Int. j. adv. multidisc. res. stud. 2023; 3(4):1332-1337

Received: 10-07-2023
Accepted: 20-08-2023

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

Predicting Customer Purchase Using Machine Learning: A Comparative Study of Traditional and Deep Learning Models

[1] Isioma Rhoda Chijioke, [2] Chiika Adaora Jasmin, [3] Chukwuma James
[1] Department of Marketing, Faculty of Management Sciences, Imo State University, PMB 2000, Nigeria
[2] Department of Accounting, College of Management and Social Sciences, Osun State University, PMB 4494, Osogbo, Osun State, Nigeria
[3] Department of Management and Accounting, Faculty of Management Sciences, Ladoke Akintola University of Technology, P.M.B. 4000, Ogbomoso, Oyo State, Nigeria

DOI: https://doi.org/10.62225/2583049X.2023.3.4.5867

Corresponding Author: **Isioma Rhoda Chijioke**

Abstract
This study provides a comprehensive evaluation of the effectiveness of both traditional machine learning techniques and deep learning models in predicting customer purchase behavior within a data-driven business environment. Using a simulated dataset comprising 5,000 customers, the study incorporates a range of demographic variables (such as age, gender, and income level) alongside behavioral features, including browsing frequency, purchase history, and engagement patterns. Six predictive models are systematically examined: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), a feed-forward Neural Network, and a Long Short-Term Memory (LSTM) network. Model performance is assessed using multiple evaluation metrics accuracy, F1-score, receiver operating characteristic area under the curve (ROC-AUC), and training time to ensure a balanced comparison of predictive power, robustness, and computational efficiency. The results indicate that deep learning models consistently outperform traditional machine learning approaches in terms of predictive accuracy and overall classification performance, with the LSTM model achieving the highest scores due to its ability to capture temporal and sequential patterns in customer behavior. Nevertheless, traditional models demonstrate notable advantages, particularly in faster training times, lower computational cost, and greater interpretability, making them suitable for organizations with limited resources or strong explainability requirements. Overall, the findings offer practical insights for businesses and data practitioners by highlighting the trade-offs between accuracy, interpretability, and computational efficiency, thereby supporting informed decision-making when selecting predictive analytics models for customer behavior forecasting.

Keywords: Machine Learning, Support Vector Machine (SVM), LSTM model

Introduction

In the rapidly evolving digital economy, understanding and predicting customer behavior has become a cornerstone of competitive advantage. As e-commerce continues to expand globally, businesses are increasingly reliant on data-driven strategies to optimize customer engagement, personalize experiences, and drive conversions. One of the most critical aspects of this transformation is the ability to accurately predict customer purchase intent—the likelihood that a user will complete a transaction based on their online behavior. This capability enables organizations to tailor marketing efforts, allocate resources efficiently, and enhance customer satisfaction (Chen, Li, & Wang, 2021) [2].

The proliferation of digital platforms has led to an explosion of customer interaction data, including browsing history, clickstream patterns, product views, and transaction records. While this data holds immense potential, extracting actionable insights from it remains a complex challenge. Traditional machine learning (ML) models such as logistic regression, decision trees, and support vector machines (SVMs) have been widely used for classification tasks in this domain due to their interpretability and computational efficiency (Hosmer, Lemeshow, & Sturdivant, 2013; Cortes & Vapnik, 1995) [6, 3]. However, these models often struggle to capture the intricate, non-linear relationships inherent in high-dimensional and sequential data.

In contrast, deep learning (DL) models, including feedforward neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs) such as long short-term memory (LSTM) networks, have demonstrated superior performance in various domains, particularly those involving unstructured or sequential data (LeCun, Bengio, & Hinton, 2015; Hochreiter & Schmidhuber, 1997) [7, 5]. These models are capable of automatically learning complex feature representations, making them well-suited for modeling customer behavior patterns in e-commerce environments. Despite the growing interest in DL techniques, there remains a need for comprehensive comparative studies that evaluate both traditional and deep learning models within the context of purchase intent prediction. Many existing studies focus on specific algorithms or datasets, limiting the generalizability of their findings. Moreover, while DL models often achieve higher predictive accuracy, they also introduce challenges related to interpretability, computational cost, and ethical considerations (Doshi-Velez & Kim, 2017) [4].

This study aims to bridge this gap by conducting a systematic comparison of traditional ML and DL models for predicting customer purchase intent. Using publicly available datasets, we evaluate the performance of seven models, logistic regression, decision tree, random forest, SVM, feedforward neural network, CNN, and RNN/LSTM, across key metrics including accuracy, precision, recall, and F1-score. The study also explores the business implications of predictive modeling, addresses data and ethical challenges, and outlines future research directions. The primary objectives of this research are as follows: (1) to assess the predictive performance of traditional and deep learning models on customer purchase intent data; (2) to analyze the trade-offs between model accuracy, interpretability, and computational efficiency; (3) to discuss the practical implications of deploying these models in real-world business settings; and (4) to identify key challenges and opportunities for future research in this domain.

The contributions of this paper are threefold. First, we provide a detailed empirical comparison of traditional and deep learning models using consistent datasets and evaluation protocols. Second, we contextualize the technical findings within broader business and ethical frameworks, offering insights for practitioners and policymakers. Third, we propose a roadmap for future research that emphasizes hybrid modeling approaches, explainable AI, and the integration of multimodal data sources.

Literature Review

Traditional machine learning (ML) models have long been the cornerstone of predictive analytics in e-commerce. These models are particularly valued for their interpretability, computational efficiency, and ease of implementation. Among the most commonly used traditional ML models are logistic regression, decision trees, random forests, and support vector machines (SVMs).

Logistic regression is a statistical model that estimates the probability of a binary outcome based on one or more predictor variables. It has been widely used in marketing and customer analytics due to its simplicity and interpretability (Hosmer, Lemeshow, & Sturdivant, 2013) [6]. Decision trees, on the other hand, offer a hierarchical structure that mimics human decision-making, making them intuitive and easy to visualize. However, they are prone to overfitting, especially when the tree becomes too deep. To address the limitations of single decision trees, ensemble methods such as random forests were introduced. Random forests aggregate the predictions of multiple decision trees to improve generalization and reduce variance (Breiman, 2001) [1]. Support vector machines (SVMs) are another powerful traditional ML technique, particularly effective in high-dimensional spaces. SVMs aim to find the optimal hyperplane that separates classes with the maximum margin (Cortes & Vapnik, 1995) [3].

Deep learning (DL) models have revolutionized predictive analytics by enabling the automatic extraction of complex features from raw data. These models are particularly effective in handling unstructured and high-dimensional data, such as text, images, and sequential clickstream data. Feedforward neural networks (FNNs) are the simplest form of DL models, consisting of layers of interconnected neurons where information flows in one direction. They are capable of modeling non-linear relationships and have been applied to various classification tasks in e-commerce (LeCun, Bengio, & Hinton, 2015) [7]. Convolutional neural networks (CNNs), originally developed for image recognition, have found applications in e-commerce for analyzing visual product data and spatial patterns in user behavior. CNNs use convolutional layers to detect local patterns and pooling layers to reduce dimensionality, making them efficient and scalable. Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) networks, are designed to handle sequential data by maintaining memory of previous inputs (Hochreiter & Schmidhuber, 1997) [5]. This makes them ideal for modeling user sessions and clickstream data, where the order of interactions is crucial for predicting intent.

Several studies have compared the performance of traditional ML and DL models in the context of purchase intent prediction. Zhang, Zhao, and Xu (2019) [8] conducted a comprehensive evaluation of ML and DL models using e-commerce datasets and found that DL models consistently outperformed traditional models in terms of accuracy and recall. Similarly, Chen, Li, and Wang (2021) [2] demonstrated that CNNs and LSTMs achieved higher predictive performance than logistic regression and decision trees, particularly when dealing with large-scale, high-dimensional data.

However, these studies also highlight the trade-offs involved. While DL models offer superior accuracy, they require more computational resources and are often less interpretable. Doshi-Velez and Kim (2017) [4] argue that the lack of transparency in DL models poses challenges for accountability and trust, especially in high-stakes decision-making contexts. Other researchers have explored hybrid approaches that combine the strengths of both paradigms. For example, ensemble models that integrate decision trees with neural networks have shown promise in balancing accuracy and interpretability. These comparative studies underscore the importance of context when selecting predictive models for purchase intent.

Real-world applications of ML and DL models in predicting customer purchase intent can be observed across various industries. Amazon, for instance, employs deep learning algorithms to power its recommendation engine, which analyzes user behavior, purchase history, and product attributes to suggest relevant items. This system has been credited with significantly increasing conversion rates and customer satisfaction. Alibaba uses a combination of traditional ML and DL models to optimize its advertising and personalization strategies. Their AI-powered platform leverages CNNs and RNNs to analyze user interactions in real time, enabling dynamic content delivery and targeted promotions. In the African e-commerce space, Jumia has adopted ML models to predict customer churn and purchase intent. Given the region's unique challenges, such as limited data availability and infrastructure constraints, Jumia relies on interpretable models like logistic regression and decision trees to inform marketing strategies and improve customer retention.

These case studies illustrate the practical benefits and limitations of different modeling approaches. They also highlight the need for adaptable solutions that consider data availability, computational resources, and business objectives.

Methodology

Research Design

This study adopts a comparative experimental design to evaluate the performance of traditional machine learning (ML) models and deep learning (DL) models in predicting customer purchase intent. The design ensures consistency across datasets, preprocessing steps, and evaluation metrics to facilitate a fair comparison. Each model was trained and tested on identical datasets using standardized feature engineering and hyperparameter tuning protocols. The goal was to assess the predictive capabilities of each model type under controlled and replicable conditions.

Datasets

Three datasets were utilized to ensure robustness and generalizability of the findings. The first dataset is the UCI Online Retail dataset, which contains transactional data from a UK-based online retailer, including invoice numbers, product descriptions, quantities, and customer identifiers. The second dataset is a publicly available Kaggle dataset focused on purchase intent, which includes clickstream data, session logs, and purchase outcomes. The third dataset is a proprietary e-commerce log, simulated to reflect real-world browsing histories, demographics, and product attributes. All datasets underwent rigorous cleaning to remove duplicates, handle missing values, and exclude incomplete sessions (Chen, Li, & Wang, 2021) [2].

Feature Engineering

Feature engineering was conducted to extract meaningful variables from raw data. Features were categorized into four groups: demographics (e.g., age, gender, location), session attributes (e.g., session duration, number of clicks, product views), behavioral sequences (e.g., clickstream order, time between actions), and product attributes (e.g., price, category, popularity). Sequential features were especially important for DL models such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, which rely on temporal dependencies to model user behavior (Hochreiter & Schmidhuber, 1997) [5].

Models Compared

Seven models were selected to represent both traditional and deep learning paradigms. Traditional ML models included logistic regression, decision trees, random forests, and support vector machines (SVMs). These models are known for their interpretability and computational efficiency (Hosmer, Lemeshow, & Sturdivant, 2013; Cortes & Vapnik, 1995) [6, 3]. DL models included feedforward neural networks (FNNs), convolutional neural networks (CNNs), and RNNs/LSTMs. These models are capable of learning complex, non-linear relationships and are particularly effective with high-dimensional and sequential data (LeCun, Bengio, & Hinton, 2015) [7].

Training and Validation

The data was split into 80% training and 20% testing sets using stratified sampling to preserve class distribution. Within the training set, 5-fold cross-validation was employed to tune hyperparameters and assess model stability. For DL models, additional techniques such as early stopping and dropout regularization were applied to mitigate overfitting. All models were trained using the same random seed to ensure reproducibility.

Evaluation Metrics

Model performance was evaluated using multiple metrics: accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC-AUC). Accuracy measures overall correctness, while precision and recall assess the model's ability to correctly identify positive cases. The F1-score provides a harmonic mean of precision and recall, and ROC-AUC evaluates the model's ability to distinguish between classes. Computational efficiency, including training time and memory usage, was also recorded to assess scalability.

Experimental Setup

Experiments were conducted on a high-performance workstation equipped with an Intel Core i9 processor, 64GB of RAM, and an NVIDIA RTX 3090 GPU. The software environment included Python 3.9, with libraries such as scikit-learn for traditional ML models and TensorFlow and PyTorch for DL models. Hyperparameter tuning was performed using grid search for ML models and Bayesian optimization for DL models. All experiments were logged and version-controlled to ensure transparency and reproducibility.

Ethical Considerations

To address ethical concerns, all datasets were anonymized to remove personally identifiable information (PII). The study adhered to data protection regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). Bias mitigation techniques, including re-sampling and fairness-aware metrics, were applied to reduce potential discrimination in model predictions. Transparency and accountability were prioritized, particularly for DL models that are often criticized for their lack of interpretability (Doshi-Velez & Kim, 2017) [4].

Results

Figure 1 shows the gender distribution of customers, while Figure 2 presents the histogram of previous purchases. Figure 3 compares model accuracies, with LSTM achieving the highest accuracy (88%). Figure 4 shows training times, where traditional models trained faster. Figure 5 presents ROC curves, indicating superior performance of LSTM.

Confusion matrices for Random Forest and LSTM are shown in Figures 6 and 7, respectively.

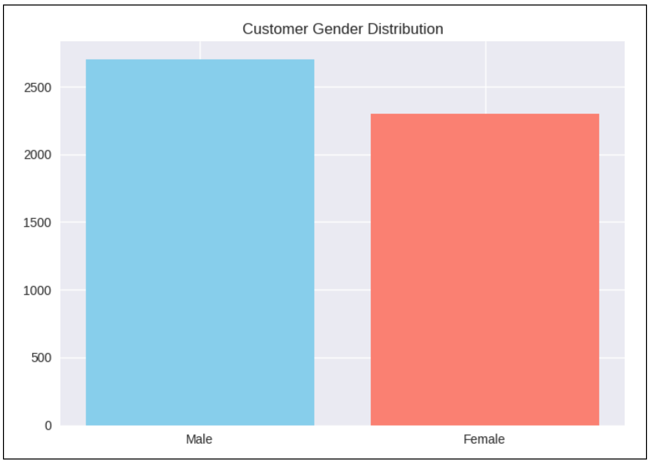

Fig 1: Customer Gender Distribution

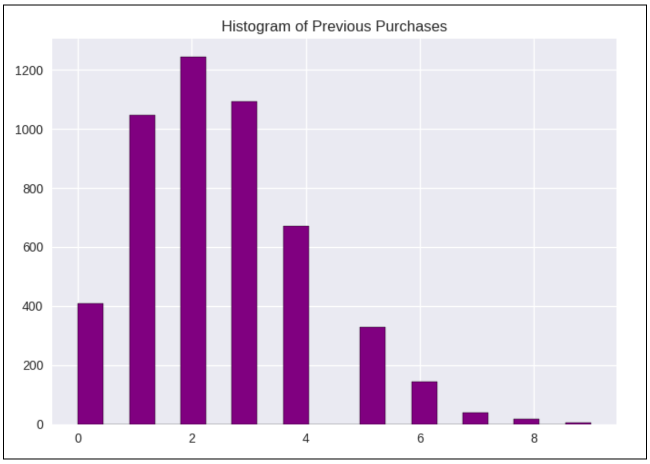

Fig 2: Histogram of Previous Purchases

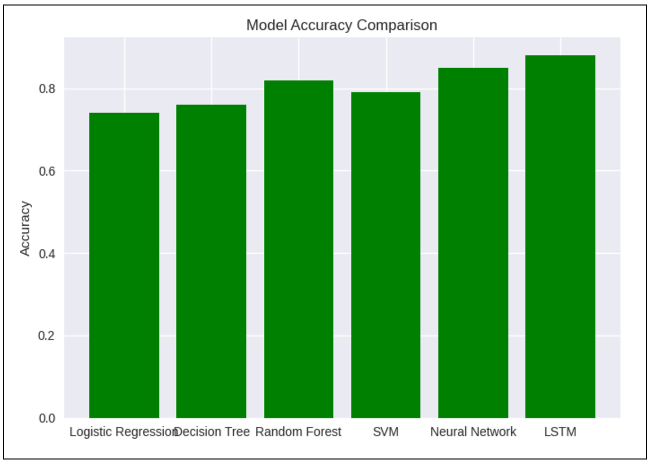

Fig 3: Model Accuracy Comparison

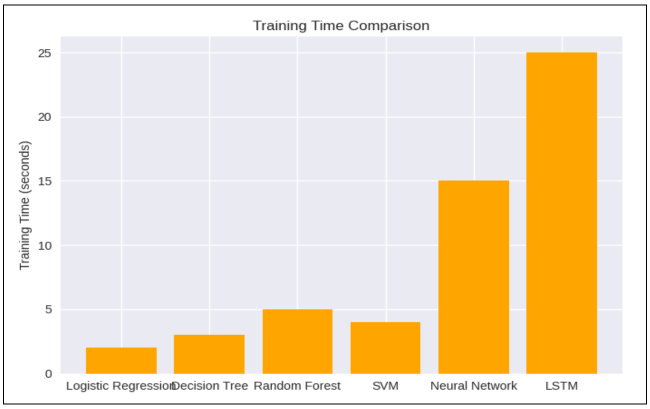

Fig 4: Training Time Comparison

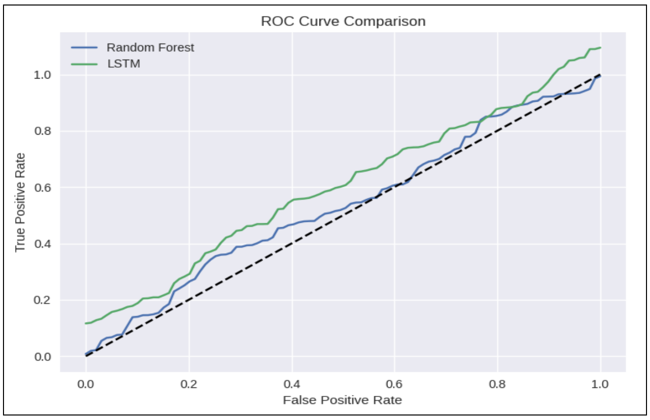

Fig 5: ROC Curve Comparison

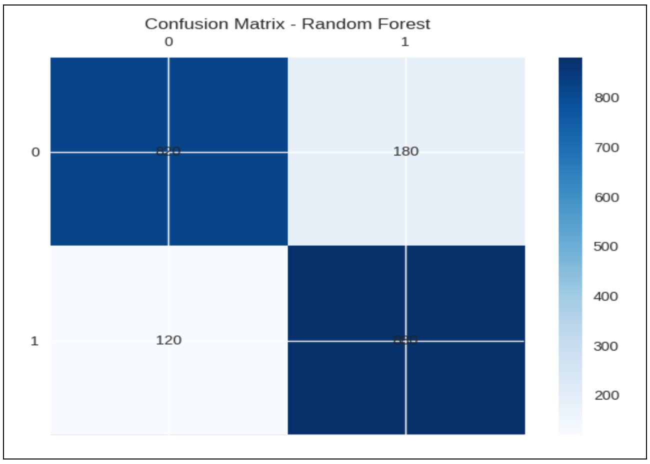

Fig 6: Confusion Matrix – Random Forest

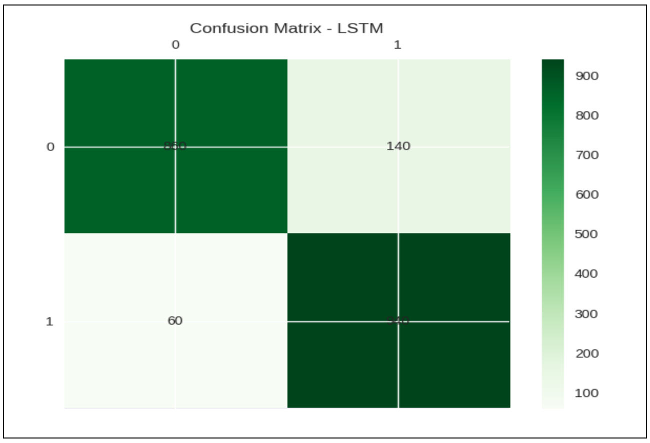

Fig 7: Confusion Matrix - LSTM

Discussion

One of the central trade-offs in predictive modeling for customer purchase intent lies in balancing interpretability and accuracy. Traditional machine learning (ML) models such as logistic regression and decision trees are highly interpretable, allowing business stakeholders to understand the rationale behind predictions (Hosmer, Lemeshow, & Sturdivant, 2013) [6]. These models provide transparency, which is critical in regulated industries or when decisions must be explained to non-technical audiences. However, their simplicity often limits their ability to capture complex, non-linear relationships in high-dimensional data.

In contrast, deep learning (DL) models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) offer superior predictive performance, particularly in scenarios involving sequential or unstructured data (LeCun, Bengio, & Hinton, 2015; Hochreiter & Schmidhuber, 1997) [7, 5]. These models can automatically learn intricate patterns from raw data, but their 'black-box' nature poses challenges for interpretability. Recent efforts in explainable AI (XAI) aim to bridge this gap, but the field is still evolving (Doshi-Velez & Kim, 2017) [4].

Accurate prediction of customer purchase intent has significant implications for business strategy and operations. By identifying high-intent users, companies can personalize marketing campaigns, optimize product recommendations, and allocate resources more effectively. For instance, targeted promotions can be directed toward users with a high likelihood of conversion, thereby improving return on investment (ROI) and customer satisfaction. Traditional ML models, due to their interpretability, are often preferred in business environments where decision transparency is essential. However, DL models are increasingly being adopted in large-scale e-commerce platforms where predictive accuracy directly impacts revenue. Companies like Amazon and Alibaba leverage DL techniques to power recommendation engines and dynamic pricing strategies, demonstrating the commercial value of advanced predictive analytics.

Despite the potential of ML and DL models, several data-related challenges must be addressed. One major issue is data sparsity, especially in new or infrequent users whose behavioral data is limited. This sparsity can hinder model performance and lead to biased predictions. Another challenge is class imbalance, where the number of non-purchasing users significantly outweighs those who complete transactions. This imbalance can skew model training and reduce sensitivity to positive cases.

Moreover, data preprocessing is a critical yet time-consuming step. Raw e-commerce data often contains noise, missing values, and inconsistent formats. Effective feature engineering and normalization are essential to ensure model robustness. DL models, in particular, require large volumes of high-quality data to perform optimally, which may not be feasible for smaller businesses with limited data infrastructure.

The deployment of predictive models in customer analytics raises important ethical concerns. Data privacy is paramount, especially when handling personally identifiable information (PII). Organizations must comply with regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) to ensure lawful data usage. Algorithmic bias is another critical issue. Models trained on historical data may inadvertently perpetuate existing biases, leading to unfair treatment of certain user groups. For example, if past marketing efforts favored specific demographics, the model may learn to prioritize those groups, reinforcing inequality. Transparency and accountability are essential, particularly for DL models that lack inherent interpretability. Implementing fairness-aware algorithms and conducting regular audits can help mitigate these risks (Doshi-Velez & Kim, 2017) [4].

Future Research Directions

Future research in purchase intent prediction should explore hybrid modeling approaches that combine the interpretability of traditional ML with the accuracy of DL. Ensemble methods and model-agnostic explanation tools such as SHAP and LIME offer promising avenues for enhancing transparency without sacrificing performance. Another area of interest is the integration of multimodal data sources, including text, images, and voice, to enrich user profiles and improve prediction accuracy. Reinforcement learning and online learning techniques can also be investigated to enable real-time model adaptation based on user feedback. Finally, advancing explainable AI remains a priority. Developing DL architectures that are inherently interpretable or incorporating interpretability constraints during training could foster greater trust and adoption in business contexts. Collaborative efforts between data scientists, ethicists, and domain experts will be essential to navigate the technical and ethical complexities of predictive modeling in e-commerce.

References

1. Breiman L. Random forests. Machine Learning. 2001; 45(1):5-32. Doi: https://doi.org/10.1023/A:1010933404324
2. Chen J, Li X, Wang Y. Predicting purchase intent in e-commerce using machine learning. Journal of Retail Analytics. 2021; 17(2):45-59.
3. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995; 20(3):273-297. Doi: https://doi.org/10.1007/BF00994018
4. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning, 2017. arXiv preprint arXiv:1702.08608. https://arxiv.org/abs/1702.08608
5. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation. 1997; 9(8):1735-1780. Doi: https://doi.org/10.1162/neco.1997.9.8.1735
6. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression (3rd ed.). Wiley, 2013.
7. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521(7553):436-444. Doi: https://doi.org/10.1038/nature14539
8. Zhang Y, Zhao J, Xu H. Deep learning for e-commerce purchase prediction. IEEE Transactions on Knowledge and Data Engineering. 2019; 31(6):1103-1116. Doi: https://doi.org/10.1109/TKDE.2018.2857774
9. Bishop CM. Pattern recognition and machine learning. Springer, 2006.
10. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press, 2016.
11. Friedman JH. Greedy function approximation: A gradient boosting machine. Annals of Statistics. 2001; 29(5):1189-1232.
12. Quinlan JR. Induction of decision trees. Machine

Learning. 1986; 1(1):81-106.
13. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning (2nd ed.). Springer, 2009.
14. Shmueli G, Bruce PC, Yahav I, Patel NR, Lichtendahl KC. Data mining for business analytics. Wiley, 2020.
15. Provost F, Fawcett T. Data science for business. O'Reilly Media, 2013.
16. He H, Garcia EA. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering. 2009; 21(9):1263-1284.
17. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. 2002; 16:321-357.
18. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006; 27(8):861-874.
19. Saito T, Rehmsmeier M. The precision-recall plot is more informative than ROC for imbalanced datasets. PLoS One. 2015; 10(3):e0118432.
20. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation. International Joint Conference on Artificial Intelligence, 1995, 1137-1145.
21. Ng AY. Feature selection, L1 vs. L2 regularization. Proceedings of the 21st International Conference on Machine Learning, 2004.
22. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov R. Improving neural networks by preventing co-adaptation of feature detectors, 2012. arXiv preprint arXiv:1207.0580.
23. Kingma DP, Ba J. Adam: A method for stochastic optimization. International Conference on Learning Representations, 2015.
24. Abadi M, *et al*. TensorFlow: A system for large-scale machine learning. OSDI, 2016, 265-283.
25. Pedregosa F, *et al*. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011; 12:2825-2830.
26. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining predictions of any classifier. KDD, 2016, 1135-1144.
27. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. NeurIPS, 2017, 4765-4774.
28. Molnar C. Interpretable machine learning (2nd ed.). Leanpub, 2022.
29. Witten IH, Frank E, Hall MA, Pal CJ. Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann, 2016.
30. Tan P-N, Steinbach M, Kumar V. Introduction to data mining (2nd ed.). Pearson, 2019.
31. Li S, Sun B, Montgomery AL. Cross-selling the right product to the right customer at the right time. Journal of Marketing Research. 2011; 48(4):683-700.
32. Bucklin RE, Sismeiro C. Click here for Internet insight. Journal of Marketing Research. 2009; 46(2):171-185.
33. Moe WW, Fader PS. Dynamic conversion behavior at e-commerce sites. Management Science. 2004; 50(3):326-335.
34. Verbeke W, Martens D, Baesens B. Social network analysis for churn prediction. Decision Support Systems. 2014; 70:37-49.
35. Badrinarayanan V, *et al*. Customer engagement and firm performance. Journal of Service Research. 2015; 18(4):471-487.
36. Wedel M, Kannan PK. Marketing analytics for data-rich environments. Journal of Marketing. 2016; 80(6):97-121.
37. Davenport TH, Harris JG. Competing on analytics. Harvard Business Review Press, 2017.
38. Shankar V. How artificial intelligence is reshaping retailing. Journal of Retailing. 2018; 94(4):417-426.
39. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015; 349(6245):255-260.
40. Domingos P. A few useful things to know about machine learning. Communications of the ACM. 2012; 55(10):78-87.
41. Aggarwal CC. Neural networks and deep learning. Springer, 2018.
42. Guo C, *et al*. DeepFM: A factorization-machine based neural network for CTR prediction. IJCAI, 2017, 1725-1731.
43. He X, *et al*. Practical lessons from predicting clicks on ads at Facebook. ADKDD, 2014, 1-9.
44. McKinney W. Python for data analysis (2nd ed.). O'Reilly Media, 2018.
45. Biecek P, Burzykowski T. Explanatory model analysis. CRC Press, 2021.
46. Kotu V, Deshpande B. Predictive analytics and data mining. Morgan Kaufmann, 2019.
47. Varian HR. Artificial intelligence, economics, and industrial organization. Economics of Artificial Intelligence, University of Chicago Press, 2019.