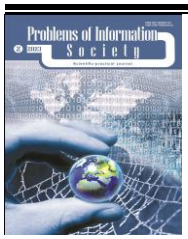

www.jpis.az

14 (2)
2023

Customer behavior analysis using big data analytics and machine learning

Leyla G. Muradkhanli[1], Zaman M. Karimov[2]

[1,2]Computer Science Department, Khazar University, Mahsati str. 41, AZ1096, Baku, Azerbaijan

[1]leyla@khazar.org, [2]zaman.karimov@khazar.org

[1]orcid.org/0000-0001-6149-4698

ARTICLE INFO

http://doi.org/10.25045/jpis.v14.i2.08

Article history:
Received 10 February 2023
Received in revised form
12 April 2023
Accepted 16 June 2023

Keywords:
Big data
Customer behavior analysis
Machine learning
Artificial intelligence
Customer lifetime value
Big data analytics

ABSTRACT

This paper delves into the utilization of big data analytics and Machine Learning (ML) in the realm of customer behavior analysis for digital marketing. It explores the practical application of ML algorithms and the ML pipeline in the development of predictive models. The primary objectives revolve around forecasting customer churn, identifying prospects with a high propensity to convert, determining optimal communication channels, and leveraging sentiment analysis to enhance the overall customer experience. Concrete real-world examples and compelling case studies are employed to illustrate the efficacy of ML in analyzing customer behavior. Moreover, the paper acknowledges the existing limitations and challenges in this domain, while also outlining potential directions for future research. By offering a comprehensive guide, the aim is to empower businesses with the knowledge and tools needed to effectively leverage big data analytics and ML for customer behavior analysis in the digital marketing landscape. The paper concludes by addressing limitations, challenges, and future research directions in this field, aiming to provide a comprehensive guide to leveraging big data analytics and ML for customer behavior analysis.

1. Introduction

In today's competitive marketplace, businesses face a growing need to understand and engage with their customers effectively. With the increasing amount of data being generated by customers through online and offline interactions, businesses have access to an unprecedented level of information about their customers. However, this data can be overwhelming and difficult to analyze, making it challenging for businesses to extract meaningful insights. Every person is unique and has various habits and personality features. Customer behavior is often consistent. We base our buying or not buying decisions on our way of life, past experiences, and emotions. No matter if it is a tiny neighborhood bakery or a massive global network of supermarkets, it is wise to understand who the clients are.

Machine Learning (ML) is useful in this situation. Traditional marketing strategies themselves became ineffective because of the expansion of digital platforms and the digitization of business. This does not imply that ML rewrites the principles of marketing and customer behavior research, but it does provide new tools and insights.

The use of ML and big data analytics can help businesses overcome these challenges and gain a deeper understanding of customer behavior. ML algorithms can analyze large and complex datasets to identify patterns and relationships, allowing businesses to make predictions about future customer behavior. Additionally, big data analytics can provide businesses with real-time insights into customer preferences and needs, enabling them to

make data-driven decisions that improve customer engagement (CE) and satisfaction.

In latest years, businesses all over the world have aggressively begun using new ML techniques to increase their competitiveness in the client acquisition market. Because of the expanding amount of data and widespread access to high-performance computing and cloud services, ML has enabled businesses to greatly improve the customer experience.

To benefit from ML models, the early investors needed to invest heavily in pricey information technology (IT) infrastructure, human resources, and significant budgets. However, the advantages of modern digital trends for small enterprises became a reality with the introduction of cloud technology and subscription-based services.

In this article, we explore the prediction of customer lifetime value (CLV) using advanced techniques and tools. Specifically, we employed Python as our primary programming language, utilizing its wide range of libraries and frameworks to perform the necessary data analysis and modeling tasks. Among these, we employed the Lifetimes package, a powerful tool developed by Cameron Davidson-Pilon, which offers methods such as the Beta-Geometric Negative Binomial Distribution (BG-NBD) model and GammaGammaFitter for estimating customer lifetime value. By leveraging these techniques, we aim to provide valuable insights into customer behavior and enable businesses to make informed decisions to maximize customer value and optimize their strategies.

2. Literature Review

The emergence of ML has sparked extensive scholarly discourse on various aspects pertaining to customer behavior analysis using analytics and ML. Within the academic community, numerous publications have delved into this subject matter, addressing a wide range of relevant topics and methodologies. In the following section, a selection of notable works in the field is presented, showcasing the breadth and depth of research undertaken by scholars and practitioners alike. Chien-Chang Hsu et al. (2004) proposed the paragraph about intelligent interface for customer behavior analysis in electronic commerce. The interface comprises three modules: the task editor, action supervisor, and behavior analyzer. The task editor allows system administrators to define business tasks and domain ontology. The action supervisor monitors customer operations, filters unnecessary actions, and identifies behavior patterns using interaction messages, Bayesian belief network, and Radial Basis Function (RBF) neural networks. The behavior analyzer generates customer behavior analysis information by measuring behavior patterns, constructing personalized domain ontology, and assessing customer skill proficiency. Pin-Liang Chen et al. (2017) proposed a study that aims to analyze the traits and possibility of making purchases of various customer groups in the Shimen shopping district in order to derive business value. The main idea of this paragraph is that the rising popularity of social networking services and mobile devices has brought new business challenges. Users' Facebook profiles and use data from a point-earning app were utilized in the study to conduct statistical analyses with a CBAS. The findings indicated that younger consumers made up the majority of the retail district's patrons and that different age and gender groupings had varying preferences. The study also showed that altering the assignments and promotions on the point-earning app based on the investigation's findings significantly increased the conversion rate. Kailash Hambarde et al. (2020) conducted investigations on implementing data analytics on a Turkey-based e-commerce company's data to classify customer behavior patterns. They used an Artificial Neural Network (ANN) model to forecast customer purchasing patterns, which can benefit the marketing department in recognizing targeted customers for specific campaigns. The ANN model using the back-propagation technique showed high accuracy in predicting customer behavior. The study was conducted in R programming environment. Stavros Anastasios Iakovou's et al. (2016) work presents a prediction model based on customer behavior using data mining techniques. The model utilizes data from a supermarket database and an additional database from Amazon to classify customers and products. The model is trained and validated with real data and is intended to be used as a tool for marketers to analyze and target consumer behavior. Meshal Alduraywish et al. (2022) discussed the competition in the fast fashion industry, specifically the emergence of online-only retailing and the use of AI and ML by these companies to enhance the online customer shopping experience. It also highlights the concern of multichannel companies regarding the performance of their physical stores and the potential for these stores to enhance customer experience. The paper aims to explore the

application of AI in ecommerce by fast fashion companies and the impact of online-only business on physical stores, as well as the future role of physical stores in the industry.

3. Marketing: Reaching customers with digital technology

Marketing refers to the process of identifying, anticipating, and satisfying customer needs and wants through the creation, promotion, and distribution of products and services (Kühl et al., 2020). It encompasses a range of activities, including market research, product development, pricing, advertising, and sales. The ultimate goal of marketing is to build long-term relationships with customers by delivering value and meeting their needs in a profitable way. Digital marketing, on the other hand, refers to the use of digital technologies, such as the internet, social media, email, and mobile devices, to promote products and services (da Silva Wegner et al, 2023). It is a subset of marketing that leverages the power of digital channels to reach and engage with customers in new and innovative ways. Digital marketing encompasses a range of tactics, including search engine optimization (SEO), pay-per-click (PPC) advertising, social media marketing, email marketing, content marketing, and more. As big data is becoming increasingly important in digital marketing as marketers seek to leverage data-driven insights to improve their strategies and tactics. By using big data effectively, digital marketers can gain a competitive advantage and better meet the needs of their customers (Ducange, et al., 2018).

4. ML for customer behavior analysis: predicting CLV use case

4.1 The problem statement

The main idea of paper is to provide an overview and assessment of the techniques and models used in a project related to customer lifetime value (CLV) analysis based on dataset that shows the sales of a British online store between 01/12/2010 and 09/12/2011 date range (https://www.kaggle.com/datasets/carrie1/ecommerce-data). The problem that stands behind of this research is that, suppose an e-commerce company that wants to segment its customers and determine marketing strategies according to these segments by the start of next year. These are the steps that should be executed in program:

- Data preprocessing
- Expected Sales Forecasting with The Beta-Geometric Negative Binomial Distribution (BG-NBD) Model
- Expected Average Profit with Gamma-Gamma (GG) Model
- Calculation of CLV with BG-NBD and GG Model
- Creating Segments by CLV

4.2 The solution

Using The Lifetime package for data preprocessing and formatting we can prepare data for further development. The Lifetimes package, developed by Cameron Davidson-Pilon, is a powerful tool used to create a CLV model. This package provides functions and methods for estimating customer lifetime metrics, such as customer lifespan and purchase frequency, based on transactional data. By utilizing the Lifetimes package, we can gain valuable insights into customer behavior and make informed decisions to enhance customer relationships and optimize business strategies. The solution also emphasizes the utilization of probabilistic approaches through the BetaGeoFitter and GammaGammaFitter models to model customer behavior. The paragraph further delves into the details of the BG-NBD model, explaining its suitability for analyzing customer purchase behavior and the role of the geometric and Poisson distributions in estimating customer "lifetime" and purchase frequency:

```
bgf = BetaGeoFitter(penalizer_coef=0.001)
bgf.fit(cltv_df['frequency'], cltv_df['recency'],
cltv_df['T'])
```
Output: <lifetimes.BetaGeoFitter: fitted with 2845 subjects, a: 0.22, alpha: 12.19, b: 3.08, r: 2.23>

The GammaGammaFitter model is introduced as a tool for estimating average transaction value. The paragraph also mentions the significance of data segmentations and z-scores in identifying high-value customers and tailoring marketing strategies accordingly:

```
ggf = GammaGammaFitter(penalizer_coef=0.01)
ggf.conditional_expected_average_profit(cltv_df
['frequency'],
cltv_df['monetary']).head(5)
```

Calculating CLV and dividing clients into segments is a crucial step in CLV analysis. By identifying high-value customers, low-value customers, and everything in between, businesses

can make more informed decisions about marketing, sales, and customer retention strategies.

To calculate CLV using the BG-NBD and GG model in Python, we first fit the BG-NBD model to the data to predict the expected number of transactions for each customer. We then fit the GG model to the data to predict the expected average profit per transaction for each customer. Using these two models, we calculate the CLV for each customer:

```
cltv_df["expected_average_profit"]=ggf.conditiona
l_expected_average_profit(cltv_df['frequency'],
cltv_df['monetary'])
cltv_df.sort_values("expected_average_profit",asce
nding=False).head(5)
```

To standardize CLV values, we use z-scores. The resulting z-scores represent the number of standard deviations that each customer's CLV value is from the mean:

```
scaler = MinMaxScaler(feature_range=(0, 1))
scaler.fit(cltv_final[['clv']])
cltv_final['scaled_clv'] =
scaler.transform(cltv_final[['clv']])
cltv_final.sort_values(by="scaled_clv",
ascending=False).head()
```

After standardizing the CLV values, we can divide customers into segments based on their z-score. For example, customers with z-score greater than 1 may be considered high-value customers, while customers with z-score less than -1 may be considered low-value customers:

```
cltv_final["segment"] =
pd.qcut(cltv_final["scaled_clv"], 4, labels=["D", "C",
"B", "A"])
```

The Figure 1 presents the overall flowchart of provided Python code.

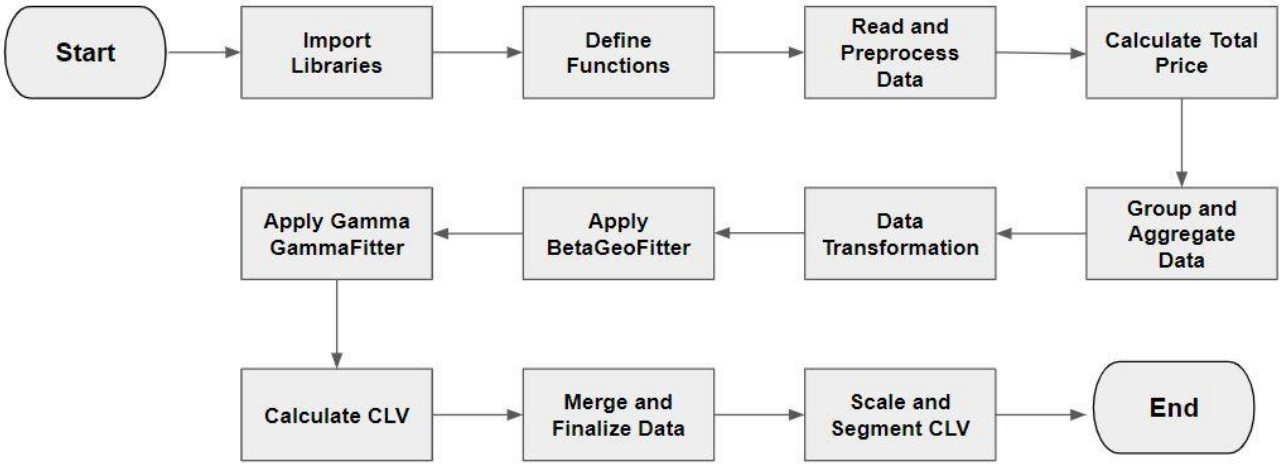

Fig. 1. The flowchart of source code

5. Results

As a result of written script, we can make a conclusion that, the use of The Lifetime package suggests preprocessed and formatted data to make it suitable for analysis. The use of BetaGeoFitter and GammaGammaFitter models implies modeled customer behavior using probabilistic approaches.

The BG-NBD model is particularly well-suited for analyzing customer purchase behavior, as it models the number of purchases made by a customer over time using a Poisson distribution, while modeling the probability that a customer is still "alive" (i.e., still making purchases) using a geometric distribution. The GammaGammaFitter model can then be used to estimate the average transaction value for each customer, which can be used in conjunction with the BG-NBD model to estimate the expected CLV for each customer.

The use of data segmentations and calculation of z-scores for customers is also noteworthy, as these techniques can help identify high-value customers and tailor marketing strategies accordingly. By segmenting customers based on their behavior and calculating their z-scores, businesses can identify customers who are most likely to make repeat purchases and are therefore likely to have a high CLV.

The results of proposed solution are displayed in the Table 1.

Table 1. Result of proposed solution

Customer ID	recency	T	frequency	monetary	expected_purc_1_week	expected_purc_1_month	expected_average_profit	clv	scaled_clv	segment
14646.0	50.428571	53.714286	74	3596.804392	1.009916	4.019927	3605.309159	45665.018393	1.000000	A
18102.0	52.285714	55.571429	60	3859.739083	0.850075	3.384130	3870.996702	41290.049223	0.904194	A
12415.0	44.714286	51.285714	21	5724.302619	0.325761	1.296453	5772.177190	23567.876738	0.516104	A
17450.0	51.285714	55.571429	46	2863.274891	0.641621	2.554284	2874.198462	23139.994099	0.506733	A
14096.0	13.857143	17.571429	17	3163.588235	0.559501	2.214638	3196.435385	21993.530135	0.481628	A

Overall, this project appears to be a well-designed and well-executed application of ML techniques to customer lifetime value analysis. The insights gained from this analysis can help businesses identify their most valuable customers and develop targeted marketing strategies to maximize their CLV.

By using this technique, businesses can better understand their customers and make more informed decisions about marketing, sales, and customer retention strategies.

6. Discussion

In this paper, we have presented a comprehensive analysis of customer behavior using big data analytics and ML techniques. However, there is always room for further research and improvement. There are some potential directions for future research in this field.

Firstly, one potential future work could be to explore the application of deep learning techniques in customer behavior analysis. Deep learning is a subset of ML that uses neural networks to model and solve complex problems (Oh et al., 2022). In recent years, deep learning has achieved significant success in image recognition, NLP, and speech recognition, among other fields. Applying deep learning techniques in customer behavior analysis could lead to more accurate and nuanced insights into customer behavior.

Secondly, a promising area for future research is to investigate the use of data visualization techniques in customer behavior analysis. Data visualization techniques such as heat maps, scatter plots, and line charts can help to uncover hidden patterns and insights in large datasets. By leveraging data visualization techniques, it may be possible to gain a better understanding of customer behavior and develop more effective marketing strategies (Li, 2022).

Thirdly, an interesting direction for future research could be to explore the use of unsupervised learning techniques in customer behavior analysis. By using unsupervised learning techniques, it may be possible to uncover hidden relationships between customer behavior and other variables that were not previously considered (Singh et al., 2020).

Another strategy is to apply real-time analysis. The ability to analyze customer behavior in real-time could enable businesses to respond quickly to changing customer needs and preferences. Future

work could explore methods for performing real-time analysis on large volumes of customer data, such as using streaming data processing techniques and ML algorithms that are optimized for speed and scalability (Walk et al., 2001).

Lastly, another interesting area for future research could be to explore the ethical implications of customer behavior analysis. As customer behavior data becomes more prevalent, there is a growing concern about privacy and data protection. Therefore, it is important to investigate the ethical implications of customer behavior analysis and develop ethical guidelines for the collection, analysis, and use of customer data (O'Leary et al., 2017).

7. Conclusion

In conclusion, customer behavior analysis using big data analytics and ML techniques holds immense potential for businesses across various industries. This field of study has gained significant attention and traction in recent years due to its ability to extract valuable insights from large volumes of customer data, enabling businesses to make informed decisions and drive strategic initiatives.

Through the integration of big data analytics and ML, businesses can uncover patterns, trends, and hidden correlations in customer behavior that were previously challenging to identify using traditional methods. The combination of advanced analytics algorithms, scalable computing power, and vast amounts of available data has revolutionized the way customer behavior is understood and leveraged for business advantage (Bhardwaj et al., 2023).

By harnessing big data analytics, businesses can capture and process diverse data types, including transactional data, web browsing behavior, social media interactions, and customer feedback. These data sources enable a comprehensive understanding of customer preferences, needs, and sentiments. ML algorithms play a pivotal role in transforming raw data into actionable insights, allowing businesses to predict customer behavior, personalize marketing campaigns, optimize pricing strategies, enhance customer experience, and ultimately, drive revenue growth.

The application of ML in customer behavior analysis has resulted in several significant advancements. For example, customer segmentation can now be performed with greater precision, allowing businesses to target specific customer groups and tailor their marketing efforts accordingly. Churn prediction models have become more accurate, enabling proactive retention strategies and minimizing customer attrition. Sentiment analysis techniques have provided deeper insights into customer sentiment, facilitating sentiment-based marketing and brand management.

Moreover, the use of ML in customer behavior analysis has demonstrated tangible benefits for businesses. Improved customer acquisition and retention rates, enhanced marketing campaign effectiveness, increased customer satisfaction, and higher customer lifetime value are just a few examples of the positive impacts observed in real-world implementations.

However, it is crucial to acknowledge that challenges and considerations exist in this field. Privacy concerns, ethical considerations, data quality issues, and the need for skilled data scientists are among the challenges that must be addressed to maximize the potential of customer behavior analysis using big data analytics and ML.

In conclusion, customer behavior analysis using big data analytics and ML represents a transformative approach to understanding and predicting customer behavior. By leveraging the power of advanced analytics and large-scale data processing, businesses can gain a competitive edge, optimize their marketing strategies, and foster long-term customer relationships. As technology continues to evolve, the potential for customer behavior analysis will only expand, offering exciting opportunities for businesses to unlock the full value of their customer data.

References

Bhardwaj, S. et al. (2023). Proposing an integrative data-analytics framework for micro, small and medium enterprises: a systematic review substantiated by evidence from two case studies. Annals of Operations Research. https://doi.org/10.1007/s10479-023-05186-9

Chien-Chang Hsu et al. (2004). An Intelligent Interface for Customer Behaviour Analysis from Interaction Activities in Electronic Commerce. Innovations in Applied Artificial Intelligence, 315-324. https://doi.org/10.1007/978-3-540-24677-0_33

da Silva Wegner et al. (2023). Performance analysis of social media platforms: evidence of digital marketing. Journal of Marketing Analytics. https://doi.org/10.1057/s41270-023-00211-z

Ducange, P. et al. (2018). A glimpse on big data analytics in the framework of marketing strategies. Soft Computing 22(1), 325–342. https://doi.org/10.1007/s00500-017-2536-4

https://www.kaggle.com/datasets/carrie1/ecommerce-data

Kailash Hambarde et al. (2020). Augmentation of Behavioral Analysis Framework for E-Commerce Customers Using MLP-Based ANN. Advances in Data Science and Management, 45-50. https://doi.org/10.1007/978-981-15-0978-0_4

Kühl, N. et al. (2020). Supporting customer-oriented marketing with artificial intelligence: automatically quantifying customer needs from social media. Electron Markets 30, 351–367. https://doi.org/10.1007/s12525-019-00351-0

Li, H. (2022). Intelligent business framework for interactive data visualization of small and medium-sized enterprises in developing countries. Annals of Operations Research, 1-17. https://doi.org/10.1007/s10479-021-04513-2

Meshal Alduraywish et al. (2022). Application of Artificial Intelligence in Recommendation Systems and Chatbots for Online Stores in Fast Fashion Industry. Proceedings of the International Conference on Intelligent Vision and Computing (ICIVC 2021), 558–567. https://doi.org/10.1007/978-3-030-97196-0_46

O'Leary, P.N., et al. (2017). Blurred Lines: Ethical Implications of Social Media for Behavior Analysts. Behavior Analysis in Practice 10, 45–51. https://doi.org/10.1007/s40617-014-0033-0

Oh, S., Ji, H., Kim, J. et al. (2022). Deep learning model based on expectation-confirmation theory to predict customer satisfaction in hospitality service. Information Technology & Tourism 24, 109–126. https://doi.org/10.1007/s40558-022-00222-z

Pin-Liang Chen et al. (2017). Social Network and Consumer Behavior Analysis: A Case Study in the Shopping District. Frontier Computing, 879–890. https://doi.org/10.1007/978-981-10-3187-8_84

Singh, N. et al. (2020). An inclusive survey on machine learning for CRM: a paradigm shift. Decision 47, 447–457. https://doi.org/10.1007/s40622-020-00261-7

Stavros Anastasios Iakovou et al. (2016). Customer Behavior Analysis for Recommendation of Supermarket Ware. Artificial Intelligence Applications and Innovations, 471–480. https://doi.org/10.1007/978-3-319-44944-9_41

Walk, N. et al. (2001). Real-time database analysis: Customer knowledge as a value-determining factor in e-commerce. Journal of Database Marketing & Customer Strategy Management 8, 143–149. https://doi.org/10.1057/palgrave.jdm.3240029