IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

Predictive Analytic for Blood Request by Using Moving Average and Linear Regression

P.Limsaiprom [1] S. Rattanachan [1] N. Phuangphairoj [2] and M. Kaenchuwongk [3]

[1]National Blood Center, Thai Red Cross Society
[2]34/45 Soi PrachaUthid 72, Thungkru, Thungkru, Bangkok 10140, Thailand
[3] Faculty of Information Technology Sripatum University, Khon Kaen

Corresponding Author: crossprajit@yahoo.com, Tel.: +6681-489-2662

DOI: https://doi.org/10.26438/ijcse/v7i5.13231329 | Available online at: www.ijcseonline.org

Accepted: 10/May/2019, Published: 31/May/2019

Abstract— The main objective of this research predicted the number of blood request in each month of the organization. In this research, we compared 3 statistical estimation methods (simple moving average, exponential moving average, and linear regression) to predict the number of requested blood in each month using historical time series data. Reducing the error of the expected requested blood would be beneficial for planning and decision making to prepare for the target of blood donation. Data used in this experiment, were the number of requested blood components (red blood cells) from 5 years which divided into two periods. The first period, from January 2014 to December 2017, was used as training set for building linear regression models. The second period, from January 2018 to December 2018, was used as test set to evaluate the regression models and to compare with simple moving average and exponential moving average. The results presented that the exponential moving average method had significantly lowered errors than both the simple moving average method and the linear regression method.

Keywords— Requested blood, Simple moving average, Exponential moving average, Linear regression

I. INTRODUCTION

When the world progresses to a big data era, the key goals of business analytics and data sciences are predicted using mathematical models, statistics, and computer science which may use one method or machine learning such as artificial neuron network or data mining to predict the phenomena that we are interested in studying [1], [2], [3]. Therefore, predictive analytics technique or predictive analysis is the method of business analytics and data sciences that is used in prediction, variable of interest, continuous variable such as CRM, sale orders, and temperature [4]. Predictive analytics are a part of advance analytics using prediction events by using data mining, statistic, modeling, machine learning and AI techniques to help predict from the introduction of historical data for a period of time used to describe the risks and opportunities that will occur in the future. Predictive analytics models show the role of variables as score and weight. They can use structure data (age, gender, education, income, etc.) or unstructured data (social media content, call center notes etc.) can be analyzed with these techniques, the organization can stay above the competition by forecasting the future as close to the truth as possible.

This blood transfusion services organization was assigned by the government to have the main duty of supplying blood to be sufficient, safe and the highest quality from blood voluntarily donors, not expecting compensation, to be used to treat patients throughout the country in the form of blood components and blood products. Therefore, the task of managing the target of supplying blood using data analytic is very important and plays an important role in providing adequate blood supply that depend on requested blood. The pain point of this blood transfusion services organization has to estimate the numbers of supplying blood that adequate for blood demand in the country. This issue needed to concern and tried to find the best solution to estimate the numbers of supplying blood for set the target of blood donations as well.

The objective of this research studied 3 statistical estimation methods (simple moving average, exponential moving average, and linear regression) to predict the target of supplying blood in each month that provided the sufficient supplying blood to hospitals.

II. RELATED WORKS

Recently, many researchers are interested to propose the framework or the methods in data analytic to accurate the performance predictive methodology: R. Daosa-nguan et al. set the comparative study of suitable models for forecasting Number of patients with epidemiological surveillance in Bangkok [5]. S. Ronachai et al. studied the forecasting the demand for power distribution units with support vector regression. 3 alternate examination forms [6]. T. Chalermchat et al. set the comparison of forecasting methods for electricity consumption in Nakhon Phanom province [7]. C. Thanaporn et al. studied the forecasting of the services of the customers in retail business. Journal of King Mongkut's University of Technology [8]. P. Komol introduced the forecasting model of manpower suitable for workload: a case study of a branch shop, CP All Public Company Limited. S [9]. Nicha et al. attempted to Forecasting water demand for consumption in the future using Arya model and Garch model [10]. E. Donald applied Dynamic Modals of Residential Water Demand [11]. E.L. Danielson presented an analysis of Residential Demand for water using Micro Time Series Data [12]. S.J. Henry et al. studied Urban Residential Demand for Water in the United States [13]. W. Charles et al. presented the Impact of Price on Residential Water Demand and Its Relation to System Design and Price Structure [14].

III. METHODOLOGY

A. Data preparation

From the database of the request of blood components (Red blood cells) with a total of 60 months, which is the data since the year 2014-2018, the researchers selected the request for blood components (Red blood cells) by dividing the request of blood components (Red blood cells). The first part, from January 2014 to December 2017, is used as training data for regression or regression learning, and part 2 data from January 2018 to December 2018 to test the error of the regression model and comparison with use Simple Moving Average and Exponential Moving Average.

B. Statistical data analysis

When studying the movement of time series from the initial data, it was found that such information does not contain elements of trends and seasons. The analysis in this research compares three methods of time series forecasting, the Simple Moving Average (SMA) method, Exponential Moving Average (EMA) and Linear Regression Analysis (LR). Each method has the following details.

- Simple Moving Average (SMA) method

  A simple moving average method is a method of generating predictions from the average value of the latest data, n numbers, with the same weight of data and does not specify the amount of data that is definitely averaged with the formula to calculate as follows;

$$\hat{y}_{t+1} = SMA_t(n) = \frac{y_t + y_{t-1} + \cdots + y_{t-n+1}}{n} \quad (1)$$

Where y_t is number of request blood components at time t

n is the periods to average (3, 5)

$\hat{y}_{t+1}$ is the forecasting request blood components at time t+1

- Exponential Moving Average (EMA)

  Exponential Moving average method is a method for creating predictive equations from the average value of data by giving importance or weight to information that is closer to the current data than the information that has passed long ago with the formula to calculate as follows;

$$\hat{y}_{t+1} = EMA_t(n) = Y_t.\frac{2}{n+1} + EMA_{t-1}(n).\left(1 - \frac{2}{n+1}\right) \quad (2)$$

Where $EMA_t(n)$ is exponential moving average n priods at time t

- Linear Regression Analysis (LR)

  Linear regression analysis is a statistical method used to study the linear relationship between an independent variable and a dependent variable. In this research used 3 and 5 independent variables, respectively. Linear regression analysis between more than one independent variable and one dependent variable is called only Multiple Linear Regression by the regression model or the regression equation that has the following characteristics;

$$\hat{y}_{t+1} = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \cdots + \beta_n Y_{t-n+1} \quad (3)$$

Where β_0 is constant of regression equation

β_i is regression coefficient of independent variables at i= 1,2, …, n

- Measurement of error in order to evaluate and compare the three statistical methods.

  In this study, the researcher used the number of periods for calculation or model creation for all 3 statistical methods with 3 and 5 periods, respectively, and using Mean Absolute Error: MAE as an error measure to evaluate and compare the 3 statistical methods by the average absolute error value can be calculated using the following equation;

$$MAE = \frac{\sum_{t=1}^{k} |\hat{Y}_t - Y_t|}{k} \quad (4)$$

Where k is the number of periods with predictive testing and error measurement

- The reason that the researchers select to use the average absolute error as a measure of error is to evaluate and compare the three statistical methods when comparing the average absolute error and the Root Mean Square Error: RMSE and found that the root mean square error is usually weighted with the aberration immobilize the higher than the low tolerance, while the average absolute expected value is an average of the absolute error with equal weight without giving extra weight to the high or low moving expectations that meet the objectives of this research.

  In comparing the results of each forecasting method will use the average value of the absolute error, the average of the amount of request of blood components (Red blood cells) and test the hypothesis of the difference in the average of the dependent two populations at the level of significance 0.05, using the statistic T, calculated from the following equation;

$$T = \frac{\bar{d} - d_0}{\frac{S_d}{\sqrt{N}}} \quad (5)$$

Where d_0 is the average difference between the two populations that want to test. The value of this study is 0.

N is the number of request blood components. The value of this study is 60.

D_i is the different number of mean square error of request blood components at time t between the comparing the average absolute error.

IV. Results and Discussion

In this study, the researchers used data on the request for blood components (Red blood cells) in 2014-2018 classified by monthly (units) as shown in Figure 1 and the number of periods for calculation and model creation for all 3 statistical methods including Simple Moving Average (SMA) method, Exponential Moving Average (EMA) method and linear regression analysis (LR) for 3 and 5 periods respectively and using Mean Absolute Error: MAE as a measure of the deviation of the request for blood components (Red blood cells) in each month, as shown in Table 1, to compare the results of each forecasting error will use the average value of the average absolute error of the request of blood components (Red blood cells) in the amount of 60 months. When comparing the number of periods used to calculate or create a model, it was found that the average value of the average error value when using the number of periods is 5 lower when using the number of periods is 3 consistent in all three statistical methods that tested but from the hypothesis testing founded. The only simple moving average method that has a significant difference. Therefore, in comparison of the three statistical methods, it is only used when the number of periods used in the calculation or creation of the model is 5.

From the comparison, it is found that the Exponential Moving Average (EMA) method, present the lowest average of the absolute average error. The hypothesis testing, it was found that the Exponential Moving Average (EMA) method present the absolute error of average, lower than the simple moving average method and significant linear regression analysis methods with significance. In which the Exponential Moving Average (EMA) presents the average value of the absolute error lower than the simple moving average method. The results show that the request for blood components (Red blood cells) in the period close to the current data are factors that influence the forecasting of the request of blood components (Red blood cells) more than the need 0f request of the blood components (Red blood cells) in periods far from current data and for the same reason, the linear regression analysis method presents an average of the average absolute error value over the other two methods because the regression model learns from the training set which has been in use for many years, from January 2014 to December 2017, when using such models to test with the test data set from January 2018 to December 2018, therefore providing a higher error than the other two methods, which uses information that is close to the current in the forecast clearly.

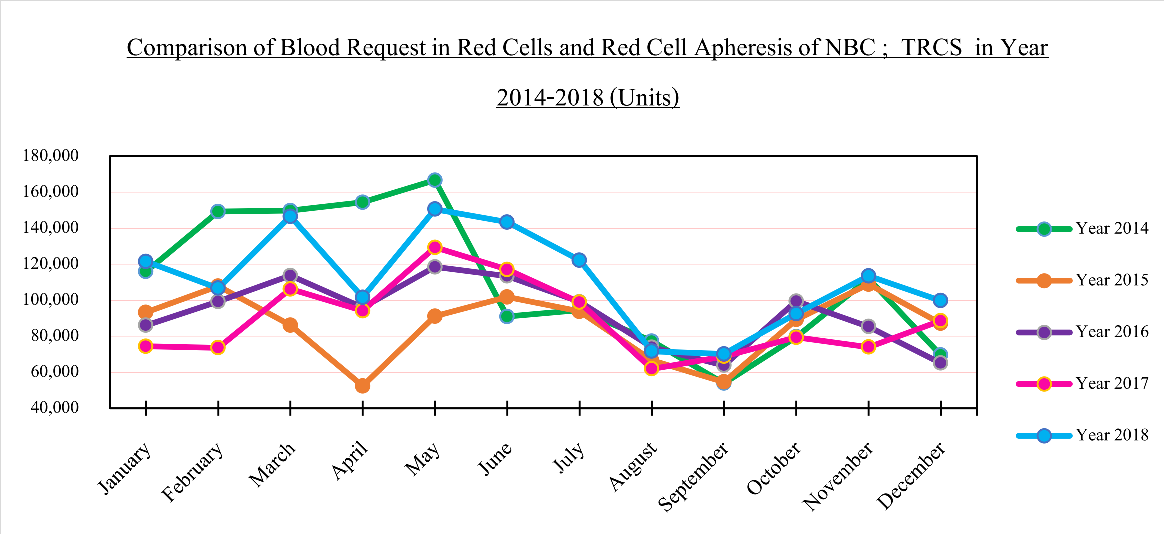

Figure 1. Shows the trend of blood request (red blood cells) year 2014-2018 classified by monthly in each year

Table 1. Shows average absolute expectation value of 3 statistical methods used to predict the blood request (red blood cells) when using the number of periods in the calculation or creating a model in 3 and 5 periods.

Month/Year	SMA		EMA		LR	
	SMA(3)	SMA(5)	EMA(3)	EMA(5)	LR(3)	LR(5)
January 2014						
February 2014						
March 2014						
April 2014	138,273.33		126,273.33		172,773.33	
May 2014	151,115.33		139,115.33		185,615.33	
June 2014	156,934.00	147,178.80	144,934.00	135,178.80	191,434.00	181,678.80
July 2014	137,337.00	142,187.40	125,337.00	130,187.40	171,837.00	176,687.40
August 2014	117,378.00	131,256.40	105,378.00	119,256.40	151,878.00	165,756.40
September 2014	87,556.67	116,748.80	75,556.67	104,748.80	122,056.67	151,248.80
October 2014	75,195.67	96,635.60	63,195.67	84,635.60	109,695.67	131,135.60
November 2014	70,215.33	79,225.20	58,215.33	67,225.20	104,715.33	113,725.20
December 2014	81,782.00	83,415.80	69,782.00	71,415.80	116,282.00	117,915.80
January 2015	87,001.00	78,409.40	75,001.00	66,409.40	121,501.00	112,909.40
February 2015	91,538.67	81,614.40	79,538.67	69,614.40	126,038.67	116,114.40
March 2015	90,164.67	92,397.20	78,164.67	80,397.20	124,664.67	126,897.20
April 2015	95,705.33	93,703.40	83,705.33	81,703.40	130,205.33	128,203.40
May 2015	82,077.67	81,791.80	70,077.67	69,791.80	116,577.67	116,291.80
June 2015	76,507.33	86,101.00	64,507.33	74,101.00	111,007.33	120,601.00
July 2015	81,720.00	87,812.20	69,720.00	75,812.20	116,220.00	122,312.20
August 2015	95,513.67	85,001.20	83,513.67	73,001.20	130,013.67	119,501.20
September 2015	87,339.67	81,081.60	75,339.67	69,081.60	121,839.67	115,581.60

October 2015	71,585.33	81,516.80	59,585.33	69,516.80	106,085.33	116,016.80
November 2015	70,100.00	81,156.80	58,100.00	69,156.80	104,600.00	115,656.80
December 2015	84,198.00	82,568.40	72,198.00	70,568.40	118,698.00	117,068.40
January 2016	95,063.33	81,246.60	83,063.33	69,246.60	129,563.33	115,746.60
February 2016	93,994.00	85,149.40	81,994.00	73,149.40	128,494.00	119,649.40
March 2016	90,792.33	94,092.60	78,792.33	82,092.60	125,292.33	128,592.60
April 2016	99,653.00	98,978.40	87,653.00	86,978.40	134,153.00	133,478.40
May 2016	102,931.33	96,389.40	90,931.33	84,389.40	137,431.33	130,889.40
June 2016	109,344.67	102,661.40	97,344.67	90,661.40	143,844.67	137,161.40
July 2016	109,243.67	108,128.20	97,243.67	96,128.20	143,743.67	142,628.20
August 2016	110,317.33	108,104.40	98,317.33	96,104.40	144,817.33	142,604.40
September 2016	95,451.00	100,140.20	83,451.00	88,140.20	129,951.00	134,640.20
October 2016	78,915.00	93,718.40	66,915.00	81,718.40	113,415.00	128,218.40
November 2016	79,012.67	89,905.20	67,012.67	77,905.20	113,512.67	124,405.20
December 2016	82,835.67	84,295.40	70,835.67	72,295.40	117,335.67	118,795.40
January2017	83,247.33	77,476.40	71,247.33	65,476.40	117,747.33	111,976.40
February 2017	74,896.00	77,572.20	62,896.00	65,572.20	109,396.00	112,072.20
March 2017	70,971.33	79,529.20	58,971.33	67,529.20	105,471.33	114,029.20
April 2017	84,711.33	80,895.60	72,711.33	68,895.60	119,211.33	115,395.60
May 2017	91,311.33	82,657.60	79,311.33	70,657.60	125,811.33	117,157.60
June 2017	109,895.00	95,517.80	97,895.00	83,517.80	144,395.00	130,017.80
July 2017	113,509.67	104,063.80	101,509.67	92,063.80	148,009.67	138,563.80
August 2017	115,079.00	109,122.20	103,079.00	97,122.20	149,579.00	143,622.20
September 2017	92,572.67	100,234.60	80,572.67	88,234.60	127,072.67	134,734.60
October 2017	76,481.67	95,166.00	64,481.67	83,166.00	110,981.67	129,666.00
November 2017	69,982.33	85,174.60	57,982.33	73,174.60	104,482.33	119,674.60
December 2017	74,025.00	76,543.80	62,025.00	64,543.80	108,525.00	111,043.80
January 2018	80,622.33	74,492.00	68,622.33	62,492.00	115,122.33	108,992.00
February 2018	94,679.33	86,438.60	82,679.33	74,438.60	129,179.33	120,938.60
March 2018	105,591.33	94,009.60	93,591.33	82,009.60	140,091.33	128,509.60
April 2018	124,919.33	107,454.20	112,919.33	95,454.20	159,419.33	141,954.20
May 2018	118,255.00	112,976.60	106,255.00	100,976.60	152,755.00	147,476.60
June 2018	132,884.67	125,367.00	120,884.67	113,367.00	167,384.67	159,867.00
July 2018	131,823.33	129,740.60	119,823.33	117,740.60	166,323.33	164,240.60
August 2018	138,725.67	132,857.20	126,725.67	120,857.20	173,225.67	167,357.20
September 2018	112,419.67	117,867.20	100,419.67	105,867.20	146,919.67	152,367.20
October 2018	87,998.33	111,586.60	75,998.33	99,586.60	122,498.33	146,086.60
November 2018	78,083.67	99,976.60	66,083.67	87,976.60	112,583.67	134,476.60
December 2018	92,050.00	94,003.20	80,050.00	82,003.20	126,550.00	128,503.20
Mean= 97,745.85	97,044.35	96,424.27	85,044.35	84,424.27	131,544.35	130,924.27

S.D.=26,962.81	21,718.11	17,629.93	20,225.07	17,003.77	32,708.84	30,929.49

V. Conclusion and Future Scope

In this research, the researchers conducted a comparative study of 3 statistical methods to predict the request of blood components (Red blood cells) that are expected in the future for each month which will be useful for planning and decision- making in preparing blood mobile unit in each month from the test results, it was found that when using the number of periods as 5 in the calculation or creating the model, the average value of the absolute error value is lower than when using the number of periods to be consistent in all three methods. The 3 statistical methods of Simple Moving Average (SMA) method, Exponential Moving Average (EMA) method and Linear Regression Analysis (LR), found that Exponential Moving Average (EMA) method had the average value of the average absolute error lower than both the simple moving average method and the linear regression analysis method with the significance. Suggestions for further study of this problem in the future may be compared with machine learning techniques that can be used for additional estimation, including regression trees, k-Nearest Neighbors and artificial neural networks etc.

Acknowledgment

Thank you to Dr. Ubonwan Jaroonruangrit, the Director of the National Blood Center, Thai Red Cross Society to support the information system of the organization which present the great progress at all times to drive the organization and achieve with strategic plans, policies, goals and objectives of the operations of the Thai Red Cross and the National Blood Center by using data analytics and blood service database that have developed and have deployed in the whole country.

References

[1] K. Saritha and S. Abraham, "*Statistical Predictability in Big Data Analytics with Data Partitioning*", International Journal of computer Science and Engineering, Vol.**06,** Issue.**06**, pp.**80-55**, **2018**.

[2] N.M. Annigeri, S. Shetty, A.P. Patil**,** "*Analysing the supervised learning methods for predicting of healthcare data in cloud environment: A Survey*", International Journal of computer Science and Engineering, Vol.**06,** Issue.**3**, pp.**434-438**, **2018**.

[3] M. Fernandes, "*Data Mining: A Comparative Study of its Various Techniques and its Process, International* Journal of computer Science and Engineering, Vol.**5,** Issue.**1**, pp.**19-23**, **2017**.

[4] J.V.N. Lakshmi, A. Sheshasaayee**,** "*A Big Data Analytical Approach for Analyzing Temperature Dataset using Machinee Learning Techniques*", International Journal of computer Science and Engineering, Vol.**5,** Issue.**3**, pp.**92-97**, **2017**.

[5] R. Daosa-nguan, C. Hansa, S.Maneerat, "*Comparative study of suitable models for forecasting Number of patients with epidemiological surveillance in Bangkok*", Research and Development Journal, KMUTT, Vol.**38,** No.**1**, pp.**33-55**, **2015**.

[6] S. Ronachai, K. Kittisak and K. Nittaya, "*Forecasting the demand for power distribution units with support vector regression. 3 alternate examination forms*", Journal of Science and Technology Ubon Ratchathani University. Vol.**19,** Issue.**1**, pp.**215-232**, **2017.**

[7] T. Chalermchat, "*Comparison of forecasting methods for electricity consumption in Nakhon Phanom province*", Narasuan University Journal: Science and Technology, Vol.**25,** Issue.**4**, pp.**124-137**, **2017.**

[8] C. Thanaporn and C. Paramet, "*Forecasting of the services of the customers in retail business*", Journal of King Mongkut's University of Technology North Bangkok, Vol.**23,** Issue.**3**, pp.**664-674**, **2013.**

[9] P. Komon, "*The forecasting model of manpower suitable for workload: a case study of a branch shop, CP All Public Company Limited*", Panyapiwat Journal. Vol.**5,** Issue.**2**, pp.**103-119**, **20142.**

[10] S. Nicha and K. Sumet, "*Forecasting water demand for consumption in the future using Arya model and Garch model*", KKU Res. J. (be), Vol.**11,** Issue.**1**, pp.**45-55**, **2012.**

[11] E. Agthe,, Donald. and R.B. Billings, "*Dynamic Modals of Residential Water Demand*", Water Resources Research, Vol.**16,** Issue.**June**, pp.**476-480**, **1980.**

[12] E.L. Danielson, "*An analysis of Residential Demand for water using Micro Time Series Data*", Water Resources Research, Vol.**15,** Issue.**August**, pp.**763-767**, **1979.**

[13] S.J. Henry, and R.B. Bruce, "*Urban Residential Demand for Water in the United States*", Land Economics, Vol.**55,** Issue.**February**, pp.**43-58**, **1979.**

[14] W. Charles and F.P. Linaweaver,Jr, "*The Impact of Price on Residential Water Demand and Its Relation to System Design and Price Structure*", Water Resources Resarch, Vol.**3,** Issue.**February**, pp.**13-32**, **1967.**

Authors Profile

Prajit Limsaiprom received the B. Sc in Applied Statistics from King Mongkut's Institute of Technology, Ladkrabang, Bangkok, THAILAND in 1989, the Master Degree in Information Technology in Business (Statistics) from Chulalongkorn University, Bangkok, THAILAND in 2001, and the Ph. D degree in Information Technology from Sripatum University, Bangkok, THAILAND in 2014. She currently is Information Technology Manager, National Blood Center, Thai Red Cross Society, Bangkok, THAILAND. Her research interests in the area of Data Mining Analysis, Information Security Visualization, Social Network Analysis and Social Networks Security. She has recorded in Who's Who in the world in Information Technology.

Sipimya Rattanachan received the B. Sc in Statistics from Srinakharinwirot University. She currently is Information Technology staff, National Blood Center, Thai Red Cross Society, Bangkok, THAILAND. Her research interests in the area of Data Mining Analysis.

Narongrit Phuangphairot received both the B.B.A in Computer Major and also Master Scient Degree in Computer Engineering

Management from Assumption University (ABAC), Ramkamhaeng, Bangkok, THAILAND in 1986 and 1993, respectively and the Ph. D. degree in Information Technology from Sripatum University, Bangkok, THAILAND in 2012. He currently is MD of Rit Solutions Company, Bangkok, THAILAND. His research interests in the area of Enterprise Resource Planning (ERP), Business Intelligent, e-Learning and e-commerce applications, and SEO. He always steps in Software Engineering-Lifecycle Profiles for Very Small Enterprises (VSE) (ISO29110).

Meesaen Kaenchuwongk received the B. Sc in Computer Science from The University of the Thai Chamber of Commerce, Bangkok, Thailaand in 1992, the Master Degree in Business Administration, College of Graduate in Management, Khon Kaen University, Khon Kaen, Thailaand in 2007, and the Ph.D degree Information Study from Khon Kaen University, Khon Kaen University, Khon Kaen, Thailaand in 2007 and 2015, respectively. He currently is Lecturer in Business Computing at Sripatum University, Khon Kaen Campus, Thailand. His research interests in the area of Information Technology, Knowledge Sharing & Transferring, Data Warehouse (MRP & ERP), Data Mining and Value Creation. He has recorded in Who's Who in the world in Information Technology and Knowledge Management.

