International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 4, July Year 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Enhancing Business Profitability Through Regression-Based Sales Forecasting

Usha R*, Vijiyalakshmi V**, Shenbaha S***
*(Mathematics, Dr. N.G.P. Arts and Science College, Coimbatore
Email: usharajkumaras@gmail.com)
**(Mathematics, Dr. N.G.P. Arts and Science College, Coimbatore
Email: kumarvijai127@gmail.com)
***(Computer Science with Data Analytics, Dr. N.G.P. Arts and Science College, Coimbatore
Email: sshenbaha@gmail.com)

--*************************--

Abstract:

Business is an uncertain field where the market fluctuates daily. To overcome this challenge, our case study explores the implementation of regression analysis in predictive modelling. From the historical data set, our model estimates the daily revenue of a coffee shop based on multiple factors such as the number of customers, average order value, marketing spends per day, and location for traffic, which impacts the daily revenue of the coffee shop. The main objective of the case study is to enhance and optimize profits through the regression model, which assists young entrepreneurs to handle the unstability of the business market. Here we choose linear and polynomial regression model as it is simple, effective, and easily understandable in predictive modelling. It's also directly correlating the relationship between independent variables to the target variable. Our case study demonstrates the performance of the model, which enables entrepreneurs to predict future revenue and also helps to adjust marketing strategy.

Keywords -- **Regression model, business, coffee shop, profits, prediction.**

--*************************--

I. INTRODUCTION

The international coffee shop market has become a significant brewing business field. In such a dynamic environment, understanding and predicting the business development is essential for the improvement of the business. In recent times coffee shops have become not just places for beverages but have also evolved into spaces for relaxing, meetings and working. As the number of coffee shops increases, the competition becomes heavy, often experiencing market swings in the business. The shop owners struggle to maintain the stability and profitability of the coffee shop. The daily revenue was affected by various factors like the average number of customers, average order value, number of working marketing strategies, etc. To address this challenge, we proposed the model in the concept of regression analysis in predictive modelling. This study explores the implementation of a regression model to predict the daily revenue of the coffee shop, which assists the business owners to enhance profitability and customer satisfaction. Regression analysis establishes the relationship between impacting key factors and daily revenue. To achieve an accurate prediction model, the study employs linear regression and polynomial regression. By using Python code, we implemented the models for the coffee shop data set. The performance of the models was evaluated by measures of errors like MAE, MSE, and R-squared values. Predicted the predicted regression model helps shop owners to understand the pattern of trends and adjust marketing strategies, which ultimately leads to enhancing and optimizing the profit of the coffee shop now.

II. LITERATURE REVIEW

Zenan Chen proposed prediction in coffee sales manufacturing by ARIMA model in the year 2024. The paper predicts coffee sales in vending machines using the Autoregressive Integrated Moving Average Model with RMSE value of 68.9416.

Coffee Shop Revenue Prediction by Sonawane Lalit, a Decision Tree model was applied which achieved an R-squared value of 88.35.

An article by Techshastri titled "Coffee Shop Daily Revenue Prediction" implemented Random Forest and Ridge Regression models, achieving a Root Mean Squared Error (RMSE) value of 262.21.

III. METHODOLOGY

To develop an effective regression model, the methodology entails collecting raw datasets and preprocessing them through data cleaning and data training.

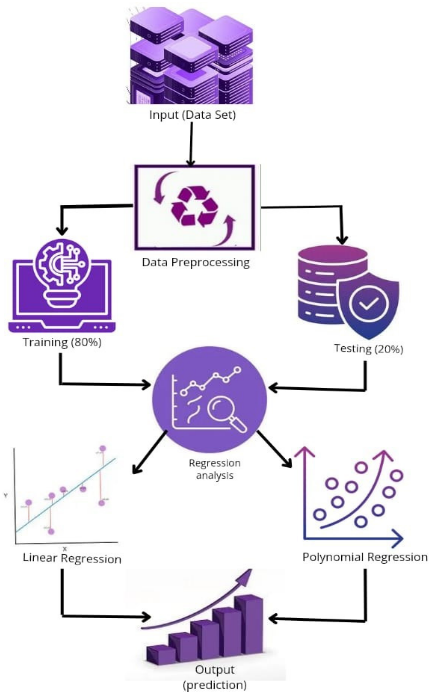

Fig. 1 Methodology flow

A. Data Collection

The dataset for the coffee shop is extracted from the Kaggle website. It consists of 2000 data entries with features such as number of customers per day, average order value, operating hours per day, number of employees, marketing spend per day and location foot traffic, which helps in forecasting the daily revenue of the coffee shop.

Number_of_Customers_Per_Day	Average_Order_Value	Operating_Hours_Per_Day	Number_of_Employees	Marketing_Spend_Per_Day	Location_Foot_Traffic	Daily_Revenue
152	6.74	14	4	106.62	97	1547.81
485	4.5	12	8	57.83	744	2084.68
398	9.09	6	6	91.76	636	3118.39
320	8.48	17	4	462.63	770	2912.2
156	7.44	17	2	412.52	232	1663.42
121	8.88	6	9	183.49	484	1155.18
238	9	11	4	331.35	156	2179.13
70	7.81	10	3	273.27	237	890.17
152	8.78	14	2	341.79	825	1704.94
171	7.73	7	5	344.51	135	2025.55
264	7.6	13	14	318.08	885	2099.66
380	7.14	16	9	140.39	892	2568.92
137	8.15	10	14	29.11	505	1367.08
422	3.69	7	4	233.23	186	1817.02
149	9.11	15	2	446.58	907	1707.13

Fig. 2 Data Collection

B. Data Cleaning

Data cleaning is a crucial process that involves techniques such as handling missing values, correcting errors, fixing inaccuracies and identifying and removing outliers in the coffee shop dataset. We apply python code to examine missing and duplicate values, rectify errors and eliminate outliers.

Per_Day	Average_Order_Value	Operating_Hours_Per_Day	Number_of_Employees	Marketing_Spend_Per_Day	Location_Foot_Traffic	Daily_Revenue
152	6.74	14	4	106.62	97	1547.81
485	4.50	12	8	57.83	744	2084.68
398	9.09	6	6	91.76	636	3118.39
320	8.48	17	4	462.63	770	2912.20
156	7.44	17	2	412.52	232	1663.42

Fig. 3 Data cleaning

C. Data Training

In data training, the dataset in split into two sections: A training set and a testing set. The training set uses 80% of the data. The testing set uses the remaining 20% of the data.

Training set - Helps to understand the patterns and the relationship between the independent and dependent variables in the model, which helps in forecasting future outcomes based on past trends.

Test set - Used to validate to assess the performance of the trained model, ensure it generalizers to the new and unseen data and also assist in evaluating the accuracy of the predicted model.

IV. REGRESSION ANALYSIS IN RESULT AND DISCUSSION

Regression analysis is a widely adopted technique in predictive modelling. Regression analysis precisely quantifies the relationship between the dependent and the independent variables. It can effectively handle both linear and nonlinear models. So, to predict the daily revenue of the coffee shop near establishes linear and polynomial models.

D. Linear Regression model

Linear regression is a basic predictive model in the statistics and machine learning field. It evaluates the association of dependent variables with independent variables. The linear regression model fits a linear form (a straight line) relationship between the dependent and the independent variables.

A multilinear regression is a modelling approach used to illustrate the correlation between the dependent and multi-independent variables.

The study explores how multi-linear regression fits a linear form (a straight line) for the coffee shop data set. It takes 6 independent variables (Number of customers per day, Average order value, operations hours per day, number of employees, marketing spend per day, location foot traffic) and Daily revenue as target (dependent) variable.

Here, the variable is represented as follows:

Independent variable

x1 -Number of customers per day
x2- Average order value
x3- Operating hours per day
x4-No. of employees
x5-Marketing spend per day
x6-Location foot traffic

Dependent variable

y -Daily revenue (target variable)

The multi linear regression model is developed by using the Python language, it aids in developing the regression equation the equation is

Daily revenue = -1451.09+5.47*x1+240.58x2-1.14x3-2.64x4+1.52*x5+0.03x

The above equation is applied to determine the predicted value of the daily revenue for the coffee shop.

To measure the efficiency of the model, we need to calculate the square value; by analysing, we can determine the performance of the model. From the predicted multi-linear regression equation, the r-square value of 0.89 indicates the model is a good fit for the prediction.

The actual and predicted value for linear regression model is Actual: 571.57, Predicted: 553.74

Actual vs predicted value in linear regression

The image for the Actual vs predicted value of the daily revenue for the coffee shop is given below:

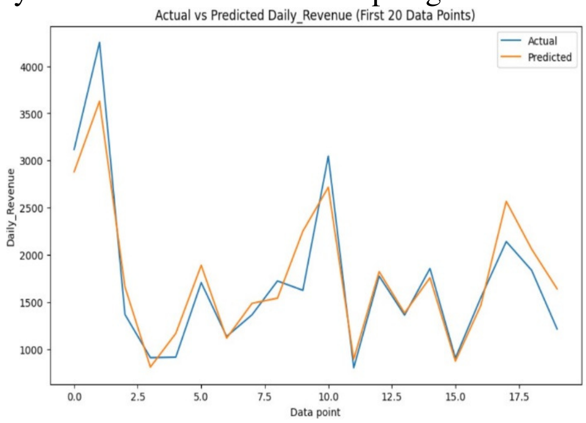

Fig. 4 Linear regression

The above graph clearly the shows the perfect fitness of the linear regression model.

E. Polynomial Regression model

Polynomial regression is a popular regression method utilized for establishing the nonlinear relationships between the variables. Polynomial regression is an expanded form of linear regression. It depicts the relationship between the multi-independent variables (x1, x2, x3...xn) and the dependent variable, (y) which is modelled as a polynomial function of the nth degree (representing nonlinear patterns) of the model.

In predicting the daily revenue of the coffee shop, to construct an accurate and optimized prediction model, we use the polynomial regression technique. Here, we take the key factors as independent variables (x1, x2, x3, x4, x5, x6), that impact the target variable (y) daily revenue, which supports forming the polynomial regression equation.

Using polynomial regression equation, we can precisely calculate the predicted value for the coffee shop ‘s daily revenue. The R^2 value is 0.95899. It shows that 95% of the variance in dependent variable determined by the independent variables in the polynomial regression model, which reflects the polynomial is effective robust model with high accuracy rate then the linear regression model.

The actual and predicted value for polynomial regression model is Actual: 942.70, Predicted: 963.09

Actual vs predicted value in linear regression

The image for the Actual vs predicted value of the daily revenue for the coffee shop is given below:

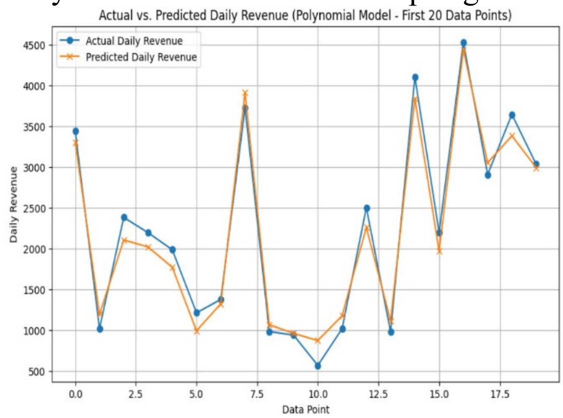

Fig. 5 Polynomial regression

The above graph clearly the shows the optimal fitness of the polynomial regression model.

F. Bar chart for Linear Regression Model and Polynomial Regression Model

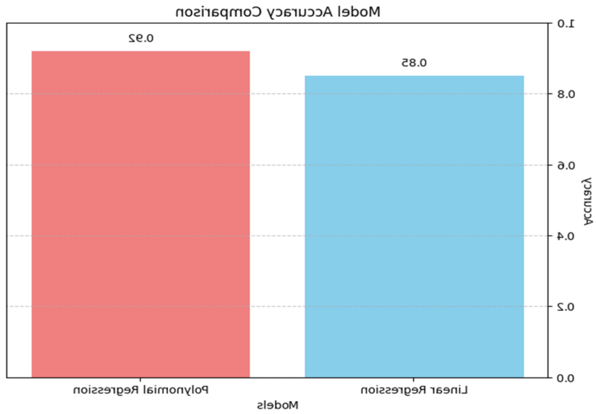

Fig. 6 Model accuracy comparison

G. Contribution

Feature	Contribution Percentage
No_of_Customers_Per_Day	20%
Average_Order_Value	30%
Operating_Hours_Per_Day	10%
Number_of_Employees	15%
Marketing_Spend_Per_Day	15%
Location_Foot_Traffic	10%

H. Pie chart of contributing percentages

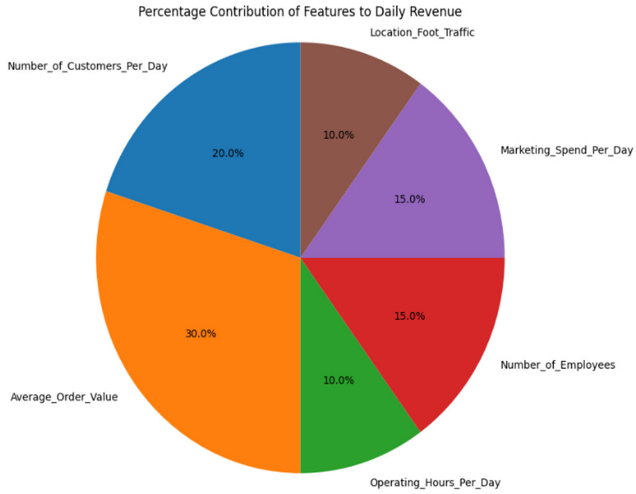

Fig. 7 Percentage contribution of features to daily revenue

V. CONCLUSION

The purpose of the study is to establish the predictive regression model for entrepreneurs and shop owners to sustain and maximize the revenue of the coffee shop. This research highlights the efficiency of regression analysis as a predictive tool for estimating the daily revenue. The dataset is analysed using linear and polynomial regression to identify the key factors that impact the daily revenue. The linear regression and polynomial regression models aid in forecasting trends and patterns. Among the models, polynomial regression yields the most accurate predicted values with an R square of 0.9588, depicting the strong correlation between the independent variables and the target variable daily revenue. The study gives insights into the contribution of key factors to the daily revenue. By leveraging these kinds of models, the businesspeople can optimize marketing and decision strategies and improve the profit of the coffee shop.

REFERENCES

[1] T. K. Hwase and A. J. Fofanah, *Machine Learning Model Approaches for Price Prediction in Coffee Market using Linear Regression,* XGB, and LSM Techniques, International Journal of Scientific Research and Technology, 2021.

[2] Z. Chen, *Coffee Sales Prediction for Vending Machines Based on ARIMA Model,* Journal of Forecasting, 2024.

[3] L. Nunno, *Stock Market Price Prediction using Linear and Polynomial Regression,* Ecological Modelling, 2006.

[4] Z. Aldott, *Predicting Specialty Coffee Auction Price using Machine Learning,* Warwick-Monash Economics Student Papers, 2021.
[5] D. Navarro, *Modelling Coffee Price using Regression*, GitHub Repository, 2021.
[6] C. Dominika, *Coffee Prices Prediction using Linear Regression Models*, GitHub Repository,2021.
[7] N.T. Mitchell, *Predicting Coffee Prices* – A Data Scientist's Notebook, Personal Blog, 2017.
[8] A. Joseph Fofanah, *Machine Learning Model Approaches for Price Prediction in Coffee Market using Linear Regression,* XGB, and LSTM Techniques, SSRN, 2021.
[9] T. K Hwase and A. J. Fofanah, *A Methodology for Coffee Price Forecasting Based on Extreme Learning Machine*, Procedia Computer Science, 2021.
[10] M. Kuhn and K. Johnson, *Applied Predictive Modelling*, Springer, 2013.
[11] S. Raschka*, Python Machine Learning*, Packet Publishing, 2015.
[12] K. Murphy, *Machine Learning*: A Probabilistic Perspective, MIT Press, 2012.
[13] C.M Bishop, *Pattern Recognition and Machine Learning,* Springer, 2006.
[14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.
[15] P. Domingos, *A Few Useful things to know about Machine Learning*, Communications of the ACM, 2012.