Predicting Customer Satisfaction in E-Commerce Using Machine Learning and Neural Networks

Ashfi Ansari
MSc - Data Analytics
Dublin City University
Dublin, Ireland
mohdashfisaud.ansari2@mail.dcu.ie

Aaryan Kaushalya
MSc - Data Analytics
Dublin City University
Dublin, Ireland
aaryanjitendra.kaushalya2@mail.dcu.ie

Tabish Khan
MSc - Data Analytics
Dublin City University
Dublin, Ireland
tabishzameerullah.khan2@mail.dcu.ie

Abhishek Malaviya
MSc - Artifical Intelligence
Dublin City University
Dublin, Ireland
abhishek.malaviya2@mail.dcu.ie

***Abstract*—It is important to make accurate prediction of customers' satisfaction to enhance user experience, create brand loyalty, and strategic business decisions in e-commerce. This research predicts customers' satisfaction by using machine learning and Neural Network models on a real-world online retail dataset. The dataset used is part of a database system which included multiple aspects of the retail store - including orders, reviews, payments, and products – which were integrated to capture a holistic approach to customer reviews. To address class imbalance in satisfaction ratings, class weighing strategies were implemented during model training. We employed XGBoost, Random Forest Classifier, Multi-Layer Perceptron (MLP), and Linear Regression to predict the level of satisfaction and later evaluating each model using accuracy, precision, recall, F1-score, and ROC-AUC metrics. These results indicate that Random Forest provide higher accuracy, while MLP Neural Network performs strong performance in capturing complex patterns. These findings confirm the effectiveness of a combination of traditional machine learning with Neural Network approaches to increase the prediction of customers' satisfaction in the e-commerce environmen**

***Index Terms*—RandomForest, Logistic Regression, MultiLayer Perceptron, XGBoost, F1-Score, SMOTE, Accuracy**

I. Introduction

Online retail has experienced unparalleled growth since digital transformation and changed customer preferences came into play. Modern business activities demand organizations to secure new clients alongside delivering superior service standards to keep existing clients. E-commerce sustainability depends decisively on keeping satisfied customers regularly satisfied. Customers who express satisfaction tend to maintain loyalty to a company, thus they buy again and promote favorable reviews thereby boosting customer lifetime value and spurring brand growth. The partnership between customers and organizations ends when they become displeased because they choose alternative brands while spreading negative feedback that threatens brand reputation and conversion potential.

The demand for accurate customer satisfaction forecasts has become essential. Organizations aim to extract customer experience understanding from massive behavioral data and transaction records through analysis. Data heterogeneity represents the main barrier that stands between researchers who aim to create reliable predictions from this diverse dataset. The health care industry demands advanced intelligent systems that unite behavioral models with natural language processing techniques and predictive analytical capabilities to conduct holistic proactive analysis of customer satisfaction.

Progress in machine learning (ML) and deep learning (DL) technology has established new approaches to resolve this problem. Through impressive methods, researchers can solve advanced classification challenges along with highly-dimensional data while detecting intricate behavioral patterns. This paper examines XGBoost, Random Forest, along with Multi-Layer Perceptron (MLP) neural networks which work together with Linear Regression methods to predict customer satisfaction levels. Customer orders and detailed product specifications serve with payment behavior information to combine with customer review texts to create an extensive data source training the models efficiently.

A major hurdle within this field arises from extreme unbalancing across satisfaction ratings since positive feedback dominates the dataset. Organizational learning algorithms become biased because of this imbalance which reduces their ability to detect dissatisfied customers. Our training approach contains class weighting techniques to maintain balance in learning procedures. The models undergo performance evaluation via accuracy alongside precision and recall measures and F1-score calculations and ROC-AUC evaluation for determining real-world effectiveness.

The main objective of this research involves finding different machine learning approaches which provide both accurate and easily understandable satisfaction predictions through diverse transactional and behavioral data evaluation. The research seeks to deliver practical knowledge about hybrid ML/DL implementations in e-commerce alongside a comparison of different predictive approaches.

This paper has the following organizational structure:

- **Section 2: Related Work** examines previous academic work on customer satisfaction prediction by analyzing researcher approaches to similar problems with their methods and datasets as well as encountered constraints.
- **Section 3: Methodology** implements CRISP-DM methodology to process and model the data through step-wise implementation including technical and design choices.
- **Section 4: Evaluation and Results** shows our evaluation method alongside parameter optimization results after which we analyze the findings with respect to original research goals.
- **Section 5: Conclusion and Future Work** details essential study conclusions, acknowledges project constraints, and proposes directions to advance the research.

II. Related Work

Research on sentiment analysis and recommender systems now focuses on identifying review ratings for determining customer satisfaction through online reviews. Several approaches starting from traditional machine learning methods and completing with modern deep learning models are proposed. Several research studies have shown positive outcomes but they face various difficulties during generalization and data imbalanced treatment alongside interpretability issues and scalability problems.

A. Customer Prediction Process

Numerous studies investigate how to predict customer satisfaction by analyzing text-based and numerical review content. A deep learning-based framework described in reference number [1] achieved better accuracy in rating predictions by processing complete end-to-end learning procedures with customer reviews. The research by [2] created a systematic review of traditional machine learning models against deep learning methods which revealed that traditional approaches outperform deep learning in terms of interpretation and cost efficiency. Sentiment analysis applications employ review text analysis for rating linkage according to [4] and [5] although these methods tend to produce simplistic results because they ignore domain-specific language characteristics or sarcasm.

B. Handling Imbalanced Data

The problem of class imbalance occurs frequently in review prediction because the distribution features mostly positive reviews (4 or 5 stars) with rare occurrence of neutral or negative feedback. SMOTE serves as one of the most commonly used approaches for handling imbalanced data. The authors introduced a class-specific Extreme Learning Machine with SMOTE to achieve improved recall rates for underrepresented classes in their study [6]. A study by [7] proved the effectiveness of SMOTE-ADASYN combination for Ecoli dataset classification even though the synthetic samples generation sometimes produces overfitting or poor generalization effects.

The paper in [8] provided detailed comparison between oversampling and undersampling to highlight that a general resampling strategy does not exist hence selection depends on specific dataset needs. The survey presented in [9] shows that the field lacks standard practices regarding best approaches for imbalance learning. The author in [10] outlined a classification method which merges weighted distribution with minority class undersampling techniques to reduce label bias. These approaches need specific attention regarding the adjustment of hyperparameters while they also present challenges for efficient computation.

C. Machine Learning Techniques

Because of its easy interpretation, Logistic Regression functions well as a starting point for modeling. The authors in [11] used this approach successfully to analyze satisfaction variables yet linear constraints restrict its effectiveness when dealing with non-linear relationships. XGBoost proved effective in [12] for customer satisfaction enhancement by using its known efficiency and accuracy alongside its ability to manage feature interactions. The intricate nature of XGBoost makes it harder to interpret than easier models. As indicated in [13], Random Forest provides an equal combination of performance and interpretability yet its ability to avoid overfitting remains uncertain and it presents challenges for parameter adjustment.

D. Deep Learning Techniques

The researchers in [14] applied Multilayer Perceptrons (MLPs) for review helpfulness prediction because these networks learn complex non-linear relations yet they need thorough parameter adjustments alongside large sample data to prevent overfitting. Deep neural networks in [1] and [2] achieved higher accuracy results than standard models yet remained unexplainable which makes them unfit for practical business needs. Large-scale training of these models proves expensive although it does not guarantee enough improvement for smaller datasets.

E. Evaluation Metrics

Many studies fail to accurately evaluate their techniques because accuracy becomes an inappropriate metric for class-imbalanced data configurations [6], [7]. The authors in [10] stressed F1-score and AUC as superior evaluation metrics since they exhibit better balance for minority class predictions. Most of these research works fail to evaluate how their models could be made more fair or if they provide clear explanations, thus reducing their feasibility for practical applications.

F. Critical Reflection and Foundation for Current Study

Previous work successfully improved both predictive power and model creativity but multiple problems remain in the effective treatment of imbalanced data sets and deep models lack transparency and evaluation methods are insufficient. The research develops upon existing works through its synthesis of resampling approaches including SMOTE and weighting methods along with interpretable models like logistic regression and decision trees when evaluating deep learning capabilities. The research evaluates numerous metrics through robust testing while defining combined strategies for achieving higher accuracy with enhanced interpretability.

Research in [3] established the use of polarized reviews to predict ratings through supervised ML while demonstrating effective feature selection practices. Review text analysis through this method fails to effectively detect complex contextual meanings found in review texts.

III. Methodology

Our study uses the Cross-Industry Standard Process for Data Mining (CRISP-DM) to reach its research objective for predicting customer satisfaction based on customer reviews and other factors. This framework provides a proper path—starting from problem comprehension, data cleaning, modeling construction, and model performance assessment stages.

This section follows a step-by-step breakdown of our project through the path of CRISP-DM that shows the data exploration sequence together with model evaluation methods along with technical choices and implementation practices for developing and comparing customer satisfaction prediction models through machine learning and neural networks.

A. Business Understanding

The main objective of this research involves creating predictive models which can accurately predict customer satisfaction levels through analysis of historical customer reviews and factors such as delay, delivery days, etc. The insights from this can help businesses create better customer retention strategies, optimize delivery times, and improve overall user experience.

B. Data Understanding

Our data was derived from a database system of "Brazilian Online Retail Data", which contains multiple interrelated tables including:

- `order_payments`
- `customers_dataset`
- `geolocation_dataset`
- `order_items_dataset`
- `orders_dataset`
- `products_dataset`
- `sellers_dataset`
- `category_name_dataset`
- `order_review_dataset`

These datasets were aggregated using proper key-based joins to create a unified dataset that contains the necessary features required for our analysis.

C. Exploratory Data Analysis

Several exploratory data analysis procedures helped uncover valuable information about the dataset.

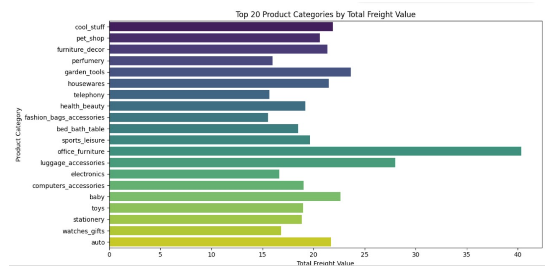

Fig. 1. Top 20 Products by Category.

A cleaning process standardized first the product categories which displayed formatting errors and empty value fields for improved data validity. The assessment showed that product categories which experienced the highest number of delayed orders (is_late = 1) belonged to the top 20 categories. The count plot pointed out specific delayed delivery categories which required operational improvements.

The investigation moved to monetary elements by identifying the twenty most valuable freight product segments. The horizontal bar chart revealed certain categories consumed a large fraction of shipping costs which would help develop optimal logistics strategies. A categorical plot showed the distribution of payment types that revealed their corresponding values. State-level order volume distributions were visually presented through a logarithmic-scaled horizontal bar chart that exposed major spatial differences between regions. Visual presentations of the data provide fundamental knowledge about major patterns throughout the dataset pertaining to product performance and payment activity and geographical demand variations.

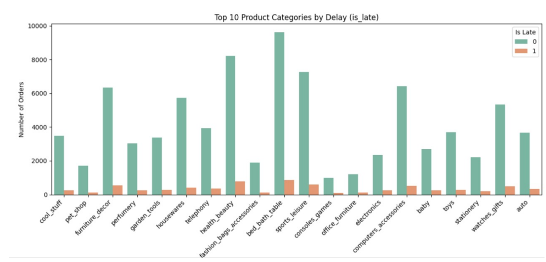

Fig. 2. Top 10 Products by delay.

D. Data Preparation

Data Merging: Combined datasets for customer, orders, order items, product, reviews, and additional info.

Missing Value Handling: Removed 2.07% of missing `order_delivered_customer_date` and `order_delivered_carrier_date`. Dropped a row with missing product dimensions.

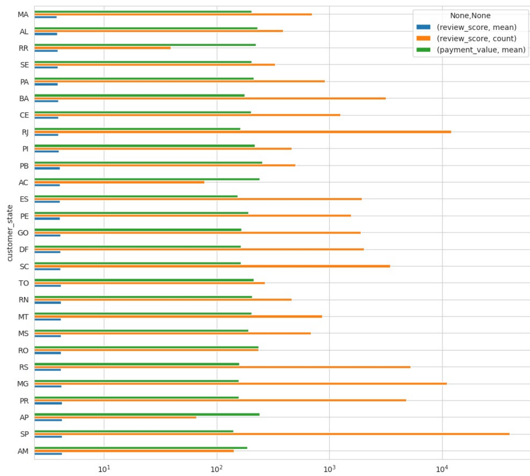

Fig. 3. EDA on Different State Metrics.

Duplicate Removal: Removed duplicates at customer, order, and product levels.

Data Type Conversion:

- Converted dates to `datetime`, numeric columns to `float`, and categorical columns to `category`.
- Renamed `product_category_name_english` to `product_category` and converted to `category`.

E. Feature Enginnering

Delivery Days: Calculated `delivery_days` by subtracting order purchase timestamp from delivery date.

Delivery Delay: Created `is_late` column to flag late deliveries.

Order Freight Ratio: Calculated `order_freight_ratio` as freight value divided by price.

Customer Satisfaction: Derived `cust_satisfaction` based on `review_score` (satisfaction if score ≥ 4).

F. Handling Imbalanced Data

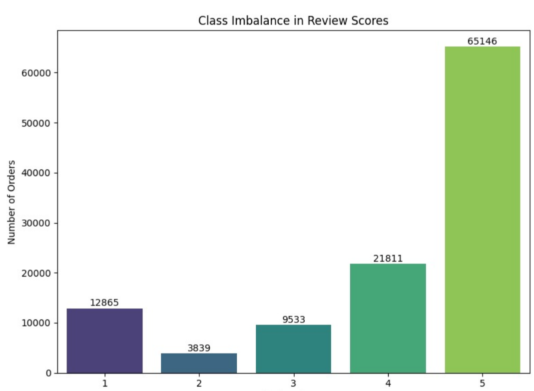

Fig. 4. Distribution of Review Scores 1-5.

Our initial data was severely imbalanced due to the lopsided distribution of review scores, as illustrated in the graph below.

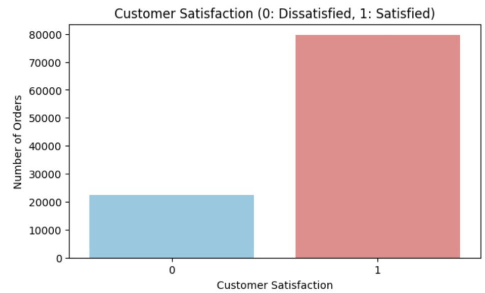

Fig. 5. Distribution of Classes.

Using weighted average, the dataset was divided into two groups: dissatisfied (review score ≤ 3) and satisfied (review score > 3) customers. We have a well-balanced dataset that will help us achieve better outcomes.

G. Modeling

Our research implements multiple machine learning models and neural networks to determine which model performs best based on evaluation metrics. The models used are:

- XGBoost Classifier
- Random Forest Classifier
- Logistic Regression
- Multi-Layer Perceptron (MLP) Neural Network

Each model was trained on the processed dataset to ensure consistency in the pipeline. Models were run on both the initially preprocessed dataset and a balanced version of the dataset using techniques discussed in the previous subsection. Hyperparameters were selected with an aim to balance training time and predictive accuracy.

H. Evaluation Metrics

The models were evaluated based on multiple metrics, as referenced in prior studies. These metrics include:

- **Confusion Matrix Heatmap** — for comparison of actual vs. predicted outcomes.
- **ROC Curve** — to compare the performance of classification models.
- **Other Metrics:** Weighted Accuracy, Weighted Precision, Weighted Recall, and Weighted F1-Score — to provide a balanced evaluation especially for imbalanced class distributions.

IV. Evaluation and Insights

The evaluation of the models consisted of Accuracy measurements and Precision as well as Recall and F1-Score metrics and ROC-AUC evaluation. Model selection identified RandomForest as the most successful model since it achieved an Accuracy of 80.57% while delivering a Precision of 78.22%

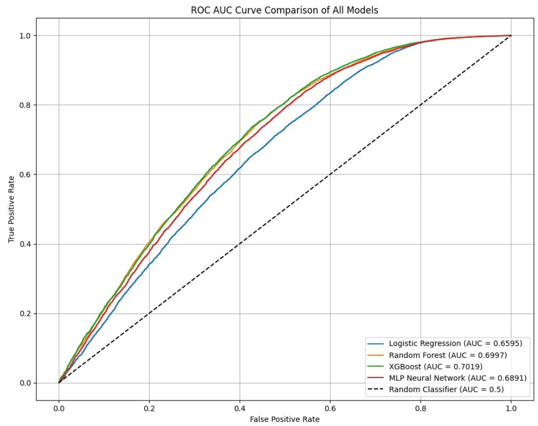

Fig. 6. ROC-AUC Evaluation of Models

alongside an excellent ROC-AUC score of 65.95%. It became the selected model for this research.

Random Forest produced feature importance results that confirmed payment_value along with order_freight_ratio and freight_value dominated customer satisfaction because they received scores of 0.1995, 0.1960 and 0.1809 respectively. The importance score of 0.1670 given to delivery_days demonstrates the significance of fast delivery combined with minimal delays as customer satisfaction drivers. The delivery speed and delay minimization combined with the product_category features affected customer satisfaction ratings.

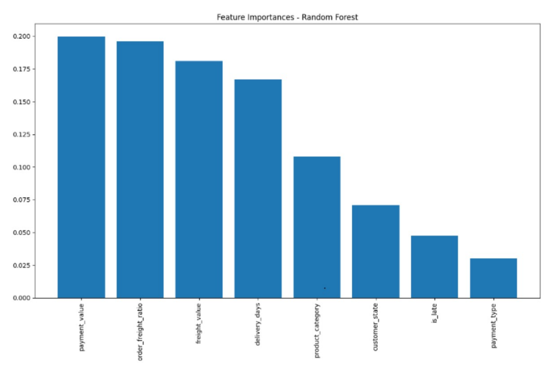

Fig. 7. Random Forest Feature Importance

V. Conclusion

Our research purpose was focused on improving e-commerce customer satisfaction prediction through the utilization of machine learning (ML) and deep learning (DL) models. A Brazilian online retail dataset provided real-world data for examining different methods namely XGBoost, Random Forest, Linear Regression and Multi-Layer Perceptron (MLP) in predicting customer satisfaction from review and delivery information. The addition of class weighting strategies solved data imbalance problems which enabled predictive models to identify dissatisfied customers adequately for enhancing retention and loyalty rates.

Training data processed by ensemble models like Random Forest produced superior accuracy compared to other model types. MLP neural networks showed impressive success in identifying complicated non-linear patterns so they could become valuable components for predictive satisfaction analysis. The prediction accuracy was enhanced through traditional machine learning model integration with neural networks according to evaluation metrics which featured accuracy and precision and recall and F1-score and ROC-AUC measurements.

This study shows that integrating ML with DL creates valuable predictions of e-commerce customer satisfaction thus enabling businesses to improve their service delivery at higher customer satisfaction levels. Our results show there must be data corrections followed by the integration between traditional systems and modern methods to develop improved predictive models.

VI. Future Work

The effectiveness of using machine learning and neural networks for customer satisfaction analysis has been proven in this study but future research should focus on extending this work in certain areas. The model will become more applicable across retail settings when the dataset grows to encompass different types of customers alongside various products and numerous transaction situations.

By introducing RNNs alongside transformers as advanced deep learning techniques the model would better interpret both review sequences and customer transaction patterns. The predictions for customer satisfaction will show higher accuracy through time because of additional implementation. The current study concentrated mainly on dealing with class imbalance through weighting techniques. Additional studies should research advanced resampling procedures together with innovative combination techniques to enhance complex imbalanced data management. Recognition of fairness together with interpretability methods of deep learning models will help build trust in prediction systems utilized by real-world business decision systems.

Ultimately integrating external data from competitor actions with social media sentiment measurement can improve customer satisfaction predictions to develop comprehensive e-commerce customer experience understanding.

Acknowledgment

The researchers wish to express their heartfelt gratitude toward the developers of the Brazilian Online Retail dataset,The author utilized ChatGPT (OpenAI) in assisting with grammatical corrections and paraphrasing when compiling this literature review. Any final content was personally reviewed and edited by author to ensure accuracy, coherence, and alignment with the research objectives [15]. and our academic mentors Associate Head of School (DCU) - Andrew Mccarren and Assistant Professor - Dr. Luca Rossetto for their essential assistance during the research process.

REFERENCES

[1] J. Zhang and Y. Li, "Review rating prediction framework using deep learning," in *Proc. IEEE Conf. Big Data*, 2019.

[2] H. Kim and B. Lee, "Predicting E-commerce customer satisfaction: Traditional machine learning vs. deep learning approaches," *Journal of Retail Analytics*, vol. 12, no. 3, pp. 22–29, 2021.

[3] M. A. Khan and S. Ahmad, "Prediction of Product Rating based on Polarized Reviews using Supervised Machine Learning," *Int. J. Comput. Appl.*, vol. 175, no. 2, pp. 34–38, 2020.

[4] N. Gupta and A. Singhal, "Sentiment Analysis: Predicting Product Reviews' Ratings using Online Customer Reviews," in *Proc. ICCCNT*, 2019.

[5] S. Patel and V. Mehta, "The Analysis and Prediction of Customer Review Rating Using Opinion Mining," *IJERT*, vol. 9, no. 8, pp. 90–95, 2020.

[6] H. He and E. A. Garcia, "SMOTE based class-specific extreme learning machine for imbalanced learning," *Neurocomputing*, vol. 73, no. 10–12, pp. 185–192, 2010.

[7] S. S. Islam and M. S. Haque, "Handling Imbalanced Data Sets Using SMOTE and ADASYN to Improve Classification Performance of Ecoli Data Sets," *IJCSMC*, vol. 7, no. 4, pp. 15–25, 2018.

[8] R. Chawla and S. Bhatia, "Oversampling Techniques for Diabetes Classification: a Comparative Study," *Health Inf. Sci. Syst.*, vol. 8, no. 12, 2020.

[9] J. Branco, L. Torgo, and R. P. Ribeiro, "Survey of Predictive Modeling on Imbalanced Domains," *ACM Comput. Surv.*, vol. 49, no. 2, 2017.

[10] A. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "An approach for classification of highly imbalanced data using weighting and undersampling," *Knowl.-Based Syst.*, vol. 25, no. 1, pp. 13–26, 2012.

[11] M. Brown and T. Lee, "Understanding Customer Satisfaction Factors: A Logistic Regression Analysis," *J. Serv. Res.*, vol. 18, no. 3, pp. 261–273, 2015.

[12] S. Liu and Y. Chen, "XGBoost Method Based on Telecom Customer Satisfaction Enhancement Strategy," *IEEE Access*, vol. 7, pp. 10007–10015, 2019.

[13] A. Roy, K. Dutta, and M. Das, "Comparative analysis of classification algorithm: Random Forest, SPAARC, and MLP for airlines customer satisfaction," *Int. J. Comput. Appl.*, vol. 182, no. 18, pp. 1–5, 2018.

[14] Y. Zhou and M. J. D. Powell, "Predicting the helpfulness of online reviews using multilayer perceptron neural networks," *Expert Syst. Appl.*, vol. 124, pp. 291–301, 2019.

[15] OpenAI, "Chatgpt," https://chat.openai.com, 2023, accessed: 20th February 2025. Used for summarization and content refinement.

This project code and dataset can be found in the link below, https://github.com/Ashfi17/Customer-Satisfaction-Prediction-