Forecasting Automobile Retail Sales Using Data Mining: The Case of Ranchi, Jharkhand, India

[1]Gyaneshwar Mahto, [2]Umesh Prasad, [3]Rajiv Kumar Dwivedi

[1]Dept. of Computer Science under faculty of Science, Vinoba Bhave University, Hazaribag, Jharkhand, India
[2]Dept. of Computer Science, BIT (Mesra) Lalpur Campus, Ranchi, Jharkhand, India
[3]Head University Dept. of Mathematics, Vinoba Bhave University, Hazaribag, Jharkhand, India

Abstract-In this article, sales forecast models for the automobile market are developed and tested. Enhanced sales forecast methodologies and models for the automobile market are presented. The methods used deliver highly accurate predictions while maintaining the ability to explain the underlying model at the same time. The representation of the economic training data is discussed, as well as its effects on the newly registered automobiles to be predicted. Our most important criteria for the assessment of these models are the quality of the prediction as well as an easy explicability.The automobile market are presented for the evaluation of the forecast models. The market demand for vehicles has increased in recent years everywhere in the world. We need suitable models to understand and forecast demand of vehicle. This study presents a singular spectrum analysis (SSA) as a univariate time-series model and vector autoregressive model (VAR) as a multivariate model. Empirical results suggest that SSA satisfactorily indicates the evolving trend and provides reasonable results. The VAR model, which comprised exogenous parameters related to the market on a monthly basis, can significantly improve the prediction accuracy. The vehicles sales in beautiful city Ranchi, Capital of Jharkhand, India, are predicted in both short term (up to December 2018) and long term (up to 2021), as proofs of the growth of the Motor Vehicles industry.

Keywords- Sales Forecast; Automobile Industry; Information Technology; Retail; Decision Making; Data Mining; Business Environment; Retail Sales Forecasting; Vehicles Sales**;**

I. INTRODUCTION

Retail, which in absolutely plain and simple terms is concerned with making available the needed goods and services to consumers has been a concept in existence over times immemorial dating back to last few centuries (Guy, C. 1994),[1].In Ranchi, vehicle emissions have been increased rapidly recent years. Strategic planning based on reliable forecasts is an essential key ingredient for a successful business management within a market-oriented company. This is especially true for the automobile industry, as it is one of the most important sectors in many countries. Reliable forecasts cannot only be based on intuitive economic guesses of the market development. Mathematical models are indispensable for the accuracy of the predictions as well as for the efficiency of their calculations, which is also supported by the increase of powerful computer resources.

Data Mining as a technique for handling large volumes of data by uncovering patterns in the data and aiming towards predicting or forecasting trends of the future has over the few decades been employed in the retail markets with satisfactory results(Dr. Umesh Prasad, 2015),[2].

Forecasting sales and demand over a monthly horizon is crucial for planning the production processes of automotive and other complex product industries. From this point of view, prediction may occasionally be superior to multivariate prediction. On the other hand, the vector autoregressive model (VAR), as a structural relationship identification methodology for multivariate time series, can identify the dynamic couplings between EV sales and economic indicators.

Therefore, the current study proposes models to forecast the quantified market demand for EV. First, an SSA model is applied to build time-series forecasting models. Second, a multivariate VAR is built by considering the external influences from six economic indicators. Finally, comparisons of the models are conducted, and the EV sales are predicted in the short and long terms.

II. DATA AND WORKFLOW

Newly registered automobiles as well as exogenous indicators are considered for both the rural and urban automobile market. In the case of the rural market, which also holds for their units and sources. These data used from, the Mahindra Finance, and Premsons Motors , whereas the

new registrations were taken from the Motor Transport Authority[3]. The feature selection performed is not taken into account here, i.e. all ten exogenous parameters are considered. The reason for this is the fact that the parameter reduction consistently delivered worse results in the case of a non-linear model. For the quarterly model, all exogenous parameters were chosen to be relevant, i.e. no parameter reduction was made. As the non-linear model turned out to be superior to the linear one, it was decided not to perform a feature selection in this work. The enhancements here are based on different approaches. However, it is not excluded that a feature.

III. LITERATURE REVIEW

Vehicle market demand
The penetration can be described with spatially explicit consumer choice model, thereby showing that it could be enhanced by the readily available estimates of the expected lifetime fuel costs, as well as increases in gasoline prices and synergetic gasoline tax. Besides, the sales and range anxiety of could be addressed through parking layout, policy, and regulation . And the market potential for the Jharkhand commercial passenger car sector in 2020. Some other researches focus on demonstrating how policy entrepreneurs influence and accelerate the introduction in local public authorities. Potential of the each models are analogized with the prediction precision, computational cost, and fuel economy. The main conclusion is that the NN-based models show the best overall performance[4]. And different scenarios were built to describe real-life cases and quantify carbon reduction of each scenario. Also, cost-optimal control model was built to maintain the best economy during charging and operation.

In recent years, a more comprehensive indicator was introduced using data displayed in Google Trends. Its performance in forecasting private consumption show that the new indicator derived from Google shows batter performance in comparison to traditional survey-based indicators[5]. There are also some studies utilizing this indicator to describe the online browsing habits and indicate consumer preferences of automobile market. The results are that this indicator not only fits the forecasting model well, but also helps to locate the turning point in sales trends. Similar conclusions were made in the experiments done by some other researchers, showing that searching data in Google Trends have strong predictive value. Therefore, this paper considers the results of previous scholars, focusing on the consumer perspective to research vehicle market demand.

Vehicle sale forecast
Automobile sales forecasting has received significant attention. Most studies have focused on time-series models. Predictions of the automobile sales can be made with

support vector regression and methods alike. Moreover, researches find that by applying data mining algorithms to model the automobile markets, the economic indicators are proved to be relate to sales. Economic indicators related to automobile market include gross domestic product, consumer price index (CPI), interest rate, unemployment rate, and gas prices with automobile sales. To address the relationship among these variables[6]. Other indicators such as subsidies, range, charge point availability, emission rate, and revenue tax are also used for forecasting the future demand for vehicle. Results by different models predict are not the same, but we can see obvious rising trends can be found in all predictions. Forecasting of automobile market is complicated by economics, finance, systems engineering and all kinds of factors, it is difficult to model using the classical tools of market forecasting. In this paper, we predict the vehicle sale from uni-variate and multi-variate methods respectively, and give the most reasonable prediction method and prediction conclusion through comparison.

IV. METHODOLOGIES

Multivariate time-series models may be expected to generate accurate forecasts. However, the univariate forecasting models may considerably outperform the multivariate models in certain conditions, such as scenarios when the prediction process were small. However, some researchers think that the univariate models perform similarly to the multivariate models. Additional comparisons of the two models can be determined in different fields, such as the prediction of emergency department demand and energy market volatility are different[7].SSA is a univariate time-series model and does not rely on a priori defined functions; however, it generates a set of components directly from the time-series under study. Unlike in traditional time-series models, the trend is any gradually varying component of the series that does not contain cyclical or seasonal components. Applying SSA to measure the nonlinear dependency of financial markets shows more accurate than conventional univariate time-series models.

Retail Sales which are an integral part of Information Retail Market which has a wider and more all-encompassing scope for any organization refers to an integrated collection of People, Market, Manufacturers etc and they are responsible for handling the information needs of any organization. Data Mining in Retail plays an extremely crucial role in any retail organization by ensuring that the information needs are met just at the right time to maximize returns and subsequent profitability (Oliver, R. L. 1981),[8].

V. SUMMARY

In summary, this study uses previous research on automobile sales prediction as basis to apply the SSA and

VAR models to forecast the market demand of vehicle by considering the characteristics of the market penetration.This illustrates that the mutual influence of the parameters is essential to achieve accurate forecasts. In order to identify the overall best models, the errors of the training and test sets are accumulated to annual values. The results for the best model based on yearly, monthly, and quarterly data ,the best model for the monthly data is significantly worse compared to both other models. The same holds for the test period. Here, the best quarterly model is significantly superior to the best yearly model, with roughly half of the test error. It can be observed that the quarterly model does not only deliver a minor test error, but at the same time provides higher information content than the yearly model. Both models generate very low errors during the training period, showing again that the set of parameters is well adapted to our problem. The only drawback of the best quarterly model is the fact that all exogenous parameters are necessary, making the model less explicable[9].

VI. Discussion and Conclusion

By looking at the different methods used in our model, answering the question be- comes more difficult. Simple and easily explicable univariate estimations are used for the seasonal, calendar and error component but a more difficult multivariate method for the greatest and most important component, the trend. In general, the chosen model is relatively simple and gives satisfying results in consideration of the quality of the forecast[10]. There can be occurrences (special effects) in the future, which are not predictable or whose effects can not be assessed. The current financial crisis, which led to lower sales in the year 2008, is an example for such an occurrence. It would be interesting to observe the newer technologies that emerge in the future that reshape businesses and their data managements and hence there is an ample scope for further research in the area.

Automobile company must lunch new model with attractive offers during festivals, India is country of Festivals During Dushara and Deepawali automobile sales increase very high. That's why it should be kept in mind to attract customers with attractive offers. The advertisement , offers and easy to finance system may affect sales of vehicle.

References
[1]. Guy, C. (1994). The retail development process: location, property, and planning. Van Nostrand Reinhold.
[2]. Dr. Umesh Prasad and et.al. Exploring The Emerging Role Of Data Mining And Related Technologies In Retail Forecasting: Contextual Issues & The Road Ahead, International Journal of mathematics, Engineering & IT (IRJMEIT) Vol-2,Issue-6 (June 2015),14-18.
[3]. Dudenhöffer, F., Borscheid, D.: Automobilmarkt-Prognosen: Modelle und Methoden. In: Automotive Management. Strategie und Marketing in der Automobilwirtschaft, pp. 192–202 (2004)
[4]. Hechenbichler, K., Schliep, K.P.: Weighted k-Nearest-Neighbor Techniques and Ordinal Classification, Discussion Paper 399, SFB 386, Ludwig–Maximilians University Munich (2004)
[5]. Groth, R. (2000). Data mining: building competitive advantage. prentice Hall PTR.
[6]. Westphal, C., & Blaxton, T. (1998). Data mining solutions: methods and tools for solving real-world problems.
[7]. Berry, M. J., & Linoff, G. (1997). Data mining techniques: for marketing, sales, and customer support. John Wiley & Sons, Inc.
[8]. Oliver, R. L. (1981). Measurement and evaluation of satisfaction processes in retail settings. Journal of retailing.
[9]. Porter, A. L., & Cunningham, S. W. (2004). Tech mining: exploiting new technologies for competitive advantage (Vol. 29). John Wiley & Sons.
[10]. Apte, C., Liu, B., Pednault, E. P., & Smyth, P. (2002). Business applications of data mining. Communications of the ACM, 45(8), 49-53.

Authors Profile:-
Gyaneshwar Mahto
Research Scholar, Department of Computer Science under faculty of Science, Vinoba Bhave University, Hazaribag, Jharkhand, India.

Umesh Prasad
Member of Faculty Department of Computer Science, BIT (Mesra) Lalpur Campus, Ranchi, Jharkhand, India.

Rajiv Kumar Dwivedi
Associate professor And Head Department of Mathematics, Vinoba Bhave University, Hazaribag, Jharkhand, India.