IJNRD.ORG ISSN : 2456-4184

INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG

An International Open Access, Peer-reviewed, Refereed Journal

International Research Journal IJNRD Research Through Innovation

Customer Segmentation Services Using Regression

Prof. Swati raut[1],

[1]Guide, Assistant Professor, Artificial intelligence Dept. JDCOEM, Nagpur

Ms.Sejal papalkar[2],Mr. gaurav butle [3], Mr. piyush mahalle[4], Mr. omkar vaidya[5]

[2,3,4,5] Author, UnderGraduate Scholar,Artificial intelligence Dept. JD College Of Engineering, Nagpur

Abstract-

A crucial tactic for companies trying to comprehend and successfully serve the wide range of wants of their customers is customer segmentation. This abstract presents the idea of customer segmentation and emphasizes the value of using regression analysis as an effective technique to do this. Businesses can identify unique consumer segments through the collection and analysis of customer data, which facilitates targeted marketing, product development, and customer interaction initiatives. This strategy improves customer experience overall, cost effectiveness, and targeting as well as client retention. Regression-based customer segmentation services give organizations the ability to make data-driven decisions that boost sales, strengthen client bonds, and give them a competitive edge in the fast-paced market of today.

keyword:Customer segmentation, Regression analysis, Data-driven, Marketing strategy, Product development Customer engagement, Targeting, Customer retention

1. Introductory

Success in today's fast-paced and cutthroat business environment depends critically on your ability to understand your clientele. Customer segmentation, which is the process of grouping customers according to common traits and behaviors, has become an essential tactic for businesses looking to customize their goods, services, and advertising to appeal to a wide range of consumer demands. Businesses use regression analysis, a statistical technique that enables them to find hidden patterns and relationships within their consumer data, to achieve effective customer segmentation.

By dividing their customer base into segments, companies can craft personalized marketing campaigns, optimize product offerings, and develop targeted customer engagement strategies. This not only enhances customer satisfaction but also contributes to increased customer retention rates and higher overall revenue. In an era where data is abundant, harnessing the potential of regression-based customer segmentation is essential for staying competitive and relevant in the ever-evolving marketplace. In the following sections, we will delve deeper into the intricacies of customer segmentation services using regression, exploring its methodologies, benefits, and real-world applications in the business landscape.

2. Study Deficit

It is evident from the literature analysis that a large number of previous studies have concentrated on customer behavior, brand awareness, purchase intention, and YouTube advertisements. Advertisers on YouTube increase their earnings 49% faster than those that do not use them. Segmentation is driven by YouTube ads. Additionally, additional variables such as the correlation between the YouTube advertising budget and sales have not been well characterized in other studies. Therefore, it is necessary to comprehend the connection between the budget for YouTube advertising and sales in social media advertisements. To improve the accuracy of our results, we are incorporating a new factor: the customer's budget.

3. Design of Research

Goal

To examine the connection between sales and previously acquired goods.

utilizing the training data set to construct a logistic regression model.

to use a logistic regression model to forecast how a test data set will be segmented.

4. Conjecture

Null Hypothesis: There isn't a statistically meaningful connection between the advertising budget and segmentation.

Alternative Hypothesis: Segmentation and budget for previously purchased items have a statistically significant relationship.

5.Setting Up the Data

The preloaded data set in the model used for this research's analysis is the Marketing dataset. This marketing dataset examines the impact of three different advertising mediums on sales: Facebook, YouTube, and newspapers. The predictor variable in this study is the budget for online marketing and YouTube advertising, whereas sales is the outcome variable. Sales are indicated by the dataset's final column, and media advertising in 1000 dollars is indicated by the remaining columns. The dataset has two hundred rows in it. Every row represents a distinct advertising experiment. Following data loading, a comment named head is used to examine a few representative samples from the dataset. Similarly, the tail comment is used to examine the end of the data setThis is done because the final few columns of some data sets may contain unrelated totals or summaries of the data. The data is then split into two sets: the training dataset (70%) and the testing dataset (30%), with 145 rows in the training dataset and 55 rows in the test dataset.

6.Understanding and analyzing data

Logistic regression, as used in data visualization, indicates a logistic relationship between the predictor and result variables.By creating a dot plot of the outcome variable (sales) against the predictor variable (YouTube advertising budget), a logistic relationship may be quickly confirmed. The relationship between the predictor and outcome variables can be shown by running the following R code.the marketing segmentation plot (marketing) and the model (col=2, lwd=3) Positively, the preceding graph shows that there is a growing logistic association between the predictor variable (budget, formerly Buyed & Age) and the outcome variable (segmentation). The following R code is used to determine the correlation coefficient between the outcome and predictor variables.

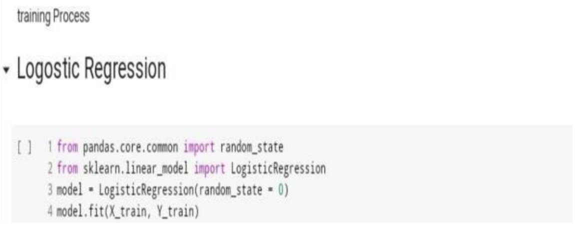
7.correlation value

The marketing segmentation and marketing $YouTube) correlation coefficient cor is 0.7822244.The correlation coefficient quantifies the degree of relationship between the predictor factors and the outcome. A perfect negative link between variables is shown by a correlation coefficient value of -1, and a perfect positive association is indicated by a value of +1. If the value is closer to zero, the two variables have a weaker link. The correlation coefficient in this study is 0.78, indicating that the predictor and outcome variables have a very strong positive association.

8.Model of Logistic Regression

In this paper, segmentation units are predicted using a straightforward logistic model based on YouTube advertising budgets. The model's beta coefficients can be obtained by running the following code. model. <lm (market train data, sales~YouTube)

Values of Coefficients



```
from pandas.core.common import random_state
from sklearn.linear_model import LogisticRegression
model = LogisticRegression(random_state = 0)
model.fit(X_train, Y_train)
```

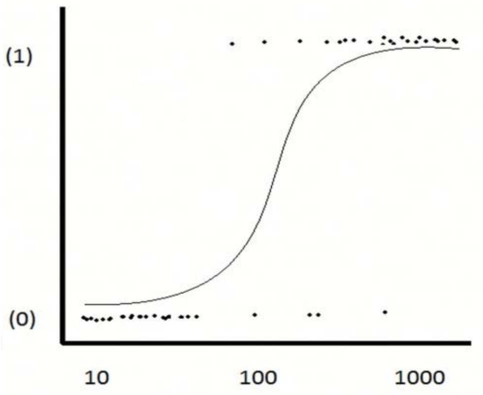


Fig. 1 Logistic Regression

The predictor variable's intercept and beta coefficient are shown in the value above. The computed equation is therefore sales = 7.82 + 0.051*PPI. THE variable's coefficient is 0.051, and the intercept (b0) is 7.82. This formula can be used to anticipate the number of sales units for each new YouTube advertising expenditure. The projected sale with a $0 YouTube advertising budget is 7.82 units. The projected sale for a $1,000 YouTube advertising spend is 7.82 + 0.051 * 1000 = 58.8 units.

9.Synopsis of the Model

By displaying the statistical summary of the model, the statistical significance of the model is evaluated prior to applying it for predictions.Before a predictive model is used to make predictions in the actual world, it is essential to evaluate its statistical significance and dependability using the model summary. It includes the kind of model that is used, the data source and preprocessing procedures, the model fit assessment using suitable metrics, the in-depth analysis of predictor variable coefficients and their significance, the validation of model assumptions, a careful look at residuals for any trends or assumptions that are broken, and a final appraisal of the model's performance. This summary acts as an essential quality control point, guaranteeing that the model is statistically sound and prepared for efficient application in forecasting and decision-making.

10.Understanding

It is possible to ascertain whether or not the null hypothesis may be rejected using the p-value of the t-static. The more important the predictor, the higher the t-statistic and the lower the p-value. The intercept and predictor variable p-values in this study are both very significant. As a result, there is a substantial correlation between the Segmentation outcome variable and the predictor variable, YouTube advertising. Thus, zero, the hypothesis is disproved. After the null hypothesis is disproved, it is typical to assess the logistic regression model's quality of fit. Regression fit's logistic quality is commonly assessed using two related metrics: the R-squared statistic and residual standard error (RSE). The residual standard error for the training data set in this study is 3.87.This indicates that there is an average deviation of 3.87 units between the Segmentation value and the real regression line. The average sales value in this training data set is 16.83, meaning that the percent error is 3.87/16.83=22%, indicating a reasonable residual error. Thus, the model better matches the data by lowering the RSE. In this study, the R square value is 0. 8684. A higher R-squared model will be superior. The R-squared number normally ranges from 0 to 1. Because of its strong R-squared value, the research's logistical regression model performs accurately.

Customer Segmentation

Using test data to make predictions, the regression model's performance is assessed. fresh data <- data. frame (c(0,1000) = YouTube) anticipate (new data, model)

In the test datasets, the model and the observed outcome value have a strong correlation. 4.06 is the computed Root mean square value for the test dataset. With 4.06 times the mean, or 24% of the test dataset, the error rate is negligible. Thus, the model's accuracy is high.

11. Resulting

The budget for YouTube advertising and sales are logistically and additively related.

The outcome variable and the predictor variable have a substantial relationship.

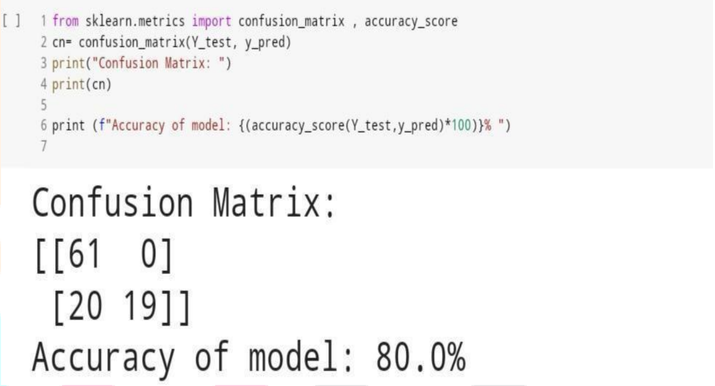
The previously purchased items budget and sales have a statistically significant link.

The research's logistic regression model matches the data extremely well.

The logistic regression model predicts that an increase of 58 units in sales will occur for the $1,000 YouTube advertising spend.

The expenditure budget will likewise drop significantly if we can narrow down the consumer pool based on previously bought goods.

12. Confusion Matrix



```
[ ]
from sklearn.metrics import confusion_matrix , accuracy_score
cn= confusion_matrix(Y_test, y_pred)
print("Confusion Matrix: ")
print(cn)

print (f"Accuracy of model: {(accuracy_score(Y_test,y_pred)*100)}% ")

```

```
Confusion Matrix:
[[61  0]
 [20 19]]
Accuracy of model: 80.0%
```

Fig 2. Confusion Matrix

TABLE I. OVERVIEW OF ISSUES APPROACHES,METHODS

serial no.	Identify	Problem Spoken	Strategy And Technique
I.	Bhattacharya, R.K., and Bose, I. [4]	machine learning from a business data mining standpoint.Data and administration	- This study compares the forecasting accuracy of several linear and nonlinear models for total retail sales.
II.	Foster J. Provost, Fawcett, and Tom [14]	Effective User Profiling with the Integration of Data Mining and Machine Learning.	- A mathematical approach for reliable production planning is presented in this study. - The model assists fashion apparel suppliers in allocating production orders to various facilities with varying lead times and manufacturing costs, as well as in scheduling and sequencing these production orders in a timely manner.
III.	Rekha Ashwini. B[5] Banjanagari, Vijaykumar	"Machine Learning for Retail Giant Sales Forecasting"	- While creating successful regression models for forecasting retail sales can benefit from the use of seasonal dummy variables. - Dummy regression models' performance might not be reliable.
IV.	Norio, Hidetsugu Suto, and Baba.[12]	"using GAs and artificial neural networks to build a sophisticated sales prediction system."	- The majority of PSA literature focuses on creating predictive models for particular sales-related choices, including calculating the likelihood that a lead would convert.

TABLE II: Synopsis OF DIFFERENT TRAJECTORY PREDICTION APPROACHES

serialnumber. .	**Call out**	**Problem Spoken**	**Estimating trajectory prediction technique**	**criteria**	**Qualities**
I	Weber, R., and Aburto, L. [1]	enhanced supply chain administration through the use of hybrid demand projections.	trajectories represented by a probabilistic model that accounts for all potential future consumer behaviors.	Customer segmentation and distance from intersection	Numerous aspects are taken into account, including variations in output and the stochastic nature of customer demand.
II	Thoben, K.D., Karimi, H.R., and Beheshti-Kashi, S.[3]	An analysis of fashion markets' retail sales forecasts and predictions.	Demand is particularly unpredictable in consumer-oriented industries like electronics and fashion.	A crucial duty in retailing is sales forecasting.	Additionally, this study examines several approaches to the predictive power of user-generated content and search terms.
III	Suniti Yadav, Khushbu Kumari,[6]	"Clinical Cardiology Statistics Curriculum, "	A component of logistics management is retail.,	We discover that in out-of-sample situations, the nonlinear models can perform better than their linear equivalents.	Time-series data, such as retail sales data, differ from the data often utilised in general regression analysis in a few key ways.
IV	Michael Giering[11]	"Predicting retail sales and suggesting items based on consumer demographics at the shop level	The goal is to use CVaR to maximise the production's overall profit.	We suggest a risk-constrained profit-expected maximisation model in the garment industry to address this unpredictability in client demand.	The production plan generated by this model takes into account the stochastic nature of client demand and specifies the quantities of each product.

13. In summary

For many marketers, YouTube is their primary marketing platform because to its incredible growth in the last ten years in social media advertising. Sales, market share, and brand awareness can all be increased online. Online branding has always been the domain of marketers. They have since taken the time to comprehend how YouTube ads affect sales. This analysis clearly shows that there is a substantial correlation between sales and the amount spent on YouTube advertising. According to this study, YouTube advertising is a more accurate indicator of a company's segmentation. As we narrow down the candidates and clients, our overall budget for YouTube and web ads will decrease swiftly.

REFERENCES

[1] Aburto, L., & Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. Applied Soft Computing, 7(1), 136–144. Retrieved July 27, 2014, from http://linkinghub.elsevier.com/retrieve/pii/S1568494605000311. doi: 10.1016/j.asoc.2005.06.001 [Crossref], [Web of science ®]Saquib HashmiSaquib Hashmi, KaushtubhSaquib Hashmi, Kaushtubh Saquib Hashmi, Kaushtubh Kumar, Siddhant Khandelwal, Real-Time License Plate Recognition from Video Streams using Deep Learning, International Journal of Information Retrieval Research, (January 2019).

[2] Ait-alla, A., Teucke, M., Lütjen, M., Beheshti-Kashi, S., & Karimi, H. R. (2014). Robust production planning in the fashion apparel industry under demand uncertainty via conditional value at risk. Mathematical Problems in Engineering, 2014, 10 pp.https://www.hindawi.com/journals/mpe/2014/901861

[3] Beheshti-Kashi, S., Karimi, H.R., Thoben, K.D., Lutjen, M., Teucke, M.: A survey on retail sales forecasting and prediction in fashion markets. Systems Science & Control Engineering 3(1),154–161(2015)https://www.researchgate.net/publication/271205922_A_survey_on_retail_sales_forecasting_and_predi ction_in_fashion_markets

[4] Bose, I., Mahapatra, R.K.: Business data mining machine learning perspective.Information & management 39(3), 211–225 (2001)3. Chu, C.W., Zhang, G.P.: A comparative study of linear and nonlinear models for aggregate retail sales forecasting. International Journal of production economics 86(3),217–231(2003)https://www.researchgate.net/publication/271205922_

[5] AshwiniRekha. Banjanagari, Vijaykumar. B, "Retail Giant Sales Forecasting using Machine Learning", International Journal of Recent Technology and Engineering (IJRTE), Vol. 8,Pgno.2408–2411.https://pubs.aip.org/aip/acp/article-abstract/2387/1/140038/1000110/Sales-prediction-using-machine learning

[6] Khushbu Kumari, Suniti Yadav, "Curriculum In Cardiology – Statistics", Journal of the practice of Cardiovascular Science. Vol-4, Issue-1, Pgno. 33–36. https://doi.org/10.4103/jpcs.jpcs_8_18

[7] Singh Manpreet, Bhawick Ghutla, Reuben Lilo Jnr, Aesaan FS Mohammed, and Mahmood A. Rashid. "Walmart's Sales Data Analysis-A Big Data Analytics Perspective." In 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), pp.114-119.IEEE,2017.https://ieeexplore.ieee.org/document/8487274

[8] Sekban, Judi. "Applying machine learning algorithms in sales prediction." (2019).

[9] Panjwani, Mansi, Rahul Ramrakhiani, Hitesh Jumnani, Krishna Zanwar, and Rupali Hande. Sales Prediction System Using Machine Learning. No. 3243. EasyChair, 2020.https://easychair.org/publications/preprint/fWt8

[10] Cheriyan, Sunitha, Shaniba Ibrahim, Saju Mohanan, and Susan Treesa. "Intelligent Sales Prediction Using Machine Learning Techniques." In 2018 International conference on Computing, Electronics & Communications Engineering (iCCECE), pp. 53-58. IEEE, 2018. https://www.researchgate.net/profile/Sunitha-Cheriyan/publication/331606166_Intelligent_Sales_Prediction_Using _Machine_Learning_Techniques

[11] Giering, Michael. "Retail sales prediction and item recommendations using customer demographics at store level." ACM SIGKDD Explorations Newsletter 10, no. 2 (2008): 84-89. https://dl.acm.org/doi/10.1145/1540276.1540301

[12] Baba, Norio, and Hidetsugu Suto. "Utilization of artificial neural networks and GAs for constructing an intelligent sales prediction system." In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, vol. 6, pp. 565-570. IEEE, 2000. https://issuu.com/irjet/docs/irjet-v7i6676/s/11461029

[13] Ragg, Thomas, Wolfram Menzel, Walter Baum, and Michael Wigbers. "Bayesian learning for sales rate prediction for thousands of retailers." Neurocomputing 43, no. 1-4 (2002): 127-144. https://www.sciencedirect.com/science/article/abs/pii/S0925231201006245 Fawcett, Tom, and Foster J. Provost. "Combining Data Mining and Machine Learning for Effective User Profiling." In KDD,pp.8-13.1996.https://www.semanticscholar.org/paper/Combining-Data-Mining-and-Machine-Learning-for-User-Fawcett-Provost/4f582a0

[14] Datamonitor, Global Apparel Retail, 2010, http://www.datamonitor.com/. R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss

distributions," Journal of Banking and Finance, vol. 26, no. 7, pp. 1443–1471, 2002. View at: Publisher Site | Google Scholar

[15] S. Zhu and M. Fukushima, "Worst-case conditional value-at-risk with application to robust portfolio management," Operations Research, vol. 57, no. 5, pp. 1155–1168, 2009. View at: Publisher Site | Google Scholar | Zentralblatt MATH | MathSciNet

[16] M. Teucke and B. Scholz-Reiter, "Improving order allocation in fashion supply chains using radio frequency identification (RFID) technologies," in Fashion Supply Chain Management Using Radio Frequency Identification (RFID) Technologies, W. K. Wong and Z. X. Guo, Eds., vol. 152 of Woodhead Publishing Series in Textiles, Woodhead Publishing, Cambridge, UK, 2014 . View at: Google Scholar

[17] S. Thomassey, "Sales forecasts in clothing industry: the key success factor of the supply chain management," International Journal of Production Economics, vol. 128, no. 2, pp. 470–483, 2010. View at: Publisher Site | Google Scholar