International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

Amazon Sales Prediction Model Using ML Algorithms

Keshav Pal[1]

ABSTRACT:
The e-commerce landscape has rapidly evolved, becoming a cornerstone of modern business with its digital paradigm for buying and selling goods or services. This dynamic and expansive platform, facilitated through websites and applications, transcends geographical boundaries, offering unprecedented access to a global consumer base. In this context, the ability to predict product sales has emerged as a critical facet for businesses operating in the e-commerce domain. The digital marketplace is characterized by fierce competition and rapid changes in consumer preferences. As businesses strive to optimize their operations, enhance profitability, and stay ahead of the curve, the predictive modelling of sales becomes invaluable. By forecasting sales, businesses can strategically adjust pricing strategies, manage inventory effectively, and tailor marketing efforts based on anticipated demand. The three fundamental factors—Price, Review, and Rating—constitute the core elements influencing consumer decisions in e-commerce. Understanding the intricate interplay of these variables is essential for businesses seeking a competitive edge. This research employs the Design of Experiment (DOE) technique to meticulously examine the relationship between Price, Review, Rating, and product sales. Leveraging a diverse dataset from Amazon, the study provides a comprehensive analysis of how these factors synergistically impact consumer behaviour and purchasing patterns. The resulting mathematical model demonstrates a promising capacity to forecast product sales accurately. In essence, the significance of predicting sales in e-commerce lies in its transformative potential for businesses, offering actionable insights that pave the way for informed decision-making, enhanced customer satisfaction, and sustained competitiveness in the ever-evolving digital marketplace.

Keywords: E-commerce, Sales, Marketplace, DOE, Mathematical Model, Forecast.

1.INTRODUCTION:

The e-commerce landscape in India has witnessed an exponential surge, transforming the way consumers engage with products and services. Among all the available digital marketplaces, Amazon is everyone's first choice, acclaimed not only for its expansive offerings but also for its ingenious customer-centric approach. At the heart of Amazon's meteoric rise in popularity lies its remarkable product backward search engine algorithm, tuned for deciphering trends and anticipating customer behaviour.

This research is an attempt to find the relationship between sales and three pivotal product parameters: Price, Review, and Rating. These factors are not only catalysts for Amazon's sophisticated search engine algorithms but also influence over consumer purchasing patterns and decision-making processes. By conducting a comprehensive analysis across diverse product categories, this study seeks to establish a mathematical relationship among these key input parameters.

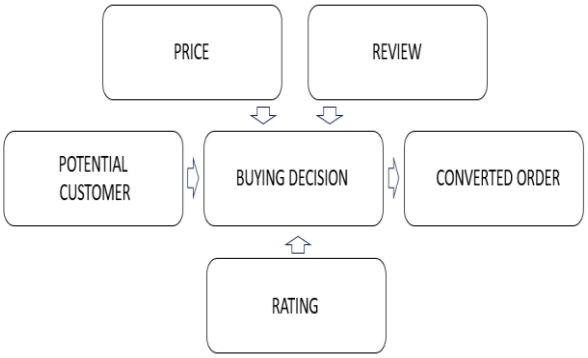

Fig 1. Factors influencing customers buying decisions.

The aim of this research is constructing a model that elucidates the intricate interplay between Price, Review, Rating, and product sales. Such a model holds immense potential for e-commerce stakeholders, offering a predictive tool that could revolutionize sales strategies. The envisioned outcome aims to empower sellers by providing actionable insights, thereby fostering a competitive advantage in the dynamic e-commerce realm.

2. LITERATURE REVIEW:

Factors like review volume, review sentiments, discount price etc are critically important as shown in an investigation carried out by Sharma S. et al. [1]. With new age technology like machine learning, mathematical model has helped people predict the demand with minimal data and more accuracy, a similar model was shown in paper by Smirnov et al. [2]. Pal K et al. [3] showed how a DOE model can be used to identify a relation between input parameters and response. Various research and model are developed by researchers in this field many concepts of regression analysis are used to develop and train models and datasets are then feed to establish a relation, a similar model was developed by Agnani D. et al. [4] for e commerce industry with accuracy of more than ninety percent. Kumar N. et al. [5] in his paper discussed the opportunities and challenges in e-commerce and emphasized the importance of demand and supply prediction. A series of research is conducted by Khanna P. to develop a predictive model using DOE techniques [6][7].

3.METHOD

This analysis follows the following step to develop a model.

1.Identification of input parameters that influence the decision making of customer.
2.Building of a design matrix.
3.Compilation of data as per the design matrix.
4.Development of a relational mathematical model.
5.Testing the adequacy of the developed model
6.Result Analysis of the results.
7.Conclusions.

3.1 Identification of input parameters that influence the decision making of customer.

After closely studying the literatures available we have shortlisted the top three input parameters namely Price, Review and Ratings and they are identified as these are found to have the highest effect on customer decision making behaviour. To analyse the behaviour, we have selected one category on Amazon India. We will check the sales variation, which will help us understand the decision-making process.

For this category the input range is shown in table 1. The range is decided as maximum and minimum value amongst the top fifty bestsellers in this category.

Category	Price Range		Review Range		Rating Range	
	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
Smart Watches	499	4999	104	37631	3.7	4.5

Table 1: Input Parameters with maximum and minimum value.

3.2 Building a design matrix.

A Design matrix is made for this investigation and analyse the results in a systematically manner. We used *Design Expert 11* software to build a standard matrix as shown in table 2.

Std	Run	Factor 1 A:Price INR	Factor 2 B:Review Nos	Factor 3 C:Rating Star
2	1	1	-1	-1
3	2	-1	1	-1
7	3	-1	1	1
4	4	1	1	-1
1	5	-1	-1	-1
6	6	1	-1	1
8	7	1	1	1
5	8	-1	-1	1

Table 2: Standard design matrix

The value -1 indicates the minimum available value of parameter and value 1 indicates the maximum value of parameter. During the analysis the actual limit values are not controllable so the values are modified and are linearly interpreted. So the converted matrix will become as shown below. The formulae used for converting the value is:

Value=[2*(Factor Actual Value-Lower Range)/(Upper Range-Lower Range)]-1 [Equation 01]

3.3 Compilation of data as per matrix.

From this category a random dataset of 8 products is selected for this test. The Price, Review and Rating and last thirty-day sale data is available on amazon website. Using the available data and equation 01 the table 2 transforms as below.

Std	Run	Factor 1 A:Price INR	Factor 2 B:Review Nos	Factor 3 C:Rating Star	Response 1 R1
2	1	-0.6	0.5	0.15	11272
3	2	-0.47	-0.5	-0.92	9530
7	3	-0.6	-0.25	-0.8	9302
4	4	-0.78	-0.5	-0.81	9285
1	5	-0.56	0	-0.19	9008
6	6	-0.73	0	-0.84	8774
8	7	-0.64	0.25	-0.35	8335
5	8	-0.62	0.25	1	7307

Table 3: Transformed Design Matrix

3.4 Developing mathematical model.

We have assumed sales (represented by Response R1) in design matrix to be a function of Price, Review and Rating and this can be shown mathematically as

Y =f (A,B,C),

where Y is Sales
A is Price
B is Review
C is Rating

This regression equation can be written in form of a derivative polynomial as below:

$Y=b_0+b_1A+b_2B+b_3C+b_{12}AB+b_{23}BC+b_{31}CA.$ [Equation 2]

Here b_0 is model coefficient; b_1,b_2,b_3 are linear coefficient; b_{12}, b_{23}, b_{31} are interaction coefficient.

The final resulted equation is $Y = 10723.77 + 5874.9*A + 6981.5*B - 3002.68*C - 12379.54*ABC$ [Equation 3]

3.5 Adequacy of the model developed.

The ANNOVA analysis was carried out to test the adequacy of the working model with help of software. The results that were observed lies in satisfactory range of R2 as shown in table 4.

Std. Dev.	861.09	**R^2**	0.7498
Mean	9101.63	**Adjusted R^2**	0.4163

Table 4: Fit statistics table

In addition to the above R2 value the predicted Vs actual scattered graph as shown in Fig 2. which clearly shows that actual values is close to the central predicted line.

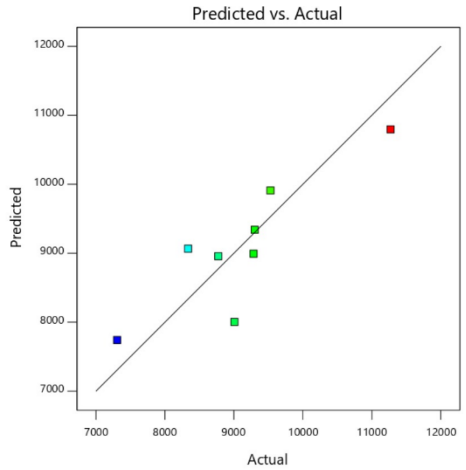

Fig 2. Predicted Vs Actual Scattered Graph

3.6 Analysis of Result

The Software was used to generate the 3D graphs which can easily describe the interaction of governing input parameters to the desired result. The graphs can be explained as below:

3.6.1 Interaction of Price and Review

The graph depicts that with increasing review magnitude the sales is increasing. For a same value of review with increase in price the sales are decreasing. This can be concluded that buying behaviour is driven by lower price.

With a higher review count even with increase in price the sale is observed to be increasing. This behaviour clearly shows that price is secondary, and the decision is driven by opinions of peer buyers.

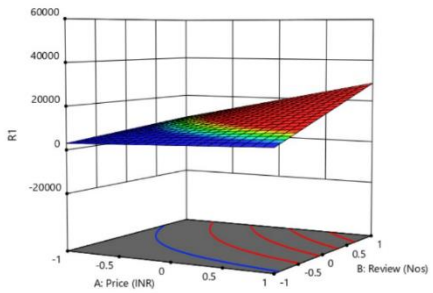

Fig 3. Interaction of Price and Review

3.6.2 Interaction of Price and Rating

With an increase in price the sale is dropped. Even with a increased rating magnitude the sale is found to be decreased with increase in price.

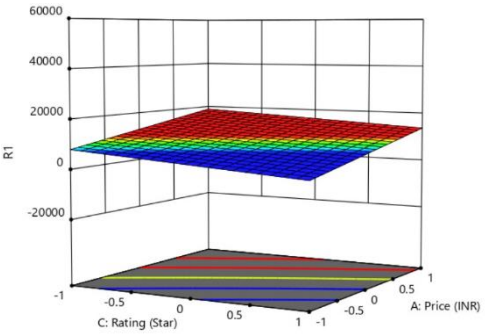

Fig 4. Interaction of Price and Rating

3.6.3 Interaction of Review and Rating

The increasing review is driving the growth in sales. As shown in the Fig 4 below the for a fixed rating with increase in review the sales follow the trend of review.

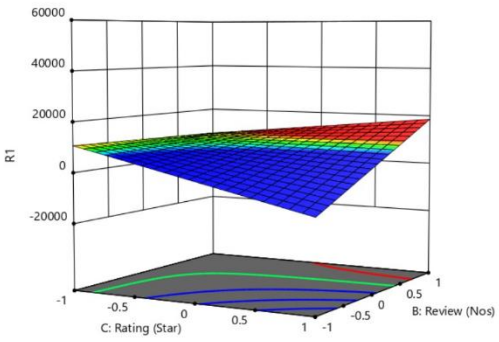

Fig 4. Interaction of Review and Rating

3.7 Conclusion:

The investigation shows that when the product is technical the price of the product does not have great effect of sales. It's the reviews count that is found to have the highest effect. The possible reason for this could be that when the product is technical the customer spends more time in making decisions, and the decisions is more calculated and has less sentiments associated when selecting the final products in this process they overlook the price and rating, they focus more on specification and opinions of peer buyers.
Below are some other conclusions that can be made from this investigation.

- The Full Factorial approach presented in this was found to be satisfactory.
- The effect of review count is found to be positive.
- Rating and price are contradictory here which has negative effect on sales. This is because of strong brand pull and technical specification.
- The Sale in smart watches category can be predicted using this equation 10723.77 + 5874.9*A + 6981.5*B -3002.68*C - 12379.54*ABC, Where A is Price, B is Review and C is Ratings

3.7 Acknowledgement

Special thanks to all the people who helped me throughout the process of this investigation. With the help of H.Kaur, equation 2 that was used to modify the design matrix was generated.

3.8 References

1. Satyendra, S., Chakraborti, S., & Jha, T. (2019). Analysis of book sales prediction at Amazon marketplace in India: A machine learning approach. Springer-Verlag GmbH Germany, part of Springer Nature.
2. Smirnov, P. S., & Sudakov, V. A. (Year). Forecasting new product demand using machine learning. 19th International Conference "Aviation and Cosmonautics" (AviaSpace-2020).
3. Bhaskar, A. K., Pal, K., & Khanna, P. (Year). Mathematical analysis of angular distortion on GTA welded hot rolled E250 grade low carbon steel plates. International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 03 | Mar 2020, pp-2273-2276.
4. Agnani, D., Bavkar, A., Salgar, S., & Ahir, S. (Year). Predicting E-commerce Sales & Inventory Management using Machine Learning. Ramrao Adik Institute of Technology, Department of Computer Science, Mumbai, India. International Conference on Automation, Computing and Communication 2022 (ICACC-2022), Vol 44.
5. Kumar, N., Anusara, J., Hossin, M. A., Sarkar, M. K., Chanthamith, B., Shah, S., & Russel, M. I. H. (Year). Challenges and Opportunities of E-Commerce in India: Pathway for Sustainable E-Commerce, International journal of Engineering, Business and Management (IJEBM) Vol-2, Issue-2, Mar-Apr, 2018 Page 13-21.
6. Aggarwal, I., Faujdar, N., Das, A., & Khanna, P. (2018). Mathematical Modeling For Predicting Angular Distortion In Tig Welding Of Stainless Steel 409L Butt Welds. International Journal of Research in Engineering and Technology, 07(06), 92-97.
7. Tiwari, I., Laksha, & Khanna, P. (2018). Mathematical Modelling of Angular Distortion in GTA Welded Low Carbon Alloy Steel Butt Welds. In Proceedings of the World Congress on Engineering and Computer Science (Vol. II, WCECS 2018, October 23-25), San Francisco, USA.