Volume 9, Issue 4, April – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR900

Customer Classification by Past Purchase Data Analysis

Sangeetha K[1]; Lauvanya R[2]; Dharshini I[3]; Amizhthini A[4]
Department of Artificial Intelligence and Machine Learning, Rajalakshmi Engineering college, Chennai, India.

Abstract:- Customer categorization is an essential strategy for companies seeking to maximize their advertising spend. Businesses can boost client involvement and sales rates of conversion substantially by identifying specific customer segments, tailoring products or services to their preferences, minimizing the hassle of irrelevant advertisements, and increasing customer satisfaction, resulting in improved long-term interactions with clients. This paper presents a classification model that uses Keras and support vector machine stacked classification passed on to a meta-learner to predict the customer segment, and RFM analysis is performed to identify the customer segment. This focused strategy lowers marketing costs and boosts income, increasing the business's efficiency. Temporal mining helps us predict the next purchase of a customer using a time series model.

Keywords:- *Customer Classification, RFM Analysis, Keras And Support Vector Machine* .

I. INTRODUCTION

Customer classification is a vital approach for companies looking to maximize their marketing efforts. Companies can significantly improve customer engagement and sales conversion rates by identifying specific customer segments and tailoring products or services to their customers. Using robust machine learning algorithms for customer classification can improve results accuracy and consistency, allowing businesses to better target their marketing campaigns and promotions.

This, in turn, reduces the annoyance of irrelevant advertisements while increasing customer satisfaction, resulting in more fruitful long-term customer relationships. Customer classification can be especially useful for insurance companies because it allows them to identify high-value customers who are more likely to purchase insurance products. This targeted approach not only boosts revenue but also reduces marketing costs, making the company more cost-effective.

Understanding consumer behavior is essential for developing individualized marketing campaigns and successful marketing strategies in the fast-paced business environment of today. The objective of our project is to use historic transactions of customers and analyze them to create an advanced customer classification system.

We intend to categorize customers into distinct segments based on their historical behavior employing the Recency, Frequency, Monetary model or RFM Analysis. This model enables us to delve deeper into customer interactions by taking into account the most recent purchases, frequency of transactions, and monetary value spent. We hope to gain a comprehensive understanding of customer preferences and habits by using this nuanced approach.

We incorporate advanced machine learning classifiers into our system to improve the precision of customer segmentation. To predict customer labels within these identified segments, we use Keras and Support Vector Machines (SVM). The collaboration of these classifiers results in more robust and accurate customer categorization.

Both Keras and SVM classifier predictive outcomes are strategically stacked and used as input for a meta-learner, such as logistic regression. This final step of refinement aims to optimize the customer classification, ensuring a nuanced and reliable representation of various customer groups.

The ultimate goal of this project is to deliver businesses with a more refined understanding of how consumers interact with them. These insights will not only streamline marketing efforts but will also foster a more profound connection with customers, contributing to the company's overall success and growth.

II. LITERATURE SURVEY

[1]The research modes of customer purchasing characteristics are divided into three distinct groups in this paper: experience-driven mode, theory-driven mode, and data-driven mode. A customer consumption behavior analysis algorithm is proposed, as well as the concept of integrating customer intake behavior factors such as Contentment and fidelity. It is demonstrated through comparison that the data-driven model is best suited to assessing aspects of online customer purchasing patterns. To classify customers, a deep neural network structure algorithm is proposed. The study extracts various kinds of essential information, such as customer habits and consumption structure, from massive amounts of consumer behavior data in order to realize the notion of personalization.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR900

[2] The aim of this research is to promote the use of intelligence services to discover prospective consumers through offering retail corporations with timely and pertinent data. The information used is the result of a thorough investigation and serves a scientific purpose in assessing customer transaction history and purchasing habits.The current research utilizes dataset categorization concepts via the K-Means Algorithm depending on the RFM method (Recency, Frequency, and Monetary) model. The Silhouette Coefficient is calculated and used to validate a variety of information set clusters. The results of the sales transactions are contrasted to various factors such as Purchase Recency, Purchase Frequency, and Purchase Volume.

[3] The purpose of this study is to examine the predictive capability of Keras' deep learning models via a trio resilient optimization techniques (stochastic gradient descent, root-average-square propagation, and adaptive moment optimizing) and two-loss functions over local risk of landslides in geographic modeling. As an instance study, low-lying landslides in the HaLong region in Vietnam were opted for. To this, a group of ten influencing factors (the slope, perspective, curves, topographic moisture index, land use, distant to road, distant to river, kind of soil, distant to failure, and rock type) and 193 landslides the polygons have been taken to consideration for the purpose to establish a Geographic Information System (GIS) a database for this research.The evaluation wraps up that the algorithm' deep neural network model is an innovative instrument for shallow in nature susceptibility visualizing in areas vulnerable to landslides.

[4] The proposed research employs a novel approach that includes data collection, preliminary processing, feature encoded data, and categorization yet tackling sentiment assessment challenges alongside three long short-term recall modifications. When analyzing such information, it is crucial to evaluate suitable gathering of information, preliminary processing, and categorization. In the studies, various literary datasets have been employed to assess the significance of the proposed models. The suggested sentiment forecasting technique achieves stronger or at least alike, outcomes while requiring fewer computations. The findings of this research demonstrate the vital role of sentiment evaluation of feedback from customers and social networking information in obtaining significant insights. The technique of deep learning is used for assessing reviews from clients.

[5] The article delves into segmenting customers, which is regarded to be among the cornerstones of an effective marketing effort. Marketing professionals place a high value on this pivotal stage in the procedure of promoting novel goods. The framework has been altered in this paper by including variance "D" as a 4th parameter, which refers to the broadening of goods bought through an individual consumer. In a retail marketplace, categorization on the basis of RFM-D is employed to identify consumer buying trends. The model that is suggested improves the accuracy of predicting customer conduct; businesses can forecast which clients will react in a positive way which is being addressed.

[6] This study looks at e-commerce data analysis for customer segmentation using an ensemble technique with various machine learning algorithms. The combined approach, which incorporates random forest modeling, k-Nearest Neighbors, as well as ,gradient booster, accomplishes an impressive 76.83% accuracy in client categorization, demonstrating the potential for improving online shopping company tactics. For customer segmentation in e-commerce, the study's algorithm employs a combination approach including Decision Tree, Logistics Regression, Support Vector Classification (SVC),Logistics Regression, k-Nearest Neighbors, Random Forest, and Gradient Boosting Classifier.By applying machine learning algorithms, particularly ensemble models, the research aims to efficiently classify customers based on e-commerce data, providing valuable insights for stakeholders to enhance business strategies and decision-making processes.

[7]The transaction forecasting challenge within the realm of online tourism is investigated in this research, which is a relatively recent and widespread online shopping application. Although numerous investigations on buying forecasts have already been carried out, not much study was conducted regarding customer buying habits for goods related to travel. Following that, a statistical evaluation is performed to deal with several appealing features related to buying habits with the goal to verify the efficacy of factors. They present co-EM logistic regression, an innovative framework that incorporates semi-supervised training and multi-view acquiring to its strategy for predicting the likelihood that or not a purchase is going to be created throughout the present exploring event. The structure acquires and fully employs the logistical modeling benefits of simplified comprehension and data that is unlabeled.

[8] The writers of the present research use a visualization based on graphs to suggest an innovative approach for client categorization in the financial and banking industries. The expenditure habits of consumers are expressed as transmission vectors identified as purchase characteristics, collecting the way consumers assign their money across various seller classifications and methods of payment, and to assess and divide consumers according to their expenditure allocations, employing enormous scale debit and credit card purchases information sets. The findings of the analysis, graph-based methods, especially the ones that use arbitrarily walk-based techniques, provide more reliable and insightful findings for segmenting clients, with potential uses in banking risk evaluation and electronic payment supervisors.

[9] The article addresses the difficulties when employing ethical advertising techniques to attract prospective clients while retaining existing ones. Companies use tools to cater to

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR900

the two kinds of customers, resulting in a greater return investment as well as increased revenue. The contributor then clarifies the concept of "the consumer grouping," which is used by companies to classify various segments of consumers and lend those various services. The current research investigates all four categories of customers, namely engaged, cold, warm, and passive, in greater depth. It has been found that such fragments are insufficient for describing advertising approaches and require additional research. The article uses the analysis of RFM to increase the size of the piece of information and then clusters the results derived from this procedure of analysis. This examination generates the necessary rules.

[10] The article presents future years consumer habits forecasting and offers vital data for effectively distributing assets to the advertising and marketing divisions. Such data aids in arranging supplies in the storage facility and the place of sale, and also in managerial choices made during production procedures. The establishment of sophisticated analytics instruments which forecast future client behavior in a non-contractual placing and anticipate whether or not a client will be making an investment at the organization across a particular period of time in the coming years is discussed in this article. Utilizing an array of data that includes over 10,000 customers as well as a total of 200 thousand payments, the gradient-based tree booster approach proves to be the most effective technique, achieving a precision rating of 89 percent and an area under the curve (AUC) of 0.

III. METHODOLOGY

A. Data Collection:

Initially, we collect a past purchased products dataset without class labels.Then we use RFM analysis by creating 6 feature columns of recency ,frequency and monetary with scores respectively then using the feature columns we create customer_segments as the class labels for each customer ID and store it in a new csv file

B. Data Preprocessing:

The data is preprocessed before training the model in enabling the model to effectively identify the characteristics and make accurate predictions. The data type of each column was preserved and the rows containing null or duplicate values were removed .The class imbalance were down sampled to one class length and the X features were selected before train test split. Additionally, the customer dataset underwent various processing steps to enhance accuracy in classification steps.

C. Model Evaluation and Selection:

Investigate and evaluate a variety of machine learning algorithms, such as LightGBM and Decision Tree as well as XGBoost, , that can be used to successfully train and categorize clients. For model performance analysis, use cross-validation and metrics such as precision, recall, specificity, then F1 rating, accuracy, and the confusion matrices.

D. Model Stacking and Meta Learner:

This phase operates stacking on a validation set by combining predictions from two base models (a neural network model and an SVM classifier). The stacking is done horizontally to create a new feature matrix (stacked_X_val).The final stacked model is used to make predictions on the test set after a logistic regression meta-model (meta_model) is trained

E. Recommendation System:

The recommendation is done based on content and collaborative filtering which uses cosine similarity for Product associations and their relevance for recommendations. correlation in user item matrix and this Aims to provide personalized and accurate product suggestions enhancing the overall user experience

F. Streamlite and Data Analysis Report:

A Streamlit script (e.g., data_entry_app.py) with input fields for new data entry. Widgets for uploading data and displaying classification results are included.Power BI reports and dashboards offer insights into your customer data. Include graphs, charts, and tables to help you visualize key metrics and trends.

G. Validation and Testing:

During the client's categorization testing and validation phase, you have to assess the model's efficacy on unknown information and assess the model's accuracy on the set of validation data using suitable metrics such as precision, recall, precision, accuracy, F1-scoring and confusion matrices. Furthermore, by thoroughly validating and testing the client categorization algorithm's hyperparameters that guarantee its validity and appropriateness for situations in the real world, resulting in improved choices based on segmentation of customers.

IV. CONCLUSION

Ultimately, the suggested system's main objective is to create a consumer classification scheme specific for advertising purposes.The systematic integration of machine learning models, Keras and SVM to logistic regression in a stacked ensemble, signifies a thoughtful and sophisticated approach to classification. This amalgamation of techniques is designed to capitalize on the unique strengths of each model, thereby optimizing the accuracy and reliability of classification outcomes. In essence, the proposed system represents a forward-thinking and data-driven solution for businesses seeking to understand and cater to the diverse needs of their customer base. Through the synergy of RFM analysis and advanced machine learning techniques, the system has the potential to revolutionize how businesses approach customer segmentation, setting the stage for more targeted and successful marketing endeavors. This collaborative approach not only aids in understanding customer behavior but also supports the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR900

development of tailored strategies to enhance customer relationships, improve marketing efforts, and drive business growth.

REFERENCES

[1]. Shijiao Yuan.,Miao Chao Chen.,The Research Article of Analysis of Consumer Behavior Data Based on Deep Neural Network Model. In:2022 Hindawi Journal of Function Spaces Volume 2022, Article ID 4938278, https://doi.org/10.1155/2022/4938278

[2]. Anitha P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. Journal of King Saud University-Computer and Information Sciences,34(5), 1785-1792 , https://doi.org/10.1016/j.jksuci.2019.12.011.

[3]. Viet-Ha Nhu., Nhat-Duc Hoang ., Hieu Nguyen., Phuong Thao Thi Ngo., Tinh Thanh Bui., Pham Viet Hoa., Pijush Samui f., Dieu Tien Bui.Effectiveness assessment of Keras based deep learning with different robust optimization algorithms for shallow landslide susceptibility.Volume 188, May 2020, 104458. https://doi.org/10.1016/j.catena.2020.104458.

[4]. AmjadIqbal.,RashidAmin.,JavedIqbal.,RoobaeaAlroobaea.,Ahmed Binmahfoudh., Mudassar Hussain.,Sentiment Analysis of Consumer Reviews Using Deep Learning .In :2022 Issue Artificial Intelligence and Digital Transformation.https://doi.org/10.3390/su141710844.

[5]. Moulay Youssef Smaili,Hanaa Hachimi.New RFM-D classification model for improving customer analysis and response prediction.In:2023 Ain Shams Engineering Journal. https://doi.org/10.1016/j.asej.2023.102254.

[6]. Im K. and Park S., (2021) "A Study on Analyzing Characteristics of Target Customers from Refined Sales Data", APIEMS. In:2016 Conference: 18th International Conference on Business Intelligence, Analytics, and Knowledge Management.Volume 9,No 4,2016. https://www.researchgate.net/publication/287815433.

[7]. Christy, A. J., Umamakeswari, A., Priyatharsini,L., & Neyaa, A. (2021).RFM ranking–An effective approach to customer segmentation. Journal of King Saud University Computer and InformationSciences,33(10),1251-1257.https://doi.org/10.1016/j.jksuci.2018.09.004.

[8]. Roggeveen, A. L., & Sethuraman, R. (2020).Customer-interfacing retail technologies in Im K. and Park S., (2021) "A Study on Analyzing Characteristics of Target Customers from Refined Sales Data".In:2020 Journal of Retailing. https://doi.org/10.1016/j.jretai.2020.08.001.

[9]. Zhu, G., Wu, Z., Wang, Y., Cao, S., & Cao, J. (2019). Online purchase decisions for tourism e-commerce.In :2019 Electronic Commerce Research and Applications, 38,100887. https://doi.org/10.1016/j.elerap.2019.100887.

[10]. Martínez, A., Schmuck, C., Pereverzyev Jr, S., Pirker, C., & Haltmeier, M. (2020).A machine learning framework for customer purchase prediction in the non-contractual setting. European Journal of Operational Research, 281(3), 588-596. https://doi.org/10.1016/j.ejor.2018.04.034.

[11]. Jair Cervantes. , Xiaoou Li. , Wen Yu. , Kang Li.,Support vector machine classification for large data sets via minimum enclosing ball clustering. In:2008 Neurocomputing Volume 71, Issues 4–6, January 2008, Pages 611-619. https://doi.org/10.1016/j.neucom.2007.07.028.

[12]. Moghaddam S Q, Abdolvand N and Harandi S R 2017 A RFMV Model and Customer Segmentation Based on Variety of Products J. Inf.Syst. Telecommun. 5 155–61 .In 2017 , Information Systems,Telecommunication.https://www.researchgate.net/publication/321195899.

[13]. Haghighat Nia S, Abdolvand N and Rajaee Harandi S 2017 Evaluating discounts as a dimension of customer behavior analysis J.Mark. Communication.In: 2017 Journal of Marketing Communications.https://doi.org/10.1080/13527266.2017.1410210.

[14]. Alessia Galdeman,Cheick Ba,Matteo Zignani,Sabrina Gaito.A Multilayer Network Perspective on Customer Segmentation Through Cashless Payment Data .In: 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA). https://doi.org/10.1109/DSAA53316.2021.9564187.

[15]. Tabianan, K., Velu, S., & Ravi, V. (2022). K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. Sustainability, 14(12), 7243. https://www.mdpi.com/2071-1050/14/12/72.