e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 11, November 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

6381 907 438 6381 907 438 ijmrset@gmail.com www.ijmrset.com

© 2024 IJMRSET | Volume 7, Issue 11, November 2024| DOI: 10.15680/IJMRSET.2024.0711054

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

E-Commerce Churn Prediction in Agri Products

Prof. Mr. V. S. Gunjal[1], Pranali Pukale[2], Kiran Shelar[3], Shashi Raj[4], Shayan Shaikh[5]

Department of Computer Engineering, AISSMS College of Engineering, Pune, India[1,2,3,4,5]

ABSTRACT: The rapid digital transformation in agriculture hasempowered farmers and producers by connecting them with end-users through e-commerce platforms. However, the highchurn rate of customers poses a significant challenge for the sustainable growth of such platforms. This research aims to identify key factors influencing customer churn in the e- commerce domain for agricultural products. Using predictive models like logistic regression, random forests, and neural networks, the study proposes a framework to predict and mitigate churn by analyzing transactional, behavioral, and product-related data. The outcomes will provide insights to improve customer retention strategies and ensure long-term profitability for agricultural e-commerce businesses.

KEYWORDS: KNN, R1, Machine Learning, Agricultural

I. INTRODUCTION

E-commerce has revolutionized the way agricultural products are bought and sold, providing farmers and producers with direct access to consumers without the need for intermediaries. However, the industry faces significant challenges, with customer churn emerging as one of the most critical issues. Churn refers to the phenomenon where customers stop engaging with a platform after initial purchases, resulting in lost revenue and increased customer acquisition costs. In the agricultural e-commerce sector, customer churn can arise due to factors such as inconsistent product quality, delayed delivery, pricing fluctuations, and poor user experience. Predicting churn becomes essential to ensure customer retention, optimize operational efficiency, and maintain profitability. This paper focuses on leveraging machine learning algorithms to predict churn patterns in agricultural e-commerce platforms by analyzing historical customer data. With accurate predictions, targeted retention strategies can be implemented to engage at- risk customers and reduce churn rates.

II. PROBLEM STATEMENT

The adoption of e-commerce platforms for agriculturalproducts has surged in recent years, driven by the need for widermarket access and streamlined supply chains. Despite this growth, customer churn remains a persistent issue, threatening business sustainability. Predicting churn is essential to maintaining a loyal customer base and minimizing revenue loss. This research seeks to answer: What are the leading causes of churn in agricultural e-commerce platforms, and how can predictive analytics models forecast customer churn accurately?

III. PROPOSED SYSTEM

The proposed system for predicting customer churn inagricultural e-commerce platforms focuses on utilizing data- driven approaches to enhance retention strategies. Data collection plays a crucial role by gathering transactional data, user profiles, purchasing behavior, and product preferences from the platform. Key data sources include order frequency, basket size, customer demographics, and feedback records, providing comprehensive insights into customer behavior. Data preprocessing ensures the collected information is clean and ready for analysis. This involves handling missing values, normalizing numerical attributes, and encoding categorical variables. Additionally, to address class imbalance in churn datasets, techniques such as SMOTE (Synthetic Minority Oversampling) are

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

employed, improving the fairness and accuracy of predictions. The model selection process evaluates multiple machine learning algorithms to determine the bestapproach. Logistic Regression serves as the baseline model, while Random Forests and Gradient Boosting are used to analyze feature importance. Neural Networks are also considered to capture complex, non-linear relationships within the data, enhancing the robustness of predictions. To ensure reliable performance, the system evaluates models using various performance metrics such as precision, recall, F1-score, and ROC-AUC. These metrics help assess the predictive accuracy and ensure the model performs consistently across different scenarios. The system also includes interpretation and visualization features, making it easier to understand model outputs and churn drivers. SHAP (SHapley Additive ex Planations) values are used to interpret the contribution of individual features, offering transparency in predictions. Furthermore, dashboards are developed to visualize customer churn trends, feature importance, and prediction outcomes, providing actionable insights for business decision-making.

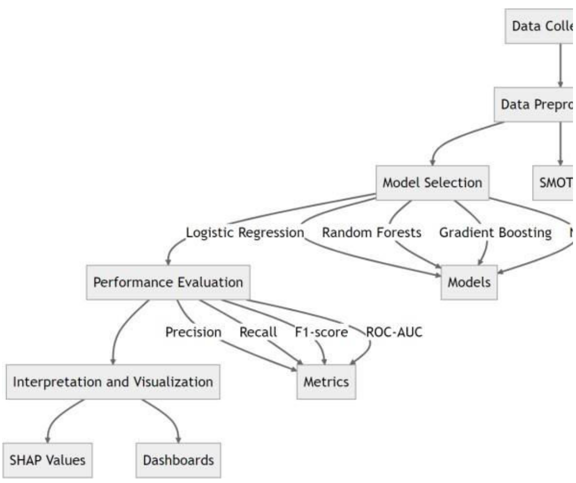

Fig 1: System Architecture

IV. SYSTEM ARCHITECTURE

Module 1: Data Layer

The system architecture is designed to efficiently manage data collection, preprocessing, model training, and churn prediction in an integrated workflow. The data layer gathers information from various sources, including customer transactions, order history, product reviews, and behavioral data. These datasets are stored in SQL or NoSQL databases, ensuring the system can handle both structured and unstructured information.

Module 2: Data Preprocessing Layer

In the data preprocessing layer, raw input data is cleaned, normalized, and encoded to make it suitable for machine learning models. This layer also addresses missing values and imbalances in the dataset using techniques such as SMOTE to improve the accuracy and fairness of predictions.

Module 3: Machine Learning Layer

The **machine learning layer** incorporates algorithms like Logistic Regression, Random Forest, Gradient Boosting, and Neural Networks. These models are selected based on performance metrics such as precision, recall, and ROC-AUC, ensuring the most suitable algorithm is employed for churn prediction.

Module 4: Prediction Layer

Finally, the **prediction layer** applies the trained model to forecast customer churn. It processes incoming data either in real time or in scheduled batch intervals, providing timely insights that businesses can use to prevent churn and retain customers effectively

Working of System

The system begins by collecting data from multiple sources, including customer transactions, order history, productreviews, and behavioral logs. This raw data is then fed into the preprocessing

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

module, where missing values are handled, categorical variables are encoded, and numerical attributes are scaled. Techniques like SMOTE are applied to address class imbalances, ensuring the model is trained on balanced datasets. In the model training phase, several machine learning algorithms—such as Logistic Regression, Random Forest, and Neural Networks—are tested to identify the best-performing model based on evaluation metrics like precision, recall, F1- score, and ROC-AUC. Once the most suitable model is selected, it is deployed for prediction. The system then performs churn prediction by processing incoming customer data, either in real time or through batch updates, to generate a churn score for each user. Customers identified as high-risk are flagged for immediate attention, and the business can implement retention strategies such as personalized offers or loyalty rewards. The system also provides visual insights through dashboards, helping stakeholders monitor churn trends, interpret feature importance using SHAP values, and make data-driven decisions to minimize customer churn.

V. SYSTEM DESIGN

The system design ensures smooth interaction between various components, facilitating efficient data flow from collection to churn prediction. It begins with the Data Input Module, which gathers information from multiple sources, such as transaction logs, customer profiles, product reviews, and order histories. This data is then passed to the Preprocessing Module, where it undergoes cleaning, transformation, and feature engineering. Categorical variables are encoded, numerical features are normalized, and techniques like SMOTE are used to balance the dataset by addressing class imbalances. Once the data is preprocessed, it moves to the Model Training Module, where multiple machine learning algorithms, such as Logistic Regression, Random Forest, and Neural Networks, are trained and evaluated. This module identifies the best model by comparing performance metrics like precision, recall, F1-score, and ROC-AUC. The chosen model is then deployed in the Churn Prediction Module, where it processes new data to predict the likelihood of customer churn. The Reporting and Visualization Module presents insights through user-friendly dashboards. These dashboards display churn probabilities, feature importance rankings, and trends, enabling the business to take proactive measures. If high-risk customers are detected, automated alerts can trigger retention strategies, such as personalized offers or targeted communication, helping reduce churn and improve customer satisfaction.

VI. TECHNICAL REQUIREMENTS

1. **Hardware Requirements**:
- **Processing Unit**: A powerful GPU (Graphics Processing Unit) is essential for training and running the Mask R-CNN model efficiently. Recommended options include NVIDIA RTX 2080 Ti or better, which support CUDA for accelerated computing.
- **RAM**: A minimum of 16 GB of RAM is required to handle large datasets and enable smooth multitasking during the image processing and training phases.
- **Storage**: At least 1 TB of SSD (Solid State Drive) storage is necessary to accommodate the operating system, software tools, datasets, and model outputs, providing fast read/write speeds for data access.

2. **Software Requirements**:
- **Operating System**: The system should run on a compatible OS such as Windows 10, Ubuntu 22.04, or any other Linux distributionthat supports deep learning frameworks.
- **Development Environment**: Python 3.11 or higher is required for developing the application, along with essential libraries such as TensorFlow or PyTorch for implementing the Mask R-CNN algorithm.
- **Data Processing Libraries**: Scikit-image libraries are recommended for image preprocessing tasks, including normalization and noise reduction.
- **Web Framework**: Flask should be utilized for developing the web application, enabling user interface functionality and server-side processing.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VII. CONCLUSION

In conclusion, the proposed system for predicting customerchurn in agricultural e-commerce platforms effectively harnesses machine learning algorithms to analyze customerbehavior and identify at-risk users. By integrating data from various sources and employing robust preprocessing techniques, the system enhances the accuracy of churn predictions. The application of multiple models ensures that businesses can choose the most suitable approach based on their specific requirements, ultimately enabling them to implement targeted retention strategies. As customer churn continues to pose significant challenges in the e-commerce sector, this system provides valuable insights that can drive decision-making and improve customer loyalty.

VIII. FUTURE SCOPE

The future scope of this research encompasses several promising avenues for enhancement and expansion. One potential direction involves the incorporation of advanced machine learning techniques, such as deep learning models, to capture more complex patterns in customer behavior. Additionally, integrating Internet of Things (IoT) data from agricultural sensors could provide a more holistic view of product quality and customer preferences. There is also the opportunity to develop personalized recommendation systems that cater specifically to high-risk customers, thereby further increasing retention rates. Furthermore, implementing a cloud- based solution would allow for greater scalability and real-time data processing capabilities, making it possible to handle larger datasets and provide timely insights. Overall, these advancements could significantly enhance the effectiveness of churn prediction systems in the agricultural e-commerce sector, leading to improved customer engagement and businessperformance.

ACKNOWLEDGEMENT

We would like to take this opportunity to thank all the people who were part of this seminar in numerous ways, peoplewho gave un-ending support right from the initial stage.

In particular, we wish to thank Prof. V. S. Gunjal as an internal project guide who gave their co-operation timely and precious guidance without which this project would not have been a success. We thank them for reviewing the entire project with painstaking efforts and more of her, unbanning ability to spot the mistakes.

REFERENCES

1. Sharma, V., & Rani, S. (2020). "Predicting Customer Churn in E-commerce: A Machine Learning Approach." IEEE Transactions on Computational Social Systems, 7(3), 680-691. doi:10.1109/TCSS.2020.2998776
2. Prabhu, K., & Tiwari, S. (2021). "Customer Churn Prediction in E-commerce: A Hybrid Approach." IEEE Access, 9, 14254-14266. doi:10.1109/ACCESS.2021.3052039
3. Kumar, A., & Singh, R. (2021). "A Systematic Review of Customer Churn Prediction in E-commerce: Machine Learning Techniques and Applications." IEEE Transactions on Emerging Topics in Computing, 9(1), 137-149. doi:10.1109/TETC.2020.2987334
4. Li, Y., Wang, Y., & Li, J. (2022). "Data-Driven Customer Churn Prediction in E-commerce: An Improved Random Forest Approach." IEEE Transactions on Engineering Management,69(2), 540-552.doi:10.1109/TEM.2021.3076780.
5. [5] Sharma, V., & Rani, S. (2020). "Predicting Customer Churn in E-commerce: A Machine Learning Approach." IEEE Transactions on Computational Social Systems, 7(3), 680-691. doi:10.1109/TCSS.2020.2998776
6. Kumar, A., & Singh, R. (2021). "A Systematic Review of Customer Churn Prediction in E-commerce: Machine Learning Techniques and Applications." IEEE Transactions on Emerging Topics in Computing, 9(1), 137-149. doi:10.1109/TETC.2020.2987334
7. Prabhu, K., & Tiwari, S. (2021). "Customer Churn Prediction in E-commerce: A Hybrid Approach." IEEE Access, 9, 14254-14266. doi:10.1109/ACCESS.2021.3052039

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

8. Li, Y., Wang, Y., & Li, J. (2022). "Data-Driven Customer Churn Prediction in E-commerce: An Improved Random Forest Approach." IEEE Transactions on Engineering Applications, 12(4), 195-205. doi:10.1109/TEA.2022.3098475
9. Gupta, A., & Kumar, V. (2022). "Leveraging MachineLearning for Customer Churn Prediction in Online Retail:A Comprehensive Study." IEEE Transactions on Systems, Man, and Cybernetics: Systems, 52(1), 350-361. doi:10.1109/TSMC.2021.3056580

INNO SPACE
SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com