



A STUDY ON THE EFFECT OF SALES FORECASTING ON THE ENTERPRISES

Antonitte Vinoline¹, M. Deepika²

Article History: Received: 30.03.2023

Revised: 13.05.2023

Accepted: 01.07.2023

Abstract

In the present period of time, when advancement is reaching unprecedented heights, every business and organization is struggling to manage both its inventory and client expectations. Sales are the main source of every business, having a general knowledge of what to expect in the future which aids in making smart, strategic sales decisions. Since revenue forecasting is typically the starting point for creating annual budgets for the business, majority of companies are still confused about how to do it. A company's estimation may be negatively impacted over time if its sales projections are regularly off. As a result, sales forecasting has an impact on the entire corporation to enhance their whole growth plan. Sale forecasting is a crucial component of every company's sales operations. An effective sales forecast is crucial for a company to provide the required quantity at the proper time. The predictions are used by executives to plan for enterprise expansion and to evaluate the future performance. In this study, we attempt to forecast the sales of a retailing company by employing the machine learning approaches, Linear regression and Naive Forecasting. A computational example is presented to show the difference between the linear regression and Naïve forecasting methods and we have discovered that the Linear regression produces superior outcomes when compared to the Naïve forecasting approaches. Also we have predicted the sales for the next 5 days by using ARIMA model for the linear regression approach.

Keywords: ARIMA model, Linear regression, Naive Forecasting, Predictions, Sales Forecasting

¹PG and Research Department of Mathematics

²Holy Cross College (Autonomous), Affiliated to Bharathidasan University Tiruchirappalli, India

Email: ¹arulavanto@gmail.com, ²manbudeepi1974@gmail.com

DOI: 10.31838/ecb/2023.12.s2.372

1. Introduction

Due to the rapid expansion of international malls and internet shopping, the competition between multiple retail centers as well as large marts is getting more fierce and violent. Each market tries to entice many consumers by offering customized and limited-time bargains, so that the number of sales for every single product may be calculated for the organization's stock control, delivery, and logistical operations. Every business depends on successful sales, so planning sales is crucial in any retail centre. In order to draw more people, every store or mall tries to offer unique, time-limited discounts. Numerous firms and enterprises monitor the statistical information about their goods in order to forecast consumer demand. This statistical data consists of the number of items sold, their features, and other details about their attributes to offer broad trends to the organization with respect to their distribution, to expand their business, and enhance their marketing plan. This information may be useful in the future if we want to develop periodic special deals to bring more clients to the business. Prior information is important since it helps to handle logistics, stocks, and transportation services along with many other things. A sales forecast assists every company in making more informed business decisions. It contributes to overall business strategy, costing, and risk management. Forecasting is a tool to calculate business income at the product scale, or at the scale of a single company, or at the level of the entire enterprise. Better forecast is always beneficial, both in establishing and upgrading sales techniques for the trade. Forecasting is essential for businesses that are opening up new markets, launching new goods or services, or experiencing quick growth. The major reason an industry creates a forecasts is to to balance marketing and promotional strategy against supply capacity

organizing. Manufacturers and enterprises expect an accurate and reliable assessment of sales information in order to avoid losses brought on by the model's inaccurate or erroneous prediction.

Sales forecasting is primarily used by businesses to evaluate two things. First, to determine the present consumer needs for the product and service. Second, to forecast future need for a firm's products or services. Future sales plans help with higher production use, planning, transportation, and stock control. This, in turn, leads to increased customer satisfaction and a reduction in production costs. Sales predictions have a direct impact on a company's marketing strategy. Before establishing an expenditure and business techniques for the future year, a standard sales assessment can help in completely analyzing the situations or events that occurred previously, and then results can be drawn about client acquisition, insufficient funding, and capabilities. There are numerous obstacles to overcome in sales forecasting, ranging from inaccurate datasets to subjective seller biases that prevent an accurate assessment of how any specific lead would behave in the future. Regardless of the situation, a deep understanding of the past is necessary to enlarge and enhance the chance of industry, particularly the external circumstance, which enables the business to plan for future demands. In order to estimate future sales demand, retailers are conducting a lot of research.

Nowadays, the advanced machine learning algorithms are used for predicting the future of an organization's sales demand. It aids in solving accessibility of inexpensive computing and storage systems. The present machine learning algorithm is very powerful and offers methods for forecasting and estimating sales for any form of organization, which is incredibly helpful in overcoming low-cost prediction techniques. Regression analysis enables enterprises to better understand the

correlations in their data so they can transform it into ideas that can be used to move their business ahead. It is a mathematical approach to determine which parameters may have an impact. The importance of regression analysis for an organization lies in its ability to assist in identifying which factors are most crucial, which can be neglected and how these components relate to one another. This ability is based on the fact that regression analysis seeks an effective statistical approach for assessing the relation between multiple variables of interest, and the data used for this analysis is made up of numbers and figures that characterize your company. Forecasting uses regression analysis to estimate and identify the causal connection among the variables. Understanding the importance of regression analysis is necessary because it enables you to analyze data in order to help you in making better decisions for your business both now and later. The regression method of prediction will assist any firm in acquiring an improved grasp of the variables that can impact its profitability in the upcoming weeks.

Linear regression is important for enterprises because it is used for predicting the future process and strategic planning. Simple linear regression is commonly used in predicting the future and financial reporting. For instance, a company finds that how a transformation in GDP might affect sales. Regression lines are employed in the financial and business sectors. Financial analysts use linear regressions to predict stock prices, commodity prices, and to value a wide range of securities. Linear regression is used in a variety of applications across companies, including trend analysis and customer experience patterns, sales forecasting, retail prices, operational risk, and more. It is frequently used in forecasting consumer behaviour and understanding the buyers, studying trend lines, and advising on which metrics to focus on to achieve the business goals. When examining trends in consumer

purchasing behaviour, linear regression estimates peaks and valleys in demand forecasting, enabling more precise budgeting. This logic can be applied to commercial spending and revenue. A significant advantage of using linear regression for business is that it aids in prioritisation and evaluates the areas of operations to concentrate on your company. It is a simple algorithm that can be studied, analysed, and interpreted by businesses and Machine Learning teams. In comparison to other algorithms, this is a modeling technique whose results can be found quickly in a time crunch.

Some forecasting methods are certainly simple and effective. One of them is the naive forecast. We basically set all forecasts to the value of the most recent observation. The naive forecasting method is the simplest of all and it is appropriate for finance and sales departments because it helps to ensure that these departments work together to improve the enterprise. In this approach, there are no computation or formulas involved, only an assumption of actual sales data. Historical sales data is the most important requirement for Naive forecasting. Salespeople will frequently use naive forecasting to set goals in order to ensure that they are always enhancing or helping to maintain their contribution to the organization. It is one of the best approach for many time series, including most stock price data.

The purpose of our study to create a model that can predict sales and examine their patterns and trends, both of which are essential for running a significant business. The remainder of the paper is subdivided as follows. Section II provides a literature review and outlines the innovative points. Section III summarises the sources of the data for our study. Section IV discusses the Data processing. Section V addresses the evaluation metrics of the study. Section VI describes the model building. Section VII provides the results of descriptive statics of study variables. Section VIII concludes the paper.

Literature review

A quick review of previous research work in sales forecasting is shown in this section. Several authors have done research on sales forecasting and analysis of sales forecasting, which is described below.

J. S. Armstrong [1] stated that the sales forecasts assist investors in choosing whether to invest in present-day businesses. They are crucial to the efficient operation of the company and can assist managers in making decisions about the size of a plant to build, how much of stock to store, the number of workers to hire, the volume of advertising to run, what price to set, and the salaries to pay salespersons. A. S. Weigend and N. A. Gershenfeld [2] analyzed that the sales forecasting is an intricate issue that is affected by both internal and external influences. The statistical technique has two significant limitations. John G. Wacker, Rhonda R. Lummus [11] stated that the foremost important managerial choices an organization can make are based on least precise estimates on forecasts. The second most helpful forecast data for resource planning is least accurate and furthermore business requires most accurate prediction is the largest forecast error. J. Martinovic, V DamnJanovic [6] discussed the qualitative and quantitative techniques in sales forecasting. In comparison to other data mining approaches, Pat Langley and Herbert A [7] noted that Rule Induction (RI) technique is the most recurrent technique used in corporate business.

Data Attributes

Variable	Description
Product ID	Unique product ID
Date	Date when the product is buyed
Customer ID	Individual client ID

Machine learning is the process by which a computer learns from data using statistical or computational methods and processes knowledge that it has acquired through experience [4]. In [3], a variety of machine learning (ML) techniques and their applications in various fields are described, and statistical and computational methodologies are investigated. Paul Goodwin [9] examined the Naive forecasts approach to determine forecast accuracy and the quality of fit of forecasts to time series data. Melvyn Hirst[5] compared the principal types of forecasting models and described a procedure for constructing model using the regression equation approach. RohitSav, PratikshaShinde, SaurabhGaikwad [10] have incorporated predictive models to estimate major market sales. ARIMA was adopted because it can predict using a time series data with different type of structure and autocorrelations between subsequent values in time series. It statistically evaluated and verified consecutive residuals (prediction errors) in the fitted ARIMA time series, and concluded that the selected ARIMA model seems to provide an effective predictive model[8]. In this study, we propose a model using Linear regression and Naive Forecasting to forecasts the sales of a Superstore sales dataset.

Data and Sources of data

The retail dataset of the global superstore is taken from the Kaggle dataset for this study. The dataset includes attributes like:

Customer Name	Username of the client
Segment	Type of buyer
Category	Defines the product
Sub Category	Defines the type of the product
Product name	Name of the product
Sales	The amount which was spend to buy the product

Row ID	Customer ID	Customer Name	Segment	Country	City	State	Product ID	Category	Sub-Category	Product Name	Sales
1	CG-12520	Claire Gite	Consumer	United Sta	Henderson	Kentucky	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.96
2	CG-12520	Claire Gite	Consumer	United Sta	Henderson	Kentucky	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	731.94
3	DV-13045	Darrin Van Huff	Corporate	United Sta	Los Angeles	California	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters by Universal	14.62
4	SO-20335	Sean O'Donnell	Consumer	United Sta	Fort Laude	Florida	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	957.5775
5	SO-20335	Sean O'Donnell	Consumer	United Sta	Fort Laude	Florida	OFF-ST-10000760	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.368
6	BH-11710	Brosina Hoffman	Consumer	United Sta	Los Angeles	California	FUR-FU-10001487	Furniture	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	48.86
7	BH-11710	Brosina Hoffman	Consumer	United Sta	Los Angeles	California	OFF-AR-10002833	Office Supplies	Art	Newell 322	7.28
8	BH-11710	Brosina Hoffman	Consumer	United Sta	Los Angeles	California	TEC-PH-10002275	Technology	Phones	Mitel 5320 IP Phone VoIP phone	907.152
9	BH-11710	Brosina Hoffman	Consumer	United Sta	Los Angeles	California	OFF-BI-10003910	Office Supplies	Binders	Dix Angle-View Binders with Locking Rings by Samsill	18.304
10	BH-11710	Brosina Hoffman	Consumer	United Sta	Los Angeles	California	OFF-AP-10002892	Office Supplies	Appliances	Belkin F5C200VTEL 6 Outlet Surge	114.9
11	BH-11710	Brosina Hoffman	Consumer	United Sta	Los Angeles	California	FUR-TA-10001539	Furniture	Tables	Chromcraft Rectangular Conference Tables	1706.184
12	BH-11710	Brosina Hoffman	Consumer	United Sta	Los Angeles	California	TEC-PH-10002033	Technology	Phones	Konftel 250 Conference Phone - Charcoal black	911.424
13	AA-10480	Andrew Allen	Consumer	United Sta	Concord	North Car	OFF-PA-10002365	Office Supplies	Paper	Xerox 1967	15.552
14	IM-15070	Irene Maddox	Consumer	United Sta	Seattle	Washington	OFF-BI-10003656	Office Supplies	Binders	Fellowes PB200 Plastic Comb Binding Machine	407.976
15	HP-14815	Harold Pawlan	Home Offi	United Sta	Fort Worth	Texas	OFF-AP-10002311	Office Supplies	Appliances	Holmes Replacement Filter for HEPA Air Cleaner, Very Large Room, HEPA Filter	68.81
16	HP-14815	Harold Pawlan	Home Offi	United Sta	Fort Worth	Texas	OFF-BI-10000796	Office Supplies	Binders	Stonex DuraTech Recycled Plastic Frosted Binders	2.544
17	PK-19075	Pete Kriz	Consumer	United Sta	Madison	Wisconsin	OFF-ST-10000486	Office Supplies	Storage	Star-D Stor Shelving Vertical 5-Shelf: 72"H x 36"W x 18 1/2"D	665.98
18	AG-10270	Alejandro Grove	Consumer	United Sta	West Jord	Utah	OFF-ST-10000107	Office Supplies	Storage	Fellowes Super Stor/Drawer	55.5
19	ZD-21925	Zuschuss Donatelli	Consumer	United Sta	San Franci	California	OFF-AR-10003056	Office Supplies	Art	Newell 341	8.56
20	ZD-21925	Zuschuss Donatelli	Consumer	United Sta	San Franci	California	TEC-PH-10001949	Technology	Phones	Cisco SPA 501G IP Phone	213.48
21	ZD-21925	Zuschuss Donatelli	Consumer	United Sta	San Franci	California	OFF-BI-10002215	Office Supplies	Binders	Wilson Jones Hanging View Binder, White, 1"	22.72
22	KB-16585	Ken Black	Corporate	United Sta	Fremont	Nebraska	OFF-AR-10000246	Office Supplies	Art	Newell 318	19.46
23	KB-16585	Ken Black	Corporate	United Sta	Fremont	Nebraska	OFF-AP-10001492	Office Supplies	Appliances	Acco Six-Outlet Power Strip, 4' Cord Length	60.34
24	SP-20065	Sandra Flanagan	Consumer	United Sta	Philadelph	Pennsylvai	FUR-CH-10002774	Furniture	Chairs	Global Deluxe Stacking Chair, Gray	71.372
25	EB-13870	Emily Burns	Consumer	United Sta	Orem	Utah	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	1044.63
26	EH-13945	Eric Hoffmann	Consumer	United Sta	Los Angeles	California	OFF-BI-10001634	Office Supplies	Binders	Wilson Jones Active Use Binders	11.648
27	EH-13945	Eric Hoffmann	Consumer	United Sta	Los Angeles	California	TEC-AC-10003027	Technology	Accessories	Imation 8GB Mini TravelDrive USB 2.0 A Flash Drive	90.57
28	TB-21520	Tracy Blumstein	Consumer	United Sta	Philadelph	Pennsylvai	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawyers Bookcase, Royale Cherry Finish	3083.43
29	TB-21520	Tracy Blumstein	Consumer	United Sta	Philadelph	Pennsylvai	OFF-BI-10000474	Office Supplies	Binders	Avery Recycled Flexi-View Covers for Binding Systems	9.618
30	TB-21520	Tracy Blumstein	Consumer	United Sta	Philadelph	Pennsylvai	FUR-FU-10004848	Furniture	Furnishings	Howard Miller 13-3/4" Diameter Brushed Chrome Round Wall Clock	124.2
31	TB-21520	Tracy Blumstein	Consumer	United Sta	Philadelph	Pennsylvai	OFF-EN-10001509	Office Supplies	Envelopes	Doherty Strite Five Envelopes	3.764

2. Methodology

Collection of data

We followed the predetermined procedure for gathering the data. Depending on the type, quantity, availability, and need of the data, the data acquisition method may differ from customer to customer.

Preprocessing and Cleaning of the data

The acquired data is put into a "purifying" procedure to make sure it is correctly segmented, identified gaps in the records are occupied with the relevant details, enabling the data suitable, and addressing faults in storage systems that may result in data redundancy.

Modeling of the data

This procedure essentially consists of analyzing the given dataset and items within it in order to clearly comprehend the needs that could assist our company strategy. Models are then developed for the project's flow based on the examination of data patterns. This flow provides improved support for using the previously decided semi-formal model that highlights the project's features. It also offers direction for understanding the connections between the data objects and other objects.

Prediction of data

This models involve training which are later tested against the data. The preprocessed dataset will then get this application. The models to be implemented in the forecast are as follows:

- Linear Regression
- Naïve Forecasting

Data Visualization

The data is then further visualized to draw conclusion and make effective decisions.

Evaluation metrics

Evaluation measures are crucial, when developing an effective model. It is useful to obtain an indicative data regarding the

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_{\text{predict}} - x_{\text{actual}}|$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (|x_{\text{predict}} - x_{\text{actual}}|)^2}$$

where n denotes the total number of error,

$|x_{\text{predict}} - x_{\text{actual}}|$ is the absolute error.

Figure.1 represents the architecture diagram of the suggested methods. Here, RMSE, MAE defines accuracy metrics for predicting sales in Superstore sales and according to the accuracy measures, the

model that may be compared and improved until we reach a high level of accuracy. The capacity to focus on a model's output and differentiate across model outputs is made possible by evaluation metrics. Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) was the evaluation measure we employed in our study. Both metrics are used as parameters for determining the accuracy of a continuous variable.

model will forecast most accurately when the MAE and RSME are kept to a minimum. By determining the metrics, we can conclude the best yielding algorithm.

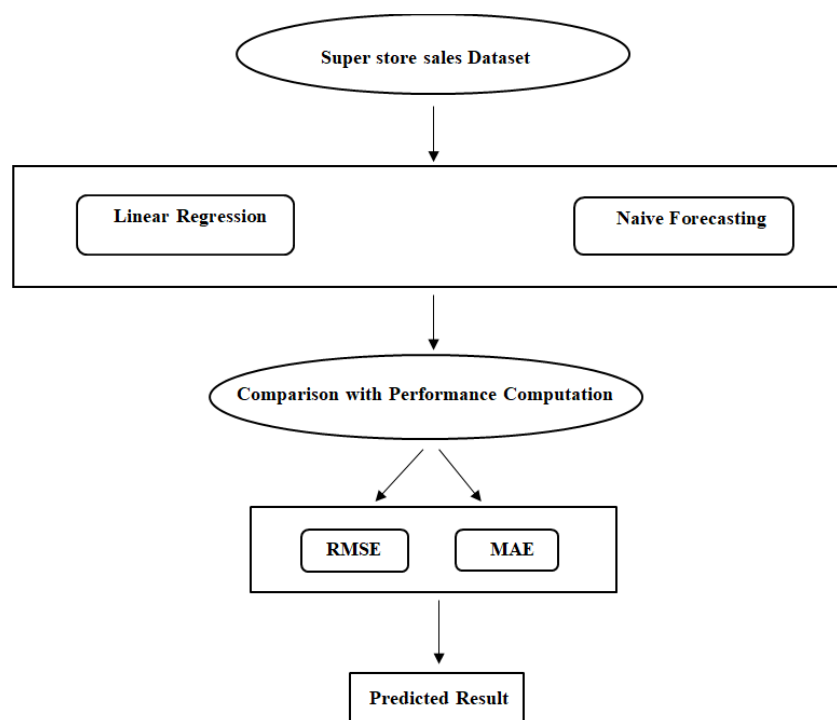


Figure.1

Model Building

Linear Regression

Linear Regression, a supervised machine learning algorithm is one such widely used predictive analysis model. It is the most effective statistical method which can be used to generate insights on consumer behaviour, enterprise understanding, and factors impacting revenue and profits. It can be employed in business to interpret trends and make estimates or prediction.

A simple linear regression investigates statistically the strength of the relation between one independent variable (x) and one dependent variable (y). An independent variable, also known as predictor variable, that altered in a study to see how it impacts the dependent variables. Patterns in data can be discovered and implemented in business operations. A Linear Regression Model was constructed by fitting a trend line to a dataset that already contains a linear relationship. In this paper, we have considered the global superstore retail dataset from the Kaggle Dataset. The Linear regression model is used on our

trained dataset which helps to make predictions about future sales values. The relation between the dependent and independent variables is described in this regression model. The dependent variable y is referred as responding variable and the independent variable x are an input factor and it is referred as explanatory or predictive variable. The linear regression equation is stated as $y=a+bx$.

Naive Forecasting

A method known as Naive forecasting uses the previous period's sales data to estimate the subsequent without making any presumptions or altering the variables. It is primarily based on historical sales or other variables such as cash flows and revenues. In this study, the retail dataset of the global superstore is taken from the Kaggle dataset. Here, Naive forecasting model is used on our trained dataset for the analysis. It is the one-step-ahead forecast which equals to the most recent actual value, $\hat{y}_t = y_t - 1$.

3. Results

Results of Descriptive Statics of Study Variables are given as follows:

Name	RMSE	MAE
Linear Regression	625.4	272.7
Naive Forecasting	874.8	337.1

The sales for the next 5 days by using ARIMA model for the linear regression approach is as follows:

Days	Sales
1.	268.3158
2.	272.719
3.	272.7935
4.	272.797

5.	272.7948
----	----------

4. Conclusion

In this study, we attempt to predict a retail business's sales by Linear Regression and Naive Forecasting methods and experiment it on the superstore sales dataset which is taken from the Kaggle dataset. Table 1 provides a summary of a comparison of the methods and further shows that, in comparison to other methods, Linear Regression approach, performs better. Here, we found that the Naive Forecasting method has the most significant RMSE and MAE of 874.8, 337.1 respectively and the method with least RMSE and MAE value is Linear regression having 625.4, 272.7 respectively. Hence we can conclude from the results that the linear regression is the most reliable predictor for the whole set because it has the lowest RMSE and MAE values. Also, we have predicted the sales for the next 5 days by using ARIMA model for the linear regression approach. Therefore, Linear Regression is the optimum algorithm for effective sales analysis. Shops, grocers, brand outlets, and other businesses primarily use this methodology. A very efficient method of managing sales is provided by the data analysis used with predictive machine learning models. It generously aids in making better decisions and formulating plans based on demands for the future. In the current era, this method is highly encouraged because it helps a lot of businesses, organizations, researchers, and brands to achieve the results that help them manage their profitability, sales, inventory, data collection, and customer needs.

5. References

1. Armstrong, J. S. (2008). Sales Forecasting. SSRN Electronic Journal. doi:10.2139/ssrn.1164602
2. S. Weigend and N. A. Gershenfeld (1994). "Time series prediction: Forecasting the future and understanding the past", Addison-Wesley.
3. Bose, I., & Mahapatra, R. K. (2001). Business data mining — a machine learning perspective. *Information & Management*, 39(3), 211–225. doi:10.1016/s0378-7206(01)00091-x
4. Domingos, P.M. (2012), A few useful things to know about machine learning. *Commun.acm* 55(10), 78–87. doi:10.1145/2347736.2347755
5. Hirst, M. (1970). Building and operating a forecasting model: the regression analysis approach. *European Journal of Marketing*, 4(3), 121–125. doi:10.1108/eum0000000005189
6. J Martinovic, V DamnJanovic, (May 17-18, 2006) INTERNATIONAL SCIENTIFIC DAYS 2006, "Competitiveness in the EU – Challenge for the V4 countries".
7. Langley, P., & Simon, H. A. (1995). Applications of machine learning and rule induction. *Communications of the ACM*, 38(11), 54–64. doi:10.1145/219717.219768
8. M Kumar, M Anand (2014). An Application Of Time Series Arima Forecasting Model For Predicting Sugarcane Production In India .*Studies in Business and Economics*, Research gate.
9. RohitSav, PratikshaShinde, SaurabhGaikwad (2021, June). Big Mart Sales Prediction using Machine Learning.2021 *International Journal of Research Thoughts (IJCRT)*. doi: 10.48175/568
10. Wacker, J. G., & Lummus, R. R. (2002). Sales forecasting for strategic resource planning. *International Journal of Operations & Production Management*, 22(9), 1014–1031. doi:10.1108/01443570210440519