

DATA MINING TECHNIQUES FOR SALES FORECASTINGS

Mehmet Yasin OZSAGLAM
Necmettin Erbakan University
School of Applied Science
Konya, Turkey
myozsaglam@konya.edu.tr

Abstract— The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. In order to support organizations in planning, data mining techniques are launched due to difficult data evaluation because of large amount of data and developments of information technologies. Data mining is an important management tool, which supports original decisions based on data and increases profitability, innovation and efficiency in resource utilization by producing information from data. Today, companies gains competitive advantage from collecting past data and using for future forecastings. Future estimates are usually based on past data and information. In this paper, the research subject is selected as the data of a Turkish consumer electronics store company whose name is hidden. Two year sales amount data of a consumer electronics was used and grouped as four quarters in a year. Next years first quarter sales are forecasted by using regression equations and naive bayes classifier methods and compared by real sales amounts. Sales forecasts results are near to the real amounts and seasonal factors are really important to some product ranges. In this context, various campaigns and marketing strategies have been proposed for the sales of company products by evaluating the forecast results.

Index Terms— Data mining, sales forecasting, business management.

I. INTRODUCTION

As global competition grows stronger, companies continuously need to come up with new advantages in their business in order to compete and survive in today's markets.[1] Predicting future sales has a heavier role to today's uncertainty and a more rapid changing today's business environment. Making predictions of the future based on past and present data and analysis of trends is the process of forecasting. Estimation of some variable of interest at some specified future date might be commonplace example of Forecasting.

Many techniques are available that can be used in forecasting economic variables. Forecasting techniques can be categorized in two broad categories: quantitative and qualitative. When quantitative information is not quite available then qualitative technique is to be relied upon for getting the required forecasts.[2][3] Qualitative forecasting

techniques are based on the opinion and judgment of consumers, experts; they are appropriate when past data are not available. They are usually applied to intermediate- or long-range decisions and subjective. In qualitative forecasting methods consumer opinions and judgement usually collected by survey datas like Complete enumeration survey, sample survey, end-use method survey. Also experts visions are based on these surveys.[4] Quantitative forecasting models are appropriate to use when past numerical data is available and when it is reasonable to assume that some of the patterns in the data are expected to continue into the future. This methods typically used for short- or intermediate-range forecastings.

Forecasting terms are planned for three time ranges. The short-range forecast usually point outs a period of three months and less. The intermediate-range forecast is points out three months to two years time period. The Long-range forecast is points out up to two years of time period. [2][5] In this paper we use quantative forecasting method for short-range forecasting.

II. METHOD

A. Regression Analysis

Regression analysis is a statistical tool for the investigation of relationships among variables. Usually, the investigator seeks to ascertain the causal effect of one variable upon another. Regression analyse includes many techniques for modeling and analyzing several variables. The focus is on the relationship between a dependent variable and one or more independent variables. Regression analysis helps to understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed.[6] Regression analysis is widely used for prediction and forecasting, also relevant with the field of machine learning.[7] Many techniques for carrying out regression analysis have been developed, simple regression, linear regression, ordinary last squares, non parametric regression methods are widely known ones. In this paper we use linear regression method for forecasting.

Linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variable) denoted X . The case of one explanatory variable is called simple linear regression. In linear regression, data are modeled using linear

predictor functions, and unknown model parameters are estimated from the data. Such models are called linear models.[8]

B. Naive Bayes Classifier

A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature.[9]

Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without believing in Bayesian probability or using any Bayesian methods.[10]

Abstractly, the probability model for a classifier is a conditional model;

$$p(C|F_1, \dots, F_n) \tag{1}$$

In Eq. 1, over a dependent class variable C with a small number of outcomes or classes, conditional on several feature variables F_1 through F_n . The problem is that if the number of features is large or when a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We therefore reformulate the model to make it more tractable. Using Bayes' theorem, we write as Eq. 2;

$$p(C|F_1, \dots, F_n) = \frac{p(C) p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)} \tag{2}$$

In plain English the above equation can be written as Eq.3;

$$posterior = \frac{prior \times likelihood}{evidence} \tag{3}$$

C. Our Method

We classify product groups by Naive Bayes classifier, using their past data which includes sales volume, price, profit, seasonal sales volume. Seasonal sales volume data are taken from the sales numbers of same quarters of 2013 and 2014 years. After classification three main product groups occurred. Each groups forecastings are analysed by linear regression methods. Two year sales amount data of a product groups was used and grouped as four quarters in a year. Next years first quarter sales are forecasted by using regression equations and naive bayes classifier methods and compared by real sales amounts. Method shown in Fig.1

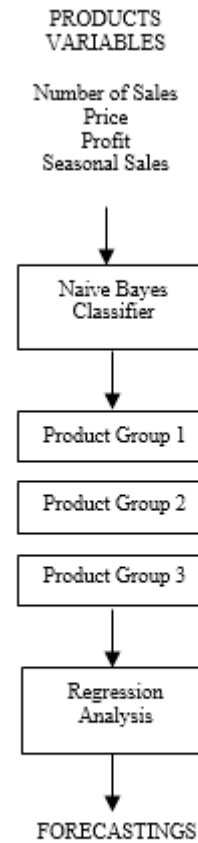


Fig.1 Diagram of our method

III. RESULTS

After naive bayes classification, consumer electronics products are divided into three main groups by probability variables. Sales volume, price of units, profit and seasonal sales volume variables probabilities are calculated and they are categorized by the average value of the probability. If the average (P) is calculated less than 0.4, product is in the Group 3, if average(P) is between 0.4 and 0.7 it is in group 2, and above the 0.7 it is in group 1. Classification results are shown in Table 1.

Sales volume forecasting is made by regression analysis for each groups. In table 2, 2013 and 2014 years first quarter and full year sales are shown. Regression analyse is depending on that datas. Also 2015's first quarters real sales numbers are given and compared with the forecasts. We see that forecasts numbers and real numbers are so close.

TABLE I. CLASSIFICATION RESULTS OF PRODUCTS

Groups	Products	P(sales F)	P(price F)	P(profit F)	P(seasonal F)	Avarage (P~)
Group 1	Tv	0.890	0.970	0.713	0.853	0.856
	Tablet	0.987	0.403	0.698	0.967	0.763
	Smart Phone	0.924	0.625	0.863	0.913	0.831
	Laptop PC	0.622	0.689	0.560	0.416	0.571
Group 2	Desktop PC	0.498	0.624	0.551	0.321	0.498
	House App.	0.701	0.519	0.535	0.498	0.563
	Digital Cameras	0.542	0.512	0.365	0.256	0.418
Group 3	Phone Periph.	0.621	0.206	0.410	0.278	0.378
	PC Periph.	0.243	0.397	0.456	0.234	0.332

TABLE II. REGRESSION ANALYSES OF SALES VOLUME

	2013		2014		2015 First quarter	
	First quarter	Full year	First quarter	Full Year	Real numbers	Forecasts
Group 1	1725	6923	2523	12350	2896	2762
Group 2	2256	8165	2103	7523	2063	2123
Group 3	2024	7632	2412	10123	2226	2461

Regression analyse is used for forecastings of monthly profits of 2015's first quarter. Forecastings results are so close to the real values. Monetary unit of profits are in shown in US dollars. For each products groups, real profits and forecasting graphics are compoared and shown in fig. 2,3 and 4.

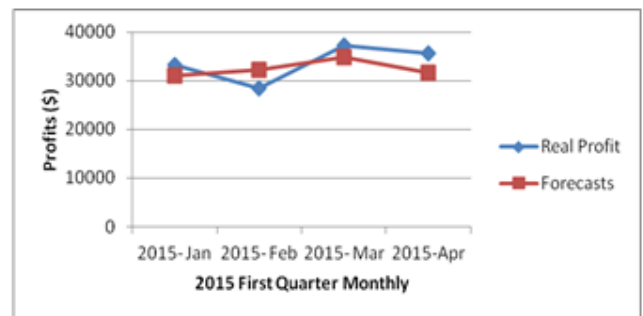


Fig. 2 Group 1 products profits in first quarter of 2015

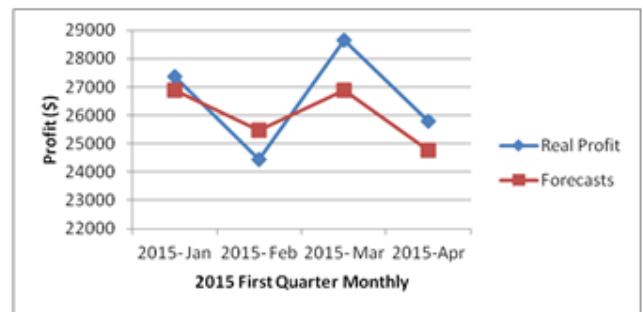


Fig. 3 Group 2 products profits in first quarter of 2015

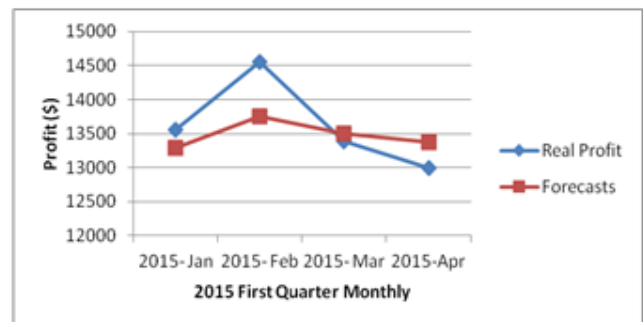


Fig. 4 Group 3 products profits in first quarter of 2015

III. CONCLUSION

In every major functional area of business management, forecasting plays an important role. But forecasting probably is more closely linked to planning and budgeting than to any other key business function. Forecasts are different from plans but supporting element for planning. A forecast is an estimate of what is expected to happen at or by some future period.[11] A plan is what management intends to do about it. Management prepares for change through planning which, in turn, requires making forecasts, establishing goals based on the forecasts, and determining how these goals are to be reached.

In this work short range forecastings are done for sales volumes and profits. Forecasting numbers are fairly close to the real numbers. With these numbers budget, supply and product storing plans can be optimized. To increase profits and sales volumes some selling campaigns can be done by seasonal factors.

REFERENCES

- [1] Karnani A. (2007), The Mirage of Marketing to the Bottom of the Pyramid, Global Competition, The William Davidson Institute at the Michigan State University, p. 90-113
- [2] Kress, G. (1985). Linguistic processes in sociocultural practice. Victoria: Deakin University Press, p. 76-77
- [3] Mentzer J.T. and Kahn K.B, (1995), Journal of Forecasting, [Volume 14, Issue 5](#), p. 465-476
- [4] Greene K.C. and Armstrong J. S. (2007), [The Ombudsman: Value of Expertise for Forecasting Decisions in Conflicts](#), p. 1-12.
- [5] Delurgio S.A. and Bhamu C.D. (1991), Forecasting systems for operations management, p. 648 -649,
- [6] Freedman D.A. (2005), Statistical Models: Theory and Practice, Cambridge University Press, p. 36
- [7] Armstrong, J. S. (2012), [Illusions in Regression Analysis](#), International Journal of Forecasting, p.689-696
- [8] Seal H. L. (1967), [The historical development of the Gauss linear model](#), Biometrika, volume 54, p.1-24
- [9] [Stuart R., Norvig P. \(2003\), Artificial Intelligence: A Modern Approach](#), Prentice Hall, p. 126-132
- [10] Murty N., Susheela D. (2011), Pattern Recognition: An Algorithmic Approach, Springer Publishing, p. 86 -102
- [11] Bednarz T. F., (2011): Sales Forecasting: Pinpoint Sales Management Skill Development Training Series, Majorium Business Press, p.46