International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 11

ISSN: 2454-4248
589 – 593

Regression Based Sales Data Forecasting for Predicting the Business Performance

[1]D. Asir Antony Gnana Singh, [2]E. Jebamalar Leavline, [3]S. Muthukrishnan, [3]R. Yuvaraj
[1]Department of Computer Science and Engineering,
[2]Department of Electronics and Communication Engineering,
[3]Department of Information Technology,
Anna University, BIT Campus, Tiruchirappalli-24 India

Abstract: Business plays a vital role in day-to-day life to bring the goods and services to the people. The profit ofa business highly depends on the sales. Forecasting thesales in business is essential since the sales forecast predicts the business performance.Moreover, sales forecasting is an estimation of futuresales in a business based on the past sales data. This forecasting to make better managerial decisions allows in business for improving the performance of the business. Furthermore, the sales forecasting helps to increase the revenue, reduce the operating cost, improve the working capital use, and increase the shareholder's values. Therefore, this paper presents a sales data forecasting to predict the business performance.

*****

1. Forecasting in Business

Business is an organization or economic system where goods and services are exchanged among one another for making profit. Every business needs some investments and enough customers to make some consistent output of profit. Sales forecasting is the process of estimating the future sales. This forecast gives the strength for the businessmen to manage the workflows, cash flows and resources etc. Accurate sales forecasts help the company to make decisions and to predict the short-term and long-term functions.Forecasting is a planning tool that helps management to predict the future events based on the past and present events. Forecasting starts with certain assumptions based on the management judgments, knowledge and experience. Forecasting is an important for planning and it is the backbone of effective operation of the business.Therefore, this paper presents sales forecasting using regression analysis tool to predict the business trends in order to improve the profit in business by changing the business activities based on the sales forecasting.

2. Types of forecasting

The forecasting is classified into five categories such as naive method, mean (simple average) method, simple moving average method, weighted moving average method and exponential smoothing method. In the naive method, the forecast for the next period (period +1) will be equal to the period of actual demand (A_t). In mean (simple average) method, the forecast for next period (period t+1) will be equal to the average of all historical demands. In the simple moving average method, the forecast of next period (period t+1) will be equal to the average of a specified number of the most recent observation, with each observation receiving the same emphasis (weight). In the weighted moving average method, the forecast for next period (period t+1) will be equal to a weighted average of a specified number of the most recent observation. In the exponential smoothing method, the forecast for the next period (period t) will be calculated as in Equation (1).

3. Literature Review

This section reviews the research works reported in the literature that are related to the presented work. XiaodanYua et al presented a newspaper sales forecasting using support vector regression method. In this approach, the real data is collected and preprocessed using the techniques such as cleaning, reduction, transformation. The support vector regression is used to forecast the sales data [1].Teresa M. McCarthy Byrne a et al presented industries sales forecasting. In this method the forecasting of sales is based on the environmental condition, use of judgment, industry feedback and training [2]. F.L. Chen, T.Y. Ou et al presented a Taguchi method of industry sales forecasting on extreme learning machine. In this method, using a Gray relation analysis mathematics (GRA) method they forecasted industry sales. The GRA has two types namely compared series and reference series which are called as relational coefficient [3].

Tsan-Ming Cho et al. presented a sales data forecasting with limited amount of time and data. In this technique, a strong and versatile artificial intelligence method called as extreme learning machine is used. In order to improve the performance of the sales forecast, they used 3f algorithm [4]. YairOrbach et al presented a product evolution sales forecast. They forecasted their sales based on the performance and availability of the product. Theyalso did

ISSN: 2454-4248
589 – 593

estimation for hybrid sales. The influence of the product is fully based on the relationship between supplier and industries. The final product is obtained based on the technology factor, past performance and roadmaps [5].

Chi-Jie Lu et al presented a computer product sales forecast based on the scheme of variable selection and support vector regression. In this method, they used min max method, normalization method and variable selection method for predicting the sales of each product sales [6]. GauriKulkarnia et al predicted sales of new product based on the online search data. In this method, the real data is collected and processed by finding the product characteristics, customer interests, box office sales and product interests. They distinguished the product by means of searching of particular data before launching and after launching [7].

Matthew J et al presented forecasting the product sales of new and existing products based on the customer reviews and random projection approach. In this approach, based on the online customer reviews and customer interests on the product, they forecasted the product sales. Random projection approach is a method used to forecast the product sales [8].José M. Merigó et al presented a sales forecasting based on the technique of aggregation system. The aggregation system provides summarized results of data. This method takes the arithmetic mean and weighted average. Here they have collected a real data from different countries and predicted the exact output [9].

Zhi-Ping Fan et al presented forecasting of sales data based on the reviews of online of product. In this method, they collected actual data based on the reviews and preprocessed the data using word segmentation and word statistics. They extracted the data and built the forecasting model using base emotional and Norton – emotional model[10]. SangeetaVhatkaraet al presentedforecasting of sales goods using neural network model. In this method, they forecast the data based on the fast moving products that are low cost. In multidimensional space, they classified and analyzed the sales data[11].JulienMostardet al presented a demand forecasting using historical data and expert judgment data. In this method, they have collected the real data and forecast the data based on demand information versus expert judgment. The role ofexpert judgment is to give demand on total sale of data and its range. [12]. NaoufelCheikhrouhou a et al presented demand forecasting for a collaborative process. There are three methods adopted such as mathematical forecasting, event-based fuzzy judgment forecasting and collaborative forecasting with fuzzy judgment. These three methods give a different forecasted data and finally concluded that collaborative forecasting gives an accurate result [13]. From these literatures, it is observed that many approaches and techniques are used to conduct the forecasting. However, this paper presents the regression-based forecasting method for sales forecast.

4. Regression based Sales Data Forecasting

This section presents the least squares based forecasting method to predict the sales data. Table 1 shows the sales data of the business for the period between the 2001 and 2015 and their respective sales data are tabulated. The year is an independent variable since it does not depend on the sales data and the sales is the dependent variable since the sales amount depends on the year variable. Figure 1 depicts the actual sales data of the shop with respect to the year.

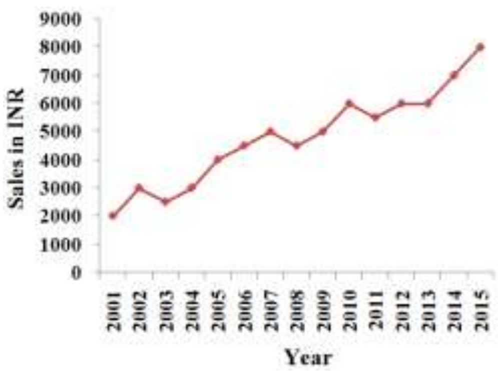

Figure 1 Actual sales data with respect to the year

Table 1 Sales with respect to the year

Year	Sales (in units)
2001	2000
2002	3000
2003	2500
2004	3000
2005	4000
2006	4500
2007	5000
2008	4500
2009	5000
2010	6000
2011	5500
2012	6000
2013	6000
2014	7000
2015	8000

Various calculations from actual sales data is shown in Table 2. These calculations are carried out with the following procedure.

ISSN: 2454-4248
589 – 593

Table 2Actual sales data

X	Y	XY	X^2	$(XY)^2$	Y^2
1	2000	2000	1	4000000	4000000
2	3000	6000	4	36000000	9000000
3	2500	7500	9	5625000	6250000
4	3000	12000	16	144000000	9000000
5	4000	20000	25	400000000	9000000
6	4500	27000	36	729000000	20250000
7	5000	35000	49	1225000000	25000000
8	4500	36000	64	1296000000	20250000
9	5000	4500	81	2025000000	25000000
10	6000	60000	100	3600000000	36000000
11	5500	60500	121	3690250000	30250000
12	6000	72000	144	5184000000	36000000
13	6000	78000	169	6084000000	36000000
14	7000	98000	196	9604000000	49000000
15	8000	120000	225	14400000000	64000000
$\sum X = 8$	$\sum Y = 4800$	$\sum XY = 45266.669$	$(\sum X) = 82.66$	$\sum(XY)$^2=3.142058339	$\sum Y^2$ =25733

The correlation coefficient r is calculated as in Equation (1) where S_{xx} is the sum of the squares of the difference between each xx and the mean xx value and S_{yy} is the sum of the squares of the difference between each xx and the mean yy value

$$r = \frac{S_{xy}}{\sqrt{S_{xx} . S_{yy}}} \qquad (1)$$

S_{xy} is sum of the product of the difference between xx its means and the difference between yy and its mean and it is calculated using Equation (2).

$$S_{xy} = n(\sum xy)\text{-}\sum x \sum y \qquad (2)$$

=678999.99 – 38400

=640599.99

S_{xx} is the sum of the squares of the difference between each xx and the mean xx valueand it is calculated using Equation (3).

$$S_{xx} = n(\sum x^2) - (\sum x)^2 \qquad (3)$$

=1239.99-64

=1175.99

Similarly S_{yy} is the sum of the squares of the difference between each xx and the mean yy valueand it is calculated using Equation (4).

$$S_{yy} = n(\sum x^2) - (\sum y)^2 \qquad (4)$$

=386000000-23040000

=362960000

From these values, correlation coefficient r is calculated using Equation (1)

$$r = \frac{640599.99}{\sqrt{1175.99 * 362960000}} = \frac{640599.99}{\sqrt{4.2683733011}} = \frac{640599.99}{653327.88}$$

r = 0.9805

$$y = mx + c \qquad (5)$$

The linear equation is of the form y = mx + c as in Equation (5) where c is constant, y is intercept, m is the slope of the straight line, y is the dependent variable and x is the independent variable. In this equation where y is the equation of the straight line, m is the slope or gradient andc is the y intercept where the line crosses the y axis. The constant c and the slope n are calculated as in Equation (6) and Equation (7).

$$c = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \qquad (6)$$

where $\sum\sum xy$ isSum of the products of paired values, $\sum\sum x$ is the sum of x values, $\sum\sum y$ is the sum of y values, $\sum\sum x^2$ is the sum of squared x values,$\sum\sum y^2$ is the sum of squared y values and n is the number of values.

$$m = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} (7)$$

$$m = \frac{396796.8 - 38400}{1239.99 - 64} = \frac{358396.8}{1175.99} = 304.761$$

m = 305

With c = 545, the forecast values can be calculated as follows.

When x=16; y = 305(16) + 545=5425

When x=17; y = 305(17) + 545=5730

When x=18; y = 305(18) + 545=6035
When x=19; y = 305(19) + 545=6340
When x=20; y = 305(20) + 545=6645
When x=21; y = 305(21) + 545=6950
When x=22; y = 305(22) + 545=7255
When x=23; y = 305(23) + 545=7560
When x=24; y = 305(24) + 545=7865
When x=25; y = 305(25) + 545=8170
When x=26; y = 305(26) + 545=8475
When x=27; y = 305(27) + 545=8780
When x=28; y = 305(28) + 545=9085
When x=29; y = 305(29) + 545=9390
When x=30; y = 305(30) + 545=9695

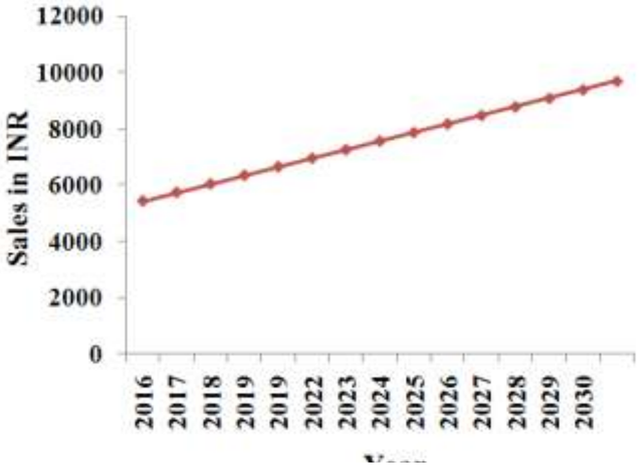

Figure 2 Predicted sales data forecasted with respect to the year

Table 3 Forecasted sales data

X	Y	XY	X^2	Y^2
16	5425	86800	256	29430625
17	5730	97410	289	32832900
18	6035	108630	324	36421225
19	6340	120460	361	40195600
20	6645	132900	400	44156025
21	6950	145950	441	48302500
22	7255	159610	484	52635025
23	7560	173880	529	57153600
24	7865	188760	576	61858225
25	8170	204250	625	66748900
26	8475	220350	676	71825625
27	8780	237060	729	77088400
28	9085	254380	784	82537225
29	9390	272310	841	88172100
30	9695	290850	900	93993025
$\sum X = 23$	$\sum Y = 7760$	$\sum XY = 54180331.67$	$(\sum X) = 547.666$	$\sum Y^2$ =54180331.67

Table 3 shows the forecast data for x=16 to x=30 and the same is plotted in Figure 2. For the predicted sales data the correlation coefficient is calculated using Equation (1) to Equation (4) as follows.

$S_{xy} = n\,(\sum xy)$ - $\sum x \sum y$ = =2693599.995 – 178480=2515119.99

$S_{xx} = n\,(\sum x^2) – (\sum x)^2$ = 8214.99-529 = 7685.99

$S_{yy} = n\,(\sum x^2) – (\sum y)^2$ = 812704975.1-60217600 = 752487375.1

Therefore, r is calculated as

$$r = \frac{2515119.99}{\sqrt{5.78361044 * 362960000}} = \frac{2515119.99}{\sqrt{5.7836104E12}} = \frac{2515119.99}{2404913.803} =$$
1.0458

So, r = 1.1. Now, the slope is calculated using Equation (7) as

$$m = \frac{(\sum y)\,(\sum x^2) - (\sum x)\,(\sum y)}{n\,(\sum x^2) - (\sum x)^2} = \frac{396796.8 - 38400}{1239.99 - 64} = \frac{358396.8}{1175.99}$$

= 304.761

5. Conclusion

Sales forecasting is an estimation of feature sales in a business based on the past sales data. This forecasting allows in business to make better managerial decisions for improving the performance of the business. Furthermore, the sales forecasting leads to increase the revenue, reduces the operating cost, improve the working capital use, and increase the shareholder's values. This paper presented a regression based sales forecasting for predicting the business performance. In order to carry out the experiment, the sales data are collected for the years from 2001 to 2016. The regression analysis is carried out on the sales data and the sales prediction is carried out for the years from 2016 to 2030.

Reference

[1] XiaodanYua, ZhiquanQib, YuanmengZhaoc, Support Vector Regression for Newspaper/Magazine SalesForecastingProcedia Computer Science, vol. 17, 2013, pp. 1055 – 1062
[2] Teresa M. McCarthy Byrne a, Mark A. Moon b, John T. Mentzer b, Motivating the industrial sales force in the sales forecasting process Industrial Marketing Management, vol. 40, 2011, pp. 128–138
[3] F.L. Chen, T.Y. Ou , Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry Expert Systems with Applications, vol. 38, 2011, pp. 1336–1345
[4] Tsan-Ming Choi, Chi-Leung Hui, Na Liu, Sau-Fun Ng, Yong Yu, Fast fashion sales forecasting with limited data and time Decision Support Systems, vol. 59, 2014, pp. 84–92
[5] YairOrbach ,GilaE.Fruchter Forecasting sales and product evolution: The case of the hybrid/electric car Technological Forecasting & Social Change, vol. 78, 2011, pp. 1210–1226
[6] Chi-Jie Lu, Sales forecasting of computer products based on variable selection scheme and support vector regression Neurocomputing, vol. 128,2014, pp. 491–499
[7] GauriKulkarni a , P.K. Kannan b, Wendy Moe b, Using online search data to forecast new product sales Decision Support Systems, vol. 52, 2012, pp. 604–611
[8] Matthew J. Schneider a 1, Sachin Gupta b, Forecasting sales of new and existing products using consumer reviews: A random projections approach International Journal of Forecasting, vol.32, 2016, pp. 243–256
[9] José M. Merigó a, Daniel Palacios-Marqués b, BelénRibeiro-Navarrete c, Aggregation systems for sales forecastingJournal of Business Research
[10] Zhi-Ping Fan, Yu-JieChe, Zhen-Yu Chen, Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis, Journal of Business Research, vol. 74, 2017, pp. 90-100
[11] SangeetaVhatkara , Jessica Diasb, Oral-Care Goods Sales Forecasting Using Artificial Neural Network Model, Procedia Computer Science, vol. 79, 2016, pp. 238 – 243
[12] JulienMostard a , Ruud Teunter b, René de Koster a, Forecasting demand for single-period products: A case study in the apparel industry, European Journal of Operational Research, vol. 211, 2011,pp.139–147
[13] NaoufelCheikhrouhou a, François Marmier b, Omar Ayadia, Philippe Wieser c, A collaborative demand forecasting process with event-based fuzzy judgements, Computers & Industrial Engineering, vol. 61, 2011, pp.409–421