iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XI **Month of publication:** November 2023

DOI: https://doi.org/10.22214/ijraset.2023.56609

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

Sales Predictor Using Simple Linear Regression and Simple Moving Average Model

Ananth S, Koulik Ghosh, Sahal P. K, Preethy Jemima P
SRMIST Ramapuram

Abstract: In the rapidly evolving business landscape, the accurate prediction of sales trends holds paramount importance for informed decision-making. Our project deals with the application of two fundamental forecasting models: SLR and SMA Models to predict future sales trajectories. The Simple Linear Average model involves the computation of a basic mean sales value, while the Simple Moving Average model incorporates a rolling average over a predefined time span. Upon successful training of the model, the model can predict the sales of the particular product with accuracy above 90%.
Keywords: Simple Linear Regression, Simple Moving Average Model, Machine Learning, Vector Auto Regression

I. INTRODUCTION

Forecasting sales poses a dilemma, for businesses across industries. However it plays a role in determining the allocation of advertising budgets. Sales prediction involves estimating the quantity of a product that people are likely to purchase taking into account factors such as advertising expenditure, target audience segmentation and the chosen advertising platform.

Accurate sales forecasting enables companies to effectively manage their workforce, cash flow and resources. Furthermore it provides insights, for businesses seeking investment capital by showcasing their growth financial stability.

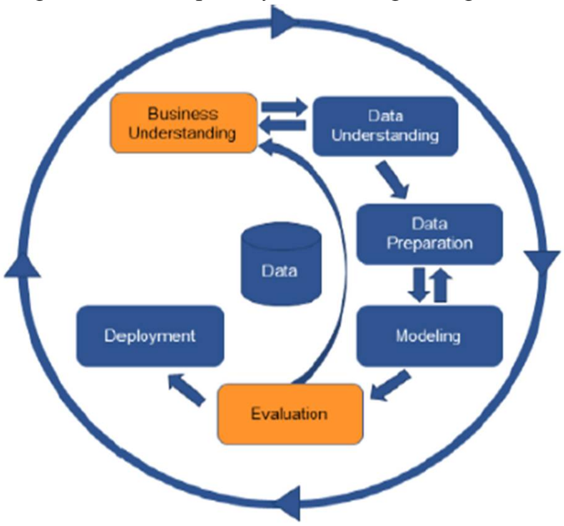

Fig 1 : Stages in Data Mining

Sales play a role, in driving the success of any business. Many companies face challenges when it comes to planning, which involves managing cash flow determining supply levels for efficient production and planning human resources to ensure the right staff is in the right positions.

These tasks include hiring employees as needed retaining staff or making reductions to minimize risks that could impact business performance. On the hand sales forecasting can guide businesses in making decisions about their inventory and contribute to overall financial planning. By analyzing sales data it enables predictions about future performance.

Therefore it is important to develop a system that utilizes SLR analysis and SMA statistical models for sales forecasting. Our project aims to predict the sales of products on any given day by considering factors such, as time and day of sale. We can even incorporate days together as input parameters to obtain our desired output. The linear regression model will perform the calculations while the moving average model will be used for sales forecasting.

II. RELATED WORKS

The primary focus of the research paper [1] revolves around combining decision analysis and predictions in a sales prediction system. The aim is to enhance the accuracy and efficiency of sales forecasting by utilizing data mining techniques. The study evaluates predictive models and concludes that the Gradient Boost Algorithm is the most suitable model for accurately predicting sales trends. The paper emphasizes the application of data mining techniques in sales forecasting highlighting their potential, for improving sales predictions through a thorough study and analysis of understandable predictive models.

In the paper [2] the researchers explore the effectiveness of using a convolution structure as the primary architecture of a convolutional neural network (CNN) for image recognition tasks. Surprisingly they achieve performance even without incorporating a pooling layer. The study also introduces a technique called **denoising autoencoder** (DAE) for unsupervised pre-training. This approach proves to be beneficial in enhancing the training effectiveness of learning models.

Furthermore the paper compares the performance of learning CNN models with that of AdaBoost models, where decision trees are used as basic classifiers to optimize AdaBoosts performance.

To construct a sales prediction model for products, data mining and deep learning techniques are utilized in this research. The authors find that employing a pre trained CNN model is more effective and adaptable, for sales forecasting.

The accuracy index is employed as an evaluation metric to assess the training performance of this classification prediction task.

In a research paper [3] the authors conducted an analysis of three different models, namely Vector Auto Regression (VAR) Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) to determine their accuracy in sales forecasting for Store XYZ during the COVID period.

They found that the SARIMA model with parameters (0,1,0) (0,1,0)12 achieved the level of accuracy with a Mean Absolute Deviation (MAD) value of 0.122. The study utilized time series data from 2017 to 2022. Applied these models to analyze sales data from a retail store. The accuracy assessment was based on comparing the forecasted values with the values using MAD.

To make sure the ARIMA model was appropriately used some pre-processing steps were undertaken. These included organizing the sales data and reducing it to one dimensional information.

The selection of SARIMA as the preferred approach was driven by its ability to account for observed variations, in historical sales data.

However it is important to note that this research paper did not delve into discussing any limitations or weaknesses associated with VAR, ARIMA or SARIMA models nor did it provide any analysis regarding drawbacks of relying solely on MAD as a measure of accuracy when comparing forecasting models.

Other evaluation measures, like Mean Squared Error (MSE) or Root Mean Squared Error (RMSE) can offer understanding and valuable insights.

In the research paper [4] there is a discussion about different machine learning algorithms that are used for sales predictions.

To predict sales, various machine learning models such as Linear Regression, K Neighbors Regressor, XGBoost Regressor and Random Forest Regressor are utilized. In addition to these models clustering methods and measures are employed to evaluate the effectiveness of approaches in predicting sales.

The paper emphasizes the potential of methods in accurately forecasting sales. It explores the use of clustering models and measures to enhance prediction outcomes. Moreover it delves into how machine learning algorithms and feature selection techniques can be applied to improve prediction results. Bayesian learning and neural network approaches are also incorporated for sales prediction purposes. The insights derived from these predictions prove valuable in constructing setups that estimate a number of outputs.

Furthermore the paper suggests directions for research including exploring other machine learning algorithms and techniques for sales prediction. It also recommends utilizing data mining techniques and data visualization to gain insights, for decision making related to sales predictions.

In a research paper titled [5] the main focus is on the importance of sales trend prediction methods in the e commerce industry. The paper emphasizes the need for data mining techniques to achieve sales revenue and efficient inventory management. It also explores the relationship between stocks and sales highlighting how accurate sales forecasting plays a crucial role.

The challenges faced by prediction systems when dealing with big data are discussed, along with the accuracy of sales forecasting. The paper utilizes data mining techniques to uncover insights from large datasets ensuring precise sales forecasting. Machine learning algorithms such as feed forward and recurrent ELM networks are mentioned as tools for developing neural network models that can forecast footwear sales.

Furthermore there is a mention of a trigger system that matches commodities with prediction models to enhance sales prediction outcomes. The paper introduces an algorithm called Extreme Learning Machine (ELM). Incorporates e commerce related indicators to improve the accuracy and reliability of sales predictions. Clustering techniques are also highlighted as tools, for making sense of vast amounts of data when it comes to predicting future sales.

III. PROPOSED WORK

The proposed system is an intelligent Sales Predictor with higher accuracy .Its primary purpose is to capture sales data and utilize it for predicting future sales using SLR analysis and SMA forecasting models. This innovative tool offers businesses a resource for making informed decisions in a timely manner. It also generates reports regarding customer orders and provides a user friendly interface for easy data input and retrieval. Moreover the system can be tailored to meet the requirements of each business ensuring optimal efficiency and accuracy in sales forecasting. With its capability to analyze sales data trends and patterns this system is indispensable, for any business aiming to outperform their competitors.

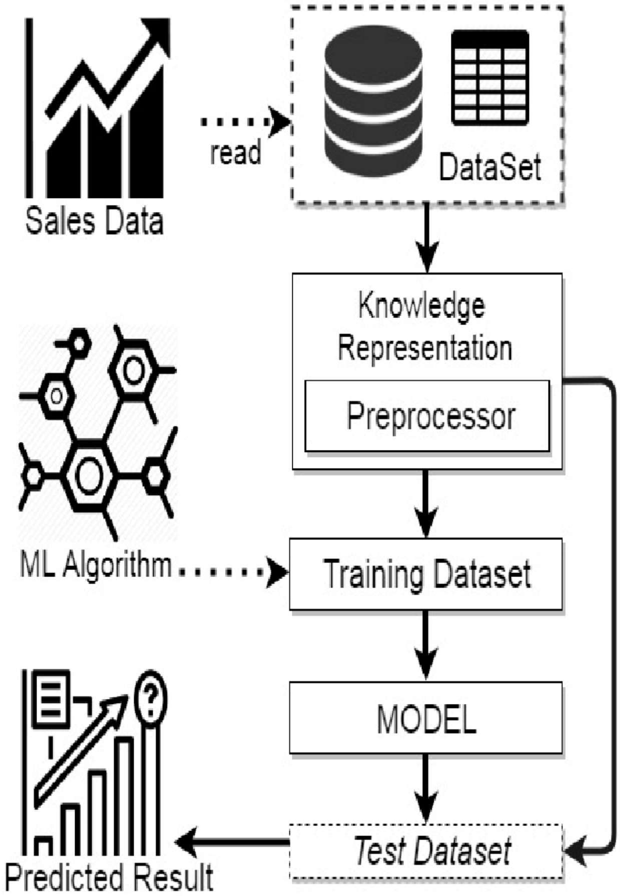

Fig 2 : Architecture Diagram

The entire prediction and forecasting of sales in based on SLR and SMA Models.
The reason for using SLR and SMA model is due to its simplicity and ease of implementation.
Though SLR is easy , its comparatively accurate as it has multiple evaluation metrices.
Now let’s look at SLR and SMA in detail.

A. Sales Forecast using SMA Models

The simple moving average method is a way to predict the period (t+1) by taking the average of a specified number of recent observations. Each observation carries the weight, which helps in smoothing out short term fluctuations in the data series. Simple moving averages are useful when the data series remains relatively stable over time. According to Hyndman (2009) a moving average is created by averaging values from another time series and its a form of mathematical convolution. Moving averages are also known as running means or rolling averages. Are considered a type of "filtering" that transforms one time series into another.

The name "moving average" comes from the fact that each average is calculated by dropping the observation and including the next one. Moving averages serve two purposes;

1) *Two Sided (Weighted):* This type of moving average is employed to smooth out a time series helping estimate or highlight its underlying trend.
2) *One Sided (Weighted):* This kind of moving averages is commonly used as a forecasting technique for time series. Although moving averages are simple they serve as the foundation for advanced methods in time series smoothing, decomposition and forecasting. Additionally a simple moving average can be calculated by taking the average of the N recent values, where N is an integer. This equation predicts the value of F at time t+1 using data up, to time t.

Moving Average Model
Ft = (sum of actual values in previous n periods)/n
(Ft+1)= (St + St-1 + St-2 + … + St_{N-1})/ N
So the three-period moving average would be:
F3 = (St + St-1 + St-2) /3
A four-period moving average would be:
F4 = (St + St-1 + St-2 + St-3) /4
A five-period moving average would be:
F5 = (St + St-1 + St-2 + St-3 + St-4) /5
and so on, for as many periods as needed for the model.

The primary benefit of the SMA is that it provides a smoothed line that is less likely to whipsaw up and down in reaction to minute, transient fluctuations in price. The shortcoming of the SMA is its delayed reaction time to sudden price swings, which frequently happen during market reversals. When it comes to traders or analysts working on larger time frames, such daily or weekly charts, the SMA is frequently preferred.

B. Sales Forecast Using Simple Linear Regression:
The simple linear regression model is given by :
$$\mathbf{y = \beta_0 + \beta_1 x + \varepsilon}$$
Where:
- y - The dependent or study variable
- x - The independent or explanatory variable.
- β_0 - The intercept parameter
- β_1 - The slope parameter
- ε - The un-observable error parameter

However, the terms β0 and β1are the parameters of the model. These parameters are usually called as regression coefficients. This parameter ε accounts for the failure of data to lie on the straight line and represents the difference between the true and observed realization of y.
$$\mathbf{\beta_0 = Y - \beta_1 X}$$
$$\mathbf{\beta_1 = \frac{\Sigma XY - n(Y)(X)}{\Sigma X2 - n(X)}}$$

IV. MODULE DESCRIPTION

The proposed method is an intelligent web-based application capable of capturing sales data and utilizing it to estimate sales using forecasting models such as Simple Moving Average and Simple Linear Regression.This would facilitate prompt decision making for the company.
The system can also generate accurate sales reports as and when needed by the customer for any particular product.
The model once trained and implemented can predict sales of a particular product or total sales on a particular day after providing a set of inputs,

The parameters used as input are :
- Past seven day sales
- Day of the week
- Date
- Season (Peak or Regular)
- Festival or Long Holidays

Required Packages:
- numpy
- pandas
- keras
- tensorflow
- csv
- matplotlib.pyplot

A. Data Collection
This module begins with the collection of historical sales data. The data should include a time series of sales figures of a considerable period.
The data would be in the form of a data-set that contains all the necessary sales figures.

B. Pre-processing
All the above parameters taken as input will have to be converted into a form that the machine understands and that is done under the pre-processing stage.
The reason for having multiple parameters is to increase the efficiency of the model.

C. Simple Moving Average (SMA) Analysis
SMA is a basic time-series analysis technique used to identify trends and smooth out fluctuations in data. The module calculates SMA for a specified period (e.g., daily, weekly, or monthly) to observe the average sales performance over time. Simple moving average can be computed by taking a simple average of the most recent N values, for some integer N.
This is the equation for predicting the value of F at time t+1 based on data up to time t:

Moving Average Model:

$$Ft = \frac{(\text{sum of actual values in previous n periods})}{n}$$

$$Ft+1 = \frac{((St) + (St-1)+(St-2)+ \ldots + (St-N-1))}{N}$$

D. Simple Linear Regression Analysis:
Simple Linear Regression is employed to establish a relationship between sales and one or more predictor variables (such as time or other relevant factors).
The module performs the following steps:
- Identifies the dependent variable (sales) and independent variable(s).
- Fits a linear regression model to the data.
- Evaluates the model's coefficients, significance, and goodness of fit.
- Predicts future sales based on the regression model.

The simple linear regression model is given by :

$$\mathbf{y = \beta_0 + \beta_1 x + \varepsilon}$$

Where:

y - The dependent or study variable

x - The independent or explanatory variable.
β_0 - The intercept parameter
β_1 - The slope parameter
ε - The un-observable error parameter

However, the terms β0 and β1are the parameters of the model. These parameters are usually called as regression coefficients. This parameter ε accounts for the failure of data to lie on the straight line and represents the difference between the true and observed realization of y.

$$\beta_0 = Y - \beta_1 X$$

$$\beta_1 = \frac{\Sigma XY - n(Y)(X)}{\Sigma X2 - n(X)}$$

E. Model Evaluation

We assess the accuracy of both the Moving Average and Simple Linear Regression models using appropriate evaluation metrics like Mean Absolute Error, Root Mean Squared Error and R-squared error etc.

1) Mean Absolute Error (MAE):

MAE measures the average absolute difference between the actual and predicted values. It quantifies the model's accuracy, with lower values indicating better performance.

Formula

$$\mathbf{MAE} = \frac{\Sigma|\text{Actual - Predicted}|}{n}$$

Where:

Σ- represents the sum over all data points.
Actual - represents the actual sales values.
Predicted - represents the predicted sales values by the SLR model.
N - the number of data points.

2) Root Mean Squared Error (RMSE):

RMSE is similar to MAE but gives more weight to larger errors. It provides a measure of the model's precision, with lower values indicating better performance.

Formula

$$\mathbf{RMSE} = \sqrt{[\Sigma(\text{Actual - Predicted})^2 / n]}$$

Where:

Σ represents the sum over all data points.
Actual represents the actual sales values.
Predicted represents the predicted sales values by the SLR model.
n is the number of data points.

3) R-squared (R^2)

R-squared measures the proportion of the variance in the dependent variable (sales) that is explained by the independent variable(s) in the SLR model. A higher R-squared value indicates a better fit, with 1.0 being a perfect fit.

Formula:

$$R^2 = 1 - (SSR / SST)$$

Where:

SSR (Sum of Squared Residuals) is the sum of the squared differences between the actual sales values and the predicted values.
SST (Total Sum of Squares) is the sum of the squared differences between the actual sales values and their mean.
Apart from the above mentioned other methods like adjusted r-squared , residual analysis , forecast error and MAPE can be used for evaluation metrics

F. Visualization

Visualizations, such as line plots and scatter plots, are generated to present historical sales data, the Moving Average trend-line, and the linear regression line. These visuals help us understand sales patterns and predictions.

G. Sales Forecasting

Based on the results of the SMA Model and SLR analyses, the module provides sales forecasts for a specified future time period. These forecasts serve as valuable guides for decision-making, inventory management, and resource allocation.

V. OUTPUT AND ACCURACY

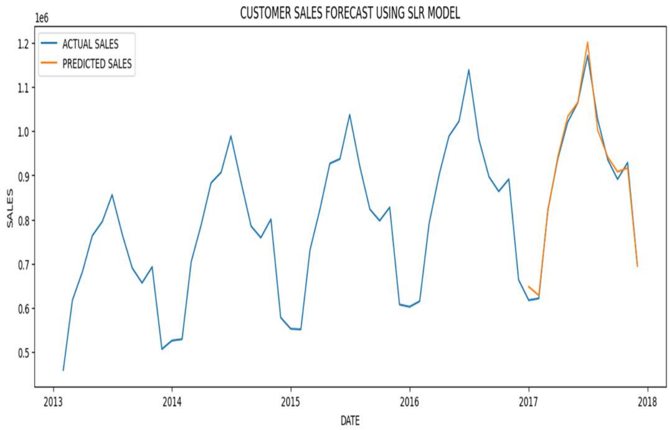

Fig 3 : Sales Forecast With The Actual and Predicted Sales

```
[ ] # print(predict_df)
    lr_mse = np.sqrt(mean_squared_error(predict_df['Linear Prediction'], monthly_sales['sales'][-12:]))
    lr_mae = mean_absolute_error(predict_df['Linear Prediction'], monthly_sales['sales'][-12:])
    lr_r2 = r2_score(predict_df['Linear Prediction'], monthly_sales['sales'][-12:])
    print("Linear Prediction MSE: ", lr_mse)
    print("Linear Prediction MAE: ", lr_mae)
    print("Linear Prediction R2: ", lr_r2)

    Linear Prediction MSE:  16221.272385416869
    Linear Prediction MAE:  12433.184266490736
    Linear Prediction R2:  0.9906152516380969
```

Fig 4 : Corresponding Values of Evaluation Metrics used in SLR

Accuracy of a SLR Model is based on the following evaluation discussed previously.
The corresponding values for our prediction model is :
*Linear Prediction MSE: 16221.272385416869
*Linear Prediction MAE: 12433.184266490736
*Linear Prediction R2: 0.9906152516380969
A regression is considered to be more precise or accurate based on its R2(R-SQUARED) Value.
Any regression model that has a relatively high R2 value (close to 1) is considered to be precise.
Our model has a R2 value of 0.99 and that clearly indicates how accurate the model is.
The same can be verified from the graph by comparing the deviation of predicted sales from the actual sales.

VI. CONCLUSION

At the end of this research work, a Web-based intelligent system application capable of capturing sales data from table and use the data to carry out sale forecast using Simple Linear Regression analysis and Simple Moving Average forecasting models was developed. This research work has provided the Management with a web-based system capable of capturing and documenting daily sales data. This web-based system for sales forecast is very flexible and open to more improvements. These improvements are expected to be primarily around the possibilities on using other algorithms and also on increasing the accuracy of the predictions made.

REFERENCES

[1] S. Cheriyan, S. Ibrahim, S. Mohanan and S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), 2018, pp. 53-58, doi: 10.1109/iCCECOME.2018.8659115.
[2] Prediction of Merchandise Sales on E-Commerce Platforms Based on Data Mining and Deep Learning (2021) Xiaoting Yin, Xiaosha Tao
[3] Predicting sales during COVID using Machine Learning Techniques Sagar Viswakarma. Dr.S.C. Solank.
[4] Sales Prediction Using Machine Learning Algorithms. Purvika Bajaj. Renesa Ray. Shivani Shedge. Shravani Vidhate. Prof. Dr. Nikhilkumar Shardoor.
[5] Intelligent Sales Prediction Using Machine Learning Techniques. (2019) Mr. Akshay Godse. Miss. Poonam Pawar. Mr.Sairaj Sawant. Miss.Shirin Mujawar. Prof. Prashant Karkalle.
[6] Using Supervised Machine Learning for B2B Sales. (2021) D. Rohaan ,E. Topan ,C.G.M Groothuis Oudshoorn
[7] K. Punam, R. Pamula and P. K. Jain, "A Two-Level Statistical Model for Big Mart Sales Prediction," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), 2018, pp. 617-620, doi: 10.1109/GUCON.2018.8675060.
[8] H. Zhu, "A Deep Learning Based Hybrid Model for Sales Prediction of E-commerce with Sentiment Analysis," 2021 2nd International Conference on Computing and Data Science (CDS), 2021, pp. 493-497, doi: 10.1109/CDS52072.2021.00091.
[9] H. V. Ramachandra, G. Balaraju, A. Rajashekar and H. Patil, "Machine Learning Application for Black Friday Sales Prediction Framework," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 57-61, doi: 10.1109/ESCI50559.2021.9396994.
[10] J. Chen, W. Koju, S. Xu and Z. Liu, "Sales Forecasting Using Deep Neural Network And SHAP techniques," 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), 2021, pp. 135-138, doi: 10.1109/ICBAIE52039.2021.9389930.