuk Pengenalan Pola Curah Hujan (Data Mining Applications Using Multiple Linear Regression For Recognize of Rainfall Pattern). Kumpulan Jurnal Ilmu Komputer (KLIK), Februari 2015; Vol.2, No.1 : p.34-44.

[7] Derick Iskandar and Yoyon Suprato. Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan Antara Algoritma C4.5 dan Naive Bayes (Comparison of Classification Accuracy of Poverty Rate Between Algorithms C4.5 and Naive Bayes). Jurnal Ilmiah NERO, 2015; Vol.2, No.1 : p.37-43.

[8] Heri Susanto and Sudiyatno. Data Mining Untuk Memprediksi Prestasi Mahasiswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan, dan Prestasi Masa Lalu (Data Mining To Predict Student Achievement Based on Socio-Economic, Motivation, Discipline, and Past Achievement). Jurnal Pendidikan Vokasi, Juni 2014; Vol.4, No.2 : p.222-231.

[9] Amira Herwindyani Hutasuhut, Wiwik Anggraeni, and Raras Tyasnurita. Pembuatan Aplikasi Pendukung Keputusan Untuk Peramalan Persediaan Bahan Baku Produksi Plastik Blowing Dan Inject Menggunakan Metode ARIMA (Auto regressive Integrated Moving Average) Di CV. Asia (Making Decision Support Approval For Forecasting Inventory of Plastic Raw Materials Production Blowing And Inject Using ARIMA Method (Auto regressive Integrated Moving Average) At CV. Asia). Jurnal Teknik Pomits, 2014; Vol 3, No. 2 : p.A-169-A-174.

[10] Douglas C.Montgomery, Cheryl L. Jennings, and Murat Kulahci. 2015. Introduction To Time Series Analysis And Forecasting. Canada: John Wiley And Sons, Inc.

[11] Ana Corber'an-Vallet, Jos'e D.Berm'udez, and Enriqueta Vercher. Forecasting Correlated Time Series With Exponential smoothing models. International Journal Of Forecasting, 2011; p.252-265.

[12] Liang-Ying Wei. A Hybrid ANFIS Model Based On Empirical Mode Decomposition For Stock Time Series Forecasting. Applied Soft Computing, 2016.

[13] Shuhaida Ismail and Ani Shabri. Time Series Forecasting using Least Square Support Vector Machine for Canadian Lynx Data. Jurnal Teknologi (Sciences and Engineering), 2014 : p.11-15.

[14] Jimmy Handoko Barus. Analisis Peramalan Ekspor Indonesia Pasca Krisis Keuangan Eropa Dan Global Tahun 2008 Dengan Metode Dekomposisi (Analysis of Indonesian Export Forecasting After Global and European Financial Crisis of 2008 With Decomposition Method). Jurnal Ekonomi Dan Keuangan, Februari 2013; Vol. 1, No.3 : p.117-133.

[15] Feby Artwodini Mutadiroh, Avia Riska Syofiani, and Terry Safiria Ramadhani. Analisis Peramalan Penjualan Semen Non-Curah (ZAK) PT Semen Indonesia (Persero) Tbk Pada Area Jawa Timur (Analysis of Forecasting of Non-Bulk Cement Sales (ZAK) of PT Semen Indonesia (Persero) Tbk In East Java Area). Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA), 28 Maret 2015; p.308-314.