International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. , No. (Mon-20..)

http://dx.doi.org/10.12785/ijcds/XXXXXX

Ret-DNN: Predictive Analytics in Retail - Enhanced Deep Learning Models for Customer Behavior Analysis

immediate

Abstract: In the competitive landscape of retail e-commerce, understanding and predicting customer behaviour is challenging for business success. This study introduces the Retail Deep Neural Network (Ret-DNN) model, a novel approach of advanced deep learning techniques to enhance predictive analytics in the e-commerce domain. The Ret-DNN model excels in predicting various aspects of customer behaviour, providing deep insights into shopping habits, and transaction distribution, and identifying popular products based on sales data. The proposed model also offers a detailed analysis of customer purchase frequency and transaction patterns by country, enabling a comprehensive understanding of customer engagement. By accurately predicting behaviours, the Ret-DNN model equips businesses with the tools to optimize marketing strategies, improve customer satisfaction, and drive significant growth in the retail e-commerce business. The proposed Ret-DNN model minimizes prediction errors and increases prediction precision, demonstrating performance with the lowest Validation Mean Absolute Error (MAE) of 0.2531 and Root Mean Square Error (RMSE) of 0.3575, along with high accuracy rates of 0.91, 0.90, and 0.92 for validation, test, and training phases, respectively. Overall, this novel Ret-DNN model achieves an average accuracy of 91%, highlighting its effectiveness in predicting customer behaviour in retail e-commerce. A future research direction is also presented as a concluding remark.

Keywords: Customer Segmentation Techniques, Retail E-commerce, Purchase Patterns, Ret-DNN

1. INTRODUCTION

In the past decades, several studies have focused on the retail business of e-commerce, where the priority is to find a predictive analysis of customer purchase behaviour and engagement with the business. The retail e-commerce refers to selling goods and providing services to the customer by using the internet [1]. This retail business model provides effective activities and strategies for selling products online e-commerce platforms. E-commerce retailers operate the store by web interaction or using a mobile application, where customers can browse, read the product description, view images, and make purchases by online payment. This online platform includes secure availability, gateway, reviews and order tracking systems.

Despite progress, it remains unclear why some predictive models outperform others in forecasting customer behaviour. This study aims to address this gap by introducing the Retail Deep Neural Network (Ret-DNN) model, which leverages advanced deep learning techniques to enhance predictive analytics in retail e-commerce[2], [3]. The primary motivation for this research is to address this gap by answering the following question: Which deep learning model most effectively predicts customer behaviour in retail e-commerce? This study aims to address this gap by introducing the Retail Deep Neural Network (Ret-DNN) model, which leverages advanced deep learning techniques to enhance predictive analytics in retail e-commerce. This research evaluates the proposed Ret-DNN model's performance in predicting customer shopping habits, transaction distribution, popular products based on sales, customer purchase frequency, and transaction patterns by country.

The data used for this study were from publicly accessible e-commerce platforms of UK retail e-commerce platforms, encompassing a diverse range of customer interactions and transactions. The experimental findings of this research clearly show that the proposed Ret-DNN model outperforms traditional models, providing deep insights into customer behaviour and equipping businesses with the tools to optimize marketing strategies, improve customer satisfaction, and drive significant growth in the retail e-commerce sector[4].

The key contributions of this research are:

- **Retail Deep Neural Network (Ret-DNN) model:** Advanced deep learning techniques for predictive analytics in retail e-commerce are proposed.
- **Model Performance:** Ret-DNN model's ability to minimize prediction errors and ensure high precision, with outstanding performance metrics.
- **Predicting Analysis:** Offering a robust framework for understanding and predicting customer behaviour, with potential applications across e-commerce platforms.

The paper is structured into several sections, each addressing a specific aspect of the research. Section 2 discusses the work on e-commerce customer behaviour.

Corresponding Author E-mail address: kamruddin@aiub.edu

Section 3 of the work outlines the terms and procedures of the method in detail. Section 4 represents deep learning models with their working algorithms. Section 5 describes the Ret-DNN proposed model architecture and working process. Section 6 also describes the evaluation matrix. The model outcome and the primitive actions that should be taken as an outcome are discussed in Section 7. Section 8 is the conclusion of the paper.

2. BACKGROUND STUDIES

For customer perception analysis using deep Learning and NLP, S. Ramaswamy and N. DeClerck [5] proposed a system where their work focused on utilizing deep learning and natural language processing techniques to analyze and understand customer perceptions. The strength of their approach lies in integrating advanced technologies like deep learning and NLP to extract valuable insights from customer feedback and comments.

A. Hybrid Models

In the context of hybrid model architecture, Kim et. al. [6] hybrid deep learning model for predicting smartphone repurchase customer behavior. The deep hybrid learning model achieved a prediction accuracy higher than 90%, indicating its effectiveness in estimating user repurchase behavior [20] of same-brand smartphones. The study only considered smartphone users who replaced their smartphones within the past two years, potentially excluding users with longer experience with their devices. In the context of identified platform engagement and consumer characteristics as key predictors of online purchases. Nisha and A. S. Singh [7] proposed that they achieved better results than conventional machine learning techniques with their approach, which made use of deep neural networks. The research only focuses on a single e-commerce platform and the need for validation across many platforms and larger datasets. C. P. Gupta and V. V. Ravi Kumar [8] proposed a system that uses sentiment analysis to identify changing customer trends and preferences. This paper shows how sentiment analysis may help businesses extract useful information from text, such as emotions and attitudes, which can be applied to a variety of industries, including marketing, customer service, and social media analysis. The limitation of this paper is making a sentiment analysis model that understands the different meanings of words in various situations. Asniar and K. Surendra [9] proposed a system that uses Predictive Analytics to forecast customer behaviour, employing a Behavior Informatics and Analytics Approach [21], [22]. They applied predictive analytics to examine customer behavior by converting transaction data into behavioral patterns, leading to improved data-driven marketing decisions. Arwa A Altameem and Alaaeldin M Hafez [15] proposed a novel hybrid model that combines an optimized fuzzy deep belief network with a deep recurrent neural network (ODRNN) to predict customer behaviors. The main work of the study involved comparing this hybrid model with traditional machine learning algorithms like K-nearest neighbor, SVM, DNN, and CNN to determine the most effective method for customer behavior prediction. The strength of their approach lies in integrating advanced clustering techniques with deep learning models, which resulted in improved accuracy in predicting customer behaviors. In contrast to simpler models like SVM, the results' interpretability and the decision-making process underlying the predictions might not be as clear.

B. Neural Network

H. Zheng [10] presented an Artificial Neural Network (ANN)-based system for predicting and analyzing customer purchase behaviour in CRM data analysis technologies. The study focuses on the creation and implementation of an Artificial Neural Network (ANN) algorithm for predicting and evaluating customer purchasing behaviour using CRM data analysis technology. Aya Hasan Abdulqader Othman et al.[11] suggested a deep learning-based prediction model (DL-PM) to forecast social media activity and customer behavior. They aim to use deep learning techniques to evaluate massive volumes of data from social media and other sources to forecast customer behavior properly. Prakash et al. [12] introduced a predictive model based on different aspects of consumer behavior using Artificial Neural Network (ANN) for Direct-to-Consumer Brands. However, a challenge they might encounter is scaling up their ANN-based predictive model for wider applications beyond Direct-to-Consumer Brands. Chaudhuri and Gupta [13] proposed a methodology that combines machine learning and deep learning approaches to predict online customers' purchase behavior [23]. The importance of this study lies in identifying key predictors of online purchases, such as platform engagement metrics and customer-related factors, and showcasing the potential of deep learning in improving predictive accuracy.The strength of their approach lies in leveraging deep learning techniques to analyze complex datasets and uncover intricate patterns that traditional methods may overlook, thereby enhancing prediction capabilities. Their work's possibly lack of generalizability across various client demographics and product categories, however, is a weakness. This suggests the need for additional research to test and extend the findings to various buying settings. Sharam Dadashnia and Joerg Evermann [14] proposed a deep neural network trained on sequences of user actions to predict future user behavior in the context of the Dutch Employee Insurance Agency UWV. The importance of their work lies in their detailed analysis of usage patterns on the UWV website, identifying changes over time, and developing a predictive model using deep learning techniques. Their approach provides valuable insights into user behavior and process optimization for the UWV. The strength of their approach is the utilization of deep learning techniques to predict user actions, enabling personalized user assistance and resource allocation. Their use of historical data for prediction, which might not consider sudden changes in user behaviour or outside variables influencing user interactions, is a drawback of their study. The predictive model's accuracy and dependability in real-time circumstances, more validation and modification might

TABLE I. Background Study on Customer Behavior Prediction Models

Study	Key Findings	Limitations	Used Methods
[5]	Utilized deep learning and NLP to analyze customer perceptions	Data dependency	Deep Learning, NLP
[6]	Achieved over 90% accuracy in predicting smartphone repurchase behavior	Timeframe bias	Deep Hybrid Learning
[7]	Improved online purchase prediction performance using deep neural networks	Single platform focus	Deep Neural Networks
[8]	Identified changing customer trends and preferences through sentiment analysis	Contextual understanding	Sentiment Analysis
[9]	Used Predictive Analytics to forecast customer behavior	Restricted scope	Predictive Analytics, Behavior Informatics
[10]	Predicted customer purchase behavior using ANN in CRM data analysis	Data limitation	Artificial Neural Networks (ANN)
[11]	Forecasted social media activity and customer behavior with deep learning	Data quality	Deep Learning
[12]	Analyzed consumer behavior for Direct-to-Consumer Brands using ANN	Scalability	Artificial Neural Networks (ANN)
[13]	Combined ML and DL to predict online customer purchase behavior	Insufficient generalization	Machine Learning, Deep Learning
[14]	Predicted user behavior at UWV using deep neural networks trained on user actions	Sudden changes influencing user interactions	Deep Neural Networks
[15]	Developed a hybrid model (ODRNN) for predicting customer behaviors	Interoperability	Optimized Fuzzy Deep Belief Network, Deep Recurrent Neural Network
[16]	Conducted gender classification using deep learning techniques	Complexity and accuracy	Deep Learning
[17]	Integrated sensor fusion with RFID and deep learning for automated retail shopping	High initial cost	Sensor Fusion, RFID
[18]	Implemented an emotional tracking system for customer data analysis using deep learning	Misclassification in emotion and demographics	Deep Learning
[19]	Created a decision support system for customer behaviour analysis in smart shopping centers using video analysis	Video quality and occlusions	Video Analysis, Object Tracking

be required. Table I represents the summary analysis of the related research on customer behavior analysis.

C. Deep Learning

Bhat et al. [16] proposed a research paper on different approaches and algorithms for gender classification using deep learning techniques. The research work experiments to customize a model that can identify multiple features of a person's appearance without increasing operating time. The approach allows for the simultaneous extraction of multiple appearance features, enhancing the efficiency of gender classification. The model architecture may be relatively shallow to prevent overfitting, potentially impacting the complexity and accuracy of gender prediction compared to deeper networks. Venkatraman et al. [17] present a revolutionary approach to retail shopping by integrating sensor fusion with RFID tags and deep learning algorithms. The combination of sensor fusion technology with RFID tags to automate the checkout process and offer customers individualized recommendations based on their behaviour is the paper's key main work. The system aims to enhance customer satisfaction and increase sales by automating processes and providing targeted recommendations. Their strategy may have a drawback because deep learning algorithms and sensor fusion technologies demand a substantial upfront cost. Generosi et al. [18] introduce an emotional tracking system where they presented in the paper involves the implementation of a non-intrusive platform consisting of modules for face expression and emotions recognition, Also age, gender recognition, gaze detection, speech recognition, and biofeedback analysis was conducted. The strength of their approach lies in the utilization of deep learning algorithms to analyze a wide range of customer data, including emotions, age, gender, and gaze coordinates, in a non-intrusive manner. The confusion matrix results, which demonstrate misclassifications in emotion states, sex, and age-range predictions, suggest that one shortcoming of their work may be the possible difficulties in precisely predicting specific emotions or demographic traits. Also, Rathnasekara [19] proposed a decision support system for analyzing customer behaviour in smart shopping centres using deep learning and video analysis. The proposed system includes object tracking, action recognition, and queue management to provide automated reports on customer actions. The significant contribution of this work lies in the integration of deep learning techniques, such as object tracking and action recognition, to analyze customer behaviour in real-time. The strength of their approach provides the proper utilization of deep learning models for accurate object tracking and action recognition, enabling real-time analysis of customer behaviour in smart shopping centres. One potential restriction of their study could be their analysis's reliance on video data, which could provide difficulties in situations with low video quality or occlusions that could compromise the precision of client behaviour tracking.

3. METHODOLOGY

In this section, our research study describes the process of selecting a dataset and its attributes, followed by the selection and evaluation of deep learning models for predictive analytics in retail based on customer behaviour. The attributes provide comprehensive insights into customer purchasing patterns, which are crucial for developing effective predictive models. For predictive analytics, we have proposed the use of a Retail Deep Neural Network (Ret-DNN) model. The Ret-DNN model is particularly well-suited for capturing complex patterns in retail transaction data, making it an ideal choice for analyzing customer purchasing behaviours and trends. By leveraging the Ret-DNN model, we aim to predict future customer behaviours and enhance decision-making processes in retail operations.

The performance of the proposed Ret-DNN model is evaluated using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics. These metrics provide a robust measure of the model's accuracy and reliability in predicting customer behaviour. Figure 1 illustrates the workflow of the methodology, including data preprocessing, model training, and evaluation phases

A. Dataset

The dataset presents a comprehensive record of transactions from a UK-based online non-store retailer specializing in unique features[24]. The dataset covers December 1, 2010, and December 9, 2011, and provides a meticulous account of all transactions without missing data points. The dataset contains nominal six-digit transaction identifiers (InvoiceNo) that include c prefixes, indicating cancellations, five-digit product identifiers (StockCode), extensive product descriptions (Description), numerical quantities of each item per transaction (Quantity), date and time of the transaction (InvoiceDate), continuous numerical unit prices in sterling (UnitPrice), customer identifiers as five-digit numbers (CustomerID), and nominal country names where the customer is based (Country) shown in Table II. This dataset provides a valuable resource for conducting in-depth sales and customer behaviour analysis, managing inventory, and developing strategic plans for primarily wholesale clientele.

B. Data Pre-processing

First, we analyze an online retail transaction dataset. To make the data more easily understandable, we remove any missing information in the CustomerID column for customer analysis purposes. Following the cleanup, we organize the data according to customer transactions. The summary of each customer's purchases includes the total quantity of products purchased, the average price, and the number of transactions. To ensure that no feature overpowers others due to differences in scale, we use StandardScaler to normalize the features, achieving a mean of zero and a standard deviation of one.

We extract the date and time from the InvoiceDate, adding valuable temporal dimensions to the dataset. This

TABLE II. Description of the Online Retail Dataset Variables

Variable	Description
InvoiceNo	A 6-digit integral number uniquely assigned to each transaction. c indicates a cancellation.
StockCode	A 5-digit integral number uniquely assigned to each distinct product.
Description	The name of the product.
Quantity	The quantities of each product per transaction.
InvoiceDate	The invoice date and time when the transaction was created.
UnitPrice	Product price per unit in selling.
CustomerID	A 5-digit integral number uniquely assigned to each customer.
Country	The name of the country where the customer creates the product order.

information helps uncover patterns such as peak sales hours or seasonal trends, which is useful for time-series analyses or profiling customers based on their purchasing habits. After completing these preprocessing steps, the data is ready for various analytical pursuits, such as exploratory analyses, customer segmentation, and predictive modelling.

C. Data Analysis

This section explores customer behaviour through statistical methods and machine learning, drawing on our preprocessed dataset. We uncover patterns and predict future behaviours, ensuring our findings are robust with an 80% training, 10% testing, and 10% validation split.

1) Time of Day Distribution Customer Purchase

Customer purchase patterns across different times of the day. In the Figure 2 chart, the Afternoon segment constitutes a significant 42% of total purchases, illustrating a preference for shopping during these hours. This could be attributed to various factors such as lunch break shopping which is shown in Table III, daytime online browsing habits, or the culmination of morning decision-making. The Noon and Morning portions are equally substantial, each accounting for approximately 28% of purchases, indicating a consistent engagement with shopping activities during work hours or early daily routines. In Figure 2, contrasting, Evening purchases are remarkably lower at 2.3%, suggesting a steep decline in shopping as customers are likely for the day or engage in leisure activities that do not involve purchasing.

2) Customer Transaction By Country

To analyze the customer segments for our e-commerce business, we first need to determine how many customers are actively making purchases and generating transaction tokens. After analyzing our dataset, we found that the United Kingdom had the highest number of transactions, with a count of 492,979. This indicates that the United Kingdom is our most active market. Additionally, other countries such as Germany, France, and EIRE have significant transactions and are represented by substantial portions in Table IV. This comprehensive overview underscores the distribution of transactional activity across various countries, providing valuable insights into the global reach of business operations.

TABLE III. Reasons for Customer Purchase Patterns at Different Times of Day

Time of Day	Findings
Morning	Start of day purchases, routine shopping.
Noon	Midday breaks, quick errands during lunch.
Afternoon	Free time for shopping, post-lunch breaks, and completion of daily tasks.
Evening	Time for personal relaxation, family, or social outings, less focus on shopping.

TABLE IV. Top 10 Countries by Number of Transactions[24]

Country	Transactions
United Kingdom	492,979
Germany	9,493
France	8,556
EIRE	8,192
Netherlands	2,367
Spain	2,532
Belgium	2,069
Switzerland	2,001
Portugal	1,519
Australia	1,256

3) Top-Selling Product Categories

Table V presents a clear and visually appealing representation of the most popular product descriptions within an inventory or sales dataset. The products are ranked based on the frequency of their sales or the number of times they have been stocked. The inventory is quite diverse, with products ranging from home decorations, such as the White Hanging Heart T-light Holder, leading the chart to the Natural Slate Heart Chalkboard anchoring the bottom of the list. Each product has a distinctly colored bar, making it easy to differentiate them at a glance.

The data shows that retro-styled products like Jumbo

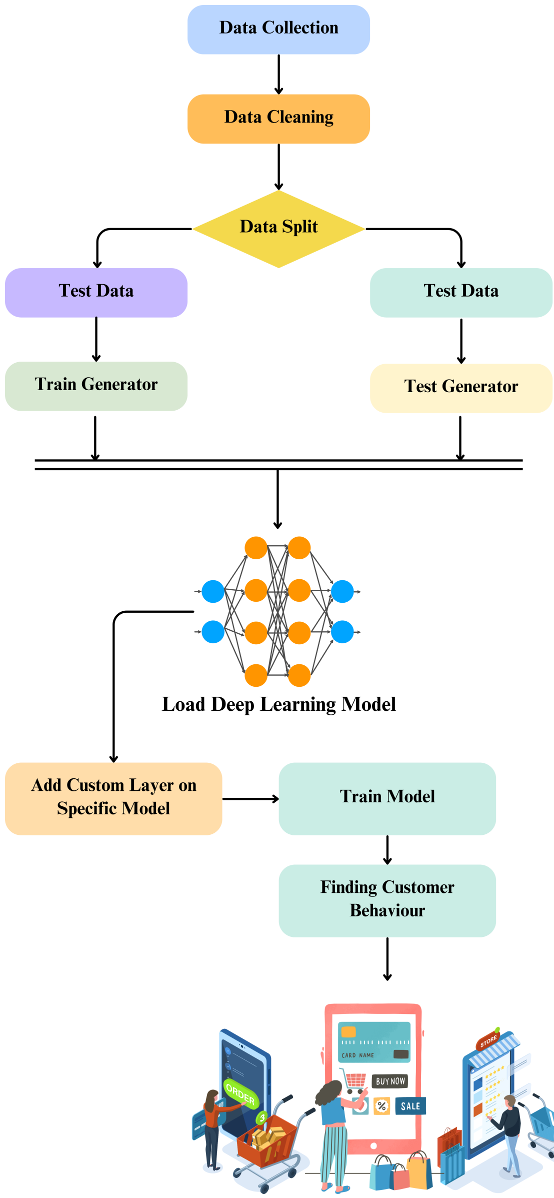


Figure 1. Methodology Workflow: The sequential stages in a data-driven decision-making process, from initial data preparation and cleaning through feature engineering and analysis to model evaluation and Propose Ret-DNN deep learning models.

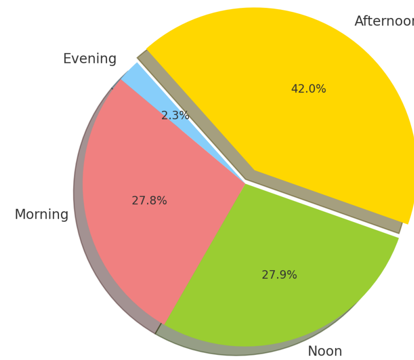


Figure 2. Customer shopping habits vary throughout the day, with distinct patterns emerging for morning, noon, afternoon, and evening periods.

TABLE V. Top 10 Product Prediction Based on Customer Purchase Frequency[24]

Product Description	Count
White Hanging Heart T-light Holder	2365
Regency Cakestand 3 Tier	2198
Jumbo Bag Red Retrospot	2156
Party Bunting	1726
Lunch Bag Red Retrospot	1638
Assorted Colour Bird Ornament	1501
Set of 3 Cake Tins Pantry Design	1473
Pack of 72 Retrospot Cake Cases	1385
Lunch Bag Black Skull.	1350
Natural Slate Heart Chalkboard	1280

Bag Red Retrospot and Lunch Bag Red Retrospot are in vogue. Kitchen items also seem popular, with Regency Cakestand 3 Tier and ‘Set of 3 Cake Tins Pantry Design’ featured prominently.

4. DEEP LEARNING MODELS

Understanding and predicting customer behaviour is paramount for sustained growth and competitiveness in the dynamic retail industry landscape. This study recognizes this imperative and introduces the Ret-DNN (Retail DNN) model. Built upon the foundation of deep neural networks (DNNs), Retail DNN extends its capabilities by integrating convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and recurrent convolutional neural networks (RCNNs). Customer behaviour analysis can help gather information about customers’ purchasing habits, interests in certain products, and purchase frequency.

A. Convolutional Neural Network

The CNN (Convolutional Neural Network) model begins by taking specific data shapes as input. This is followed by two sequences of a Conv1D layer and a MaxPooling1D layer which help in extracting and down-sampling features[25] . In the model architecture, the output is flattened and passed through two Dense layers. The first Dense layer contains 255 neurons with ReLU activation and the second Dense layer contains 128 neurons with ELU activation. Dropout layers with a 0.5 rate are interspersed between the dense layers to prevent overfitting. Finally, the model concludes with an output-dense layer that contains a single neuron with ELU activation to make the final predictions.

B. Recurrent Convolutional Neural Network

The RCNN model consists of various layers that enable it to process sequential data [26]. The model starts with a Conv1D layer that has 64 filters and ReLU activation, followed by a MaxPooling1D layer. Then, an LSTM layer

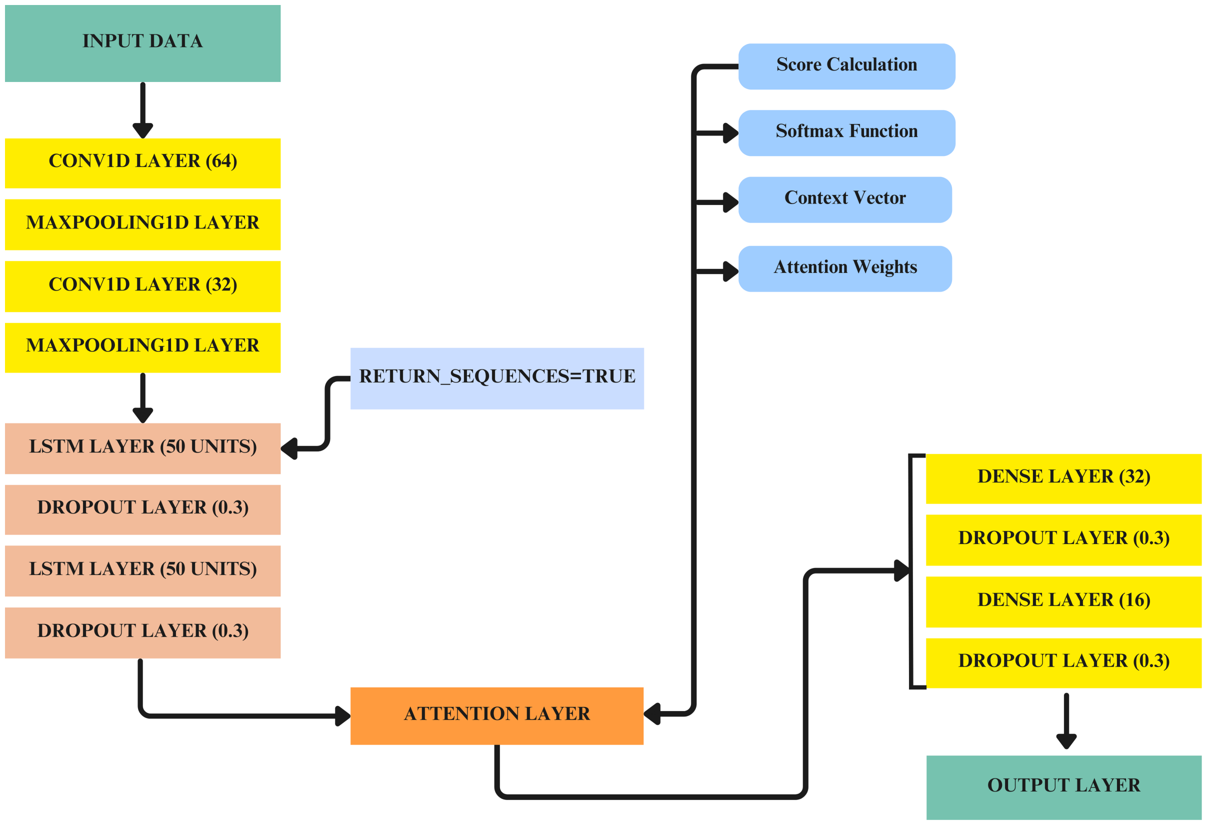


Figure 3. Architecture of the Retail Deep Neural Network (Ret-DNN) model. The model consists of an input layer that accepts scaled features, followed by two dense layers with 32 and 16 neurons, respectively, each using ReLU activation.

is used to capture the sequential information, and a Bidirectional LSTM layer is employed for temporal processing. To prevent overfitting, a Dropout layer is added, and a Dense layer with 64 neurons is used for feature extraction. Finally, the output is generated by a linear activation output layer.

C. Long Short-Term Memory

In the LSTM (Long Short-Term Memory) model architecture comprises an input layer, which is followed by an LSTM layer[27]. The LSTM layer is crucial for capturing long-term dependencies in sequential data. To prevent overfitting, a dropout layer is often included, and for more complex tasks, additional LSTM or dense layers may also be incorporated. The output layer generates predictions based on the task, with a single neuron for regression or multiple neurons for classification.

5. PROPOSED MODEL ARCHITECTURE

In Algorithm 1, the Retail Deep Neural Network (Ret-DNN) model architecture is designed to analyze customer behavior effectively. By configuring the model architecture in Table VI, the model begins with an input layer that accepts scaled features. It then processes these features through convolutional layers, recurrent layers, and attention mechanisms. The initial Conv1D layers, followed by MaxPooling1D layers, capture spatial features from the input data. Subsequently, LSTM layers capture temporal dependencies in the data, with dropout layers incorporated to prevent overfitting.

The architecture also includes an attention mechanism to enable the model to focus on the most relevant parts of the input sequence. Finally, dense layers with 32 and 16 neurons are utilized with ReLU activation functions, interspersed with dropout layers to maintain model regularization. The output layer, consisting of a single neuron with linear activation, predicts the scaled output quantity. This comprehensive architecture ensures robust feature extraction and prediction capabilities tailored to the nuances of customer behaviour analysis.

A. Feature Extraction (Spatial Features)

The model processes the input data through convolutional layers to extract spatial features. The first Conv1D layer applies 64 filters with a kernel size of 3 and uses the ReLU activation function to introduce non-linearity. This is followed by a MaxPooling1D layer with a pool size of 2 to reduce the dimensionality and retain important features. The second Conv1D layer, with 32 filters and a kernel size of 3, further processes the data, followed by another MaxPooling1D layer to continue dimensionality reduction.

TABLE VI. Ret-DNN Model Layer Configuration

Layer Type	Filters/Units	Kernel Size/Rate	Activation/Function
Input Layer	-	-	Scaled Features
Conv1D Layer	64 filters	Kernel Size: 3	ReLU
MaxPooling1D Layer	-	Pool Size: 2	-
Conv1D Layer	32 filters	Kernel Size: 3	ReLU
MaxPooling1D Layer	-	Pool Size: 2	-
LSTM Layer	50 units	-	Return Sequences: True
Dropout Layer	-	Rate: 0.3	-
LSTM Layer	50 units	-	-
Dropout Layer	-	Rate: 0.3	-
Attention Layer	-	-	-
Dense Layer	32 units	-	ReLU
Dropout Layer	-	Rate: 0.3	-
Dense Layer	16 units	-	ReLU
Dropout Layer	-	Rate: 0.3	-
Output Layer	1 unit	-	Linear

B. Sequence Modeling (Temporal Features)

After extracting spatial features, the model uses recurrent layers to capture temporal dependencies in the data. The first LSTM layer, with 50 units and `return_sequences=True`, processes the spatial features and outputs the full sequence. A dropout layer with a 30% dropout rate is applied to prevent overfitting. The second LSTM layer, also with 50 units, further processes the sequence, followed by another dropout layer with a 30% dropout rate.

C. Attention Mechanism

To enhance the model's ability to focus on the most relevant parts of the sequence, an attention mechanism is introduced. In customer behaviour analysis, the attention layer plays a crucial role by allowing the model to prioritize different parts of the input data based on their importance. For instance, in a sequence of customer interactions, not all interactions are equally informative for predicting future behavior. The attention mechanism assigns a weight to each interaction, indicating its importance in the sequence. The attention mechanism works as follows:

- **Score Calculation:** For each element in the input sequence, a score is calculated to determine its importance. This can be done using various methods, such as dot-product attention or additive attention. The score in Equation 1, e_i for the i-th element can be calculated as:
$$e_i = \text{score}(h_t, h_i) \tag{1}$$
where h_t is the hidden state of the target word and h_i is the hidden state of the input word.
- **Softmax Function:** The scores are passed through a softmax function to convert them into probabilities, which sum to 1. This helps in normalizing the importance of weights:
$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{T} \exp(e_j)} \tag{2}$$
In Equation 2, where α_i is the attention weight for the i-th element and T is the total number of elements in the sequence.
- **Context Vector:** The input sequence elements are weighted by their respective probabilities, and a context vector is created by summing these weighted elements. The context vector c_t captures the most relevant information from the sequence by Equation 3 as
$$c_t = \sum_{i=1}^{T} \alpha_i h_i \tag{3}$$
- **Attention Weights:** The context vector and the attention weights are used by the model to make the final prediction. The attention weights highlight which parts of the sequence had the most influence on the prediction, providing insights into customer behavior patterns.

D. Layer Architecture

1) (Dense Layers Final Processing)

The Ret-DNN model begins with the input layer, which receives the scaled features from the dataset. This layer can process an unspecified number of samples, each with a fixed number of features. For instance, if we have customer attributes such as age, income, and purchase history, the input layer will accept these features for further processing. Mathematically, the input shape is denoted as $(\text{None}, 10)$, where 10 represents the number of features. This flexible structure allows the model to handle varying batch sizes efficiently, as shown in Figure 3.

Following the input layer is the first dense layer, com-

Algorithm 1 The proposed Ret-DNN Model Architecture

```
Input Layer: Accepts the scaled features.
Convolutional Layers:
    Conv1D Layer:
    - Filters: 64
    - Kernel Size: 3
    - Activation Function: ReLU
    MaxPooling1D Layer:
    - Pool Size: 2
    Conv1D Layer:
    - Filters: 32
    - Kernel Size: 3
    - Activation Function: ReLU
    MaxPooling1D Layer:
    - Pool Size: 2
Recurrent Layers:
    LSTM Layer:
    - Units: 50
    - Return Sequences: True
    Dropout Layer:
    - Dropout Rate: 30%
    LSTM Layer:
    - Units: 50
    Dropout Layer:
    - Dropout Rate: 30%
Attention Mechanism:
    Attention Layer
Dense Layers:
    Dense (32) Layer:
    - Neurons: 32
    - Activation Function: ReLU
    Dropout (0.3) Layer:
    - Dropout Rate: 30%
    Dense (16) Layer:
    - Neurons: 16
    - Activation Function: ReLU
    Dropout (0.3) Layer:
    - Dropout Rate: 30%
Output Layer:
    Dense Layer:
    - Neurons: 1
    - Activation Function: Linear
    - Purpose: Predicts the scaled quantity.
Compile the model:
    Loss Function: Mean Squared Error (MSE)
    Optimizer: Adam
```

prising 32 neurons. This layer performs a linear transformation on the input data using a weight matrix and a bias vector, then applying the ReLU (Rectified Linear Unit) activation function. The ReLU function introduces non-linearity by outputting the input directly if it is positive and zero otherwise. This operation can be expressed mathematically as:

$$z_1 = W_1 \cdot x + b_1 \tag{4}$$

$$a_1 = \mathrm{ReLU}(z_1) \tag{5}$$

In the Equation 4,5, where W_1 is the weight matrix, x is the input vector, b_1 is the bias vector, z_1 is the weighted sum, and a_1 is the activation output. This layer transforms the input features into a higher-dimensional space, enabling the model to learn more complex patterns. The second dense layer further processes the data using 16 neurons. Similar to the first dense layer, it applies a linear transformation followed by the ReLU activation function:

$$z_2 = W_2 \cdot a_1' + b_2 \tag{6}$$

$$a_2 = \mathrm{ReLU}(z_2) \tag{7}$$

This layer reduces the number of neurons and distils the most relevant features extracted from the previous layer by Equation 6 and 7. This compression helps highlight the significant patterns and relationships within the data, making it easier for the subsequent layers to perform accurate predictions.

E. Drop Out Layer

The Ret-DNN model includes a dropout layer after the first dense layer to prevent overfitting. Overfitting occurs when the model performs exceptionally well on training data but fails to generalize to new, unseen data. The dropout layer mitigates this by randomly setting 30% of the neurons to zero during each training iteration. This regularization technique can be represented as:

$$a_1' = a_1 \cdot \mathrm{Dropout}(0.7) \tag{8}$$

where Dropout(0.7) is a mask that retains 70% of the neurons and sets the remaining 30% to zero in Equation 8. A second dropout layer also follows the second dense layer, applying the same 30% dropout rate to enhance regularization. This additional dropout layer reinforces the model's generalisation ability by ensuring it does not overfit the training data, maintaining a balance between model complexity and robustness. This process ensures the model does not become overly reliant on any single neuron, promoting robustness and generalization.

F. Output Layer

The final component of the Ret-DNN model is the output layer, consisting of a single neuron with a linear activation function. This layer is responsible for generating the model's prediction, which is a continuous value suitable for regression tasks. The linear activation function outputs the weighted sum of its inputs:

$$\hat{y} = W_3 \cdot a_2' + b_3 \tag{9}$$

In Equation 9, (W_3) is the weight matrix, (b_3) is the bias vector, and (haty) is the predicted output. This layer's simplicity is crucial for tasks that require predicting a continuous quantity, such as sales forecasting or customer lifetime value estimation.

6. PERFORMANCE MATRICS

To assess the model's performance, the study employed two commonly used evaluation metrics, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE)[28]. RMSE measures the differences between predicted and observed values by taking the square root of the average of the squared differences. At the same time, MAE calculates the average of the absolute differences between predicted and observed values are shown in Table VII. These metrics provide valuable insights into the accuracy and precision of the model's predictions.

A. MAE (Mean Absolute Error)

Measures the average absolute difference between the predicted values and the actual values [29] in Equation 10. Calculated using the formula:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{10}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of samples.

B. RMSE (Root Mean Square Error)

Equation 11 calculates the square root of the average of the squared differences between the predicted values and the actual values[30]. Computed using the formula:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{11}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of samples. Calculating the accuracy in Equation 12, True Positives (TP) are instances correctly identified as positive, while True Negatives (TN) are correctly identified as negative. False Positives (FP) occur when negatives are incorrectly labeled as positive, and False Negatives (FN) when positives are mislabeled as negative. These metrics collectively assess a model's classification accuracy [31].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

7. EXPERIMENTAL RESULT AND DISCUSSION

The performance metrics over the training and testing datasets, as presented in Table VII and Table VIII, provide a detailed analysis of various deep learning models' effectiveness in predictive analytics for the retail business. This table includes performance metrics, which are discussed in Section 6, where validation loss and Mean Absolute Error (MAE) for both training and testing phases help identify the strengths and weaknesses of each model.

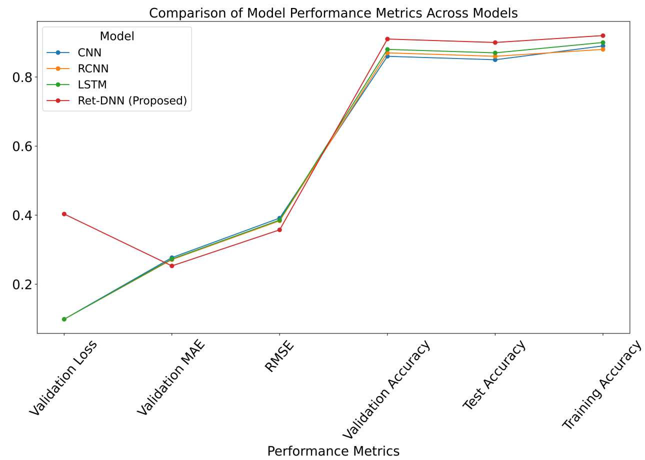


Figure 4. Comparative Analysis of Model Performance Across four different models (CNN, RCNN, LSTM, and Ret-DNN)

In analyzing customer behavior in retail e-commerce businesses using deep learning models, CNN, RCNN, LSTM, and the proposed Ret-DNN model, the Ret-DNN model performs better than others. It boasts the lowest Validation MAE of 0.2531, indicating that model predictions are more precise with the actual values. Its Root Mean Square Error (RMSE) is also the lowest at 0.3575, demonstrating its ability to minimize significant errors effectively. The Ret-DNN model's accuracy is equally impressive, with a Validation Accuracy of 0.91, Test Accuracy of 0.90, and Training Accuracy of 0.92 in Figure 4. As depicted in Figure 5, all the model results highlight the model's robust capabilities and practical learning from the training data, making it a perfect choice for predictive analytics in the retail industry.

Conversely, the CNN model presents a low validation loss of 0.0986 and performs the highest RMSE at 0.3915. The RCNN and LSTM models show balanced performance metrics, with the RCNN model having a slightly better Validation MAE of 0.2713 compared to LSTM's 0.2730. Both models exhibit solid accuracy, with RCNN achieving a Validation Accuracy of 0.87 and Test Accuracy of 0.86, while LSTM achieves a Validation Accuracy of 0.88 and Test Accuracy of 0.87. These metrics indicate that both models are reliable for predictions, although their performance varies slightly.

The proposed Ret-DNN model is novel because it uses advanced features like attention mechanisms, convolutional layers, and sequence modeling, which are not commonly used in existing research (Table IX). Unlike Ramaswamy and DeClerck [5], who only use deep learning techniques, the Ret-DNN model's attention mechanism helps it focus on important parts of the input data. Kim et al. [6] and Altameem and Hafez [15] use hybrid models but do not include convolutional layers, which limits their ability to analyze spatial features. Nisha and Singh [7] do not use

TABLE VII. Experimental Deep Learning Model Results

Model	Training Loss	Validation Loss	Test Loss	Training MAE	Validation MAE	Test MAE
CNN	0.0529	0.0979	0.0990	0.1911	0.2766	0.2775
	0.0428	0.0954	0.0965	0.1706	0.2721	0.2730
	0.0399	0.0954	0.0965	0.1641	0.2785	0.2794
	0.0353	0.1030	0.1041	0.1430	0.2808	0.2817
	0.0414	0.0986	0.0997	0.1687	0.2770	0.2779
RCNN	0.0937	0.1008	0.1020	0.2593	0.2715	0.2724
	0.0905	0.0995	0.1007	0.2643	0.2714	0.2723
	0.0935	0.0991	0.1003	0.2653	0.2714	0.2723
	0.0953	0.0988	0.1000	0.2659	0.2713	0.2722
	0.0902	0.1001	0.1013	0.2604	0.2715	0.2724
	0.0964	0.1025	0.1037	0.2675	0.2717	0.2726
LSTM	0.0926	0.1014	0.1026	0.2639	0.2728	0.2737
	0.0913	0.1007	0.1019	0.2619	0.2729	0.2738
	0.1010	0.1002	0.1014	0.2665	0.2729	0.2738
	0.0994	0.0990	0.1002	0.2705	0.2730	0.2739
	0.0954	0.0986	0.0998	0.2662	0.2730	0.2739
	0.0999	0.0983	0.0995	0.2716	0.2731	0.2740
Proposed Ret-DNN	**0.4576**	**0.4087**	**0.4100**	**0.3109**	**0.2535**	**0.2544**
	0.4608	**0.4074**	**0.4087**	**0.3173**	**0.2534**	**0.2543**
	0.3905	**0.4061**	**0.4074**	**0.2280**	**0.2534**	**0.2543**
	0.4608	**0.4048**	**0.4061**	**0.3021**	**0.2533**	**0.2542**
	0.4421	**0.4034**	**0.4047**	**0.2769**	**0.2531**	**0.2540**

TABLE VIII. Comparison of Model Performance Metrics

Model Name	Validation Loss	Validation MAE	RMSE	Validation Accuracy	Test Accuracy	Training Accuracy
CNN	0.0986	0.2770	0.3915	0.86	0.85	0.89
RCNN	0.0988	0.2713	0.3837	0.87	0.86	0.88
LSTM	0.0986	0.2730	0.3858	0.88	0.87	0.90
Ret-DNN (Proposed)	**0.4034**	**0.2531**	**0.3575**	**0.91**	**0.90**	**0.92**

TABLE IX. Comparative Analysis of Proposed Ret-DNN Model With Existing Study

Study	Attention Layer	Conv. Layers	Recurrent Layers	Dense Layers	Deep Learning	Hybrid Model	Sequence Layer
[5]	✗	✗	✗	✗	✓	✗	✗
[6]	✗	✗	✗	✓	✓	✓	✗
[7]	✗	✗	✗	✗	✓	✗	✗
[8]	✗	✗	✗	✗	✗	✗	✗
[15]	✗	✗	✓	✗	✓	✓	✗
Ret-DNN	✓	✓	✓	✓	✓	✓	✓

recurrent layers, which are important for understanding time-based data.

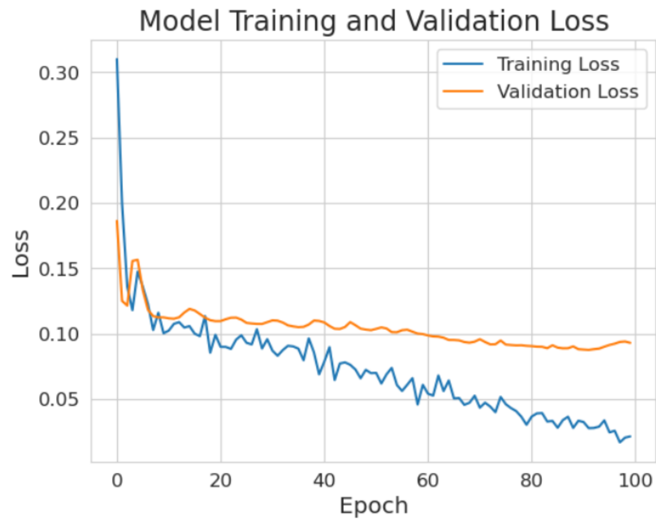
8. Conclusion And Future Work

The comparative analysis of deep learning models for predictive analytics in the retail business demonstrates the superiority of the DNN model over CNN, RCNN, and LSTM. As shown in Table VIII, the performance metrics highlight that the DNN model achieves the lowest Validation MAE of 0.2531 and the lowest RMSE of 0.3575. These metrics indicate that the DNN model provides the most accurate and reliable predictions for customer behaviour and engagement in e-commerce.

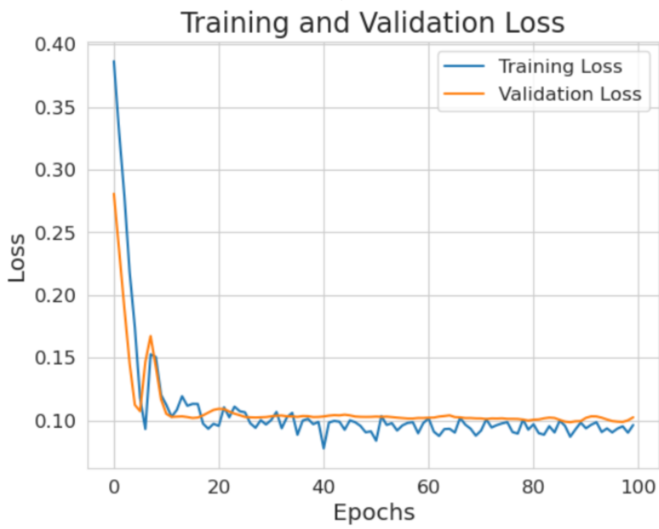
The literature review summary overview in Table I offers valuable insights into previous research findings. Studies [5] and [6] present the importance of deep learning and hybrid learning approaches but have limitations, such as data dependency and time frame bias. Other research [9] and [13], highlight the effectiveness of predictive analytics and the combined use of machine learning and deep learning models but encountered challenges in scope restriction and insufficient generalization.

To address the previous research, our study introduces a novel application of the Ret-DNN model for retail predictive analytics, addressing several limitations identified in previous research. Firstly, it mitigates data dependency by effectively handling various data types. Secondly, it addresses time frame bias by incorporating a more extensive

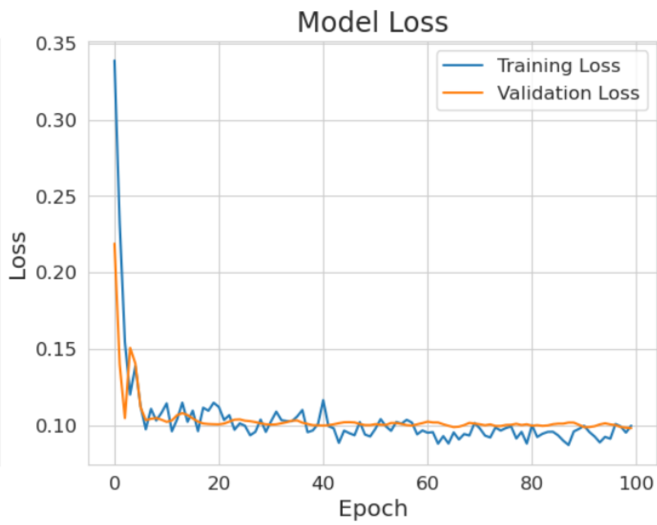
(a) CNN Model Training and Validation Loss



(b) Ret-DNN Model Training and Validation Loss



(c) RCCN Model Training and Validation Loss



(d) LSTM Model Training and Validation Loss

Figure 5. Comparative Analysis Graphs for Deep Learning Models With Proposed Ret-DNN

dataset over different periods. The study also overcomes the single platform focus of previous research by using data from multiple platforms, enhancing scalability and applicability. Additionally, the Ret-DNN model effectively manages issues with data quality and scalability and shows strong generalization capabilities, making it suitable for diverse retail scenarios.

With the advancements demonstrated by the Ret-DNN model, several open challenges remain for future research. Implementing real-time predictive analytics without compromising accuracy is a significant challenge in ensuring data privacy and security. Addressing these challenges will further enhance the effectiveness and applicability of predictive analytics in retail, paving the way for more advanced and reliable customer behaviour predictions.

References

[1] M. L. Arnumukti, S. Sudianto, and U. Athiyah, “Product layout recommendations based on customer behavior and data mining,” in *2023 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, 2023, pp. 330–334.

[2] S. Koli, R. Singh, R. Mishra, and P. Badhani, “Imperative role of customer segmentation technique for customer retention using machine learning techniques,” in *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, 2023, pp. 243–248.

[3] S.-S. Chen, T.-L. Li, Y.-C. Wu, and V. Singh, “An algorithm-based approach for mapping customer journeys by identifying customer browsing behaviors on e-commerce clickstream data,” in *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2023, pp. 1–8.

[4] A. Agatić, E. Tijan, S. Aksentijević, and A. Pucihar, “Internet of

behavior – the transformation of customer relationship management in logistics," in *2023 46th MIPRO ICT and Electronics Convention (MIPRO)*, 2023, pp. 1411–1415.

[5] S. Ramaswamy and N. DeClerck, "Customer perception analysis using deep learning and nlp," *Procedia Computer Science*, vol. 140, pp. 170–178, 2018.

[6] J. Kim, H. Ji, S. Oh, S. Hwang, E. Park, and A. P. del Pobil, "A deep hybrid learning model for customer repurchase behavior," *Journal of Retailing and Consumer Services*, vol. 59, p. 102381, 2021.

[7] Nisha and A. S. Singh, "Customer behavior prediction using deep learning techniques for online purchasing," in *2023 2nd International Conference for Innovation in Technology (INOCON)*, 2023, pp. 1–7.

[8] C. P. Gupta and V. V. Ravi Kumar, "Sentiment analysis and its application in analysing consumer behaviour," in *2023 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*, 2023, pp. 332–337.

[9] Asniar and K. Surendro, "Predictive analytics for predicting customer behavior," in *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*, 2019, pp. 230–233.

[10] H. Zheng, "Customer purchase behavior prediction and analysis based on crm data analysis technology," in *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, 2020, pp. 1374–1378.

[11] A. H. Abdulqader Othman, M. A. Alkhafaji, R. Hussam, N. H. Haroon, M. G. Majeed, and T. R. Al-Shaikhli, "Predicting consumer behaviour and results using social media and deep learning," in *2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS)*, 2023, pp. 1–7.

[12] Prakash, S. M. Babu, P. P. Kumar, S. Devi, K. P. Reddy, and M. Satish, "Predicting consumer behaviour with artificial intelligence," in *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, 2023, pp. 698–703.

[13] N. Chaudhuri, G. Gupta, V. Vamsi, and I. Bose, "On the platform but will they buy? predicting customers' purchase behavior using deep learning," *Decision Support Systems*, vol. 149, p. 113622, 2021.

[14] S. Dadashnia, T. Niesen, P. Hake, P. Fettke, N. Mehdiyev, and J. Evermann, "Identification of distinct usage patterns and prediction of customer behavior," *BPI Challenge*, 2016.

[15] A. A. Altameem and A. M. Hafez, "Behavior analysis using enhanced fuzzy clustering and deep learning," *Electronics*, vol. 11, no. 19, p. 3172, 2022.

[16] S. F. Bhat, A. W. Lone, and T. A. Dar, "Gender prediction from images using deep learning techniques," in *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2019, pp. 1–6.

[17] S. Venkatraman, G. N. K. R., and H. K. Machado, "i-shopping with sensor fusion for finding customer behavior using deep learning algorithm," in *2018 4th International Conference for Convergence in Technology (I2CT)*, 2018, pp. 1–6.

[18] A. Generosi, S. Ceccacci, and M. Mengoni, "A deep learning-based system to track and analyze customer behavior in retail store," in *2018 IEEE 8th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, 2018, pp. 1–6.

[19] H. Rathnasekara, "Customer behavior analysis in smart shopping centers using deep learning."

[20] A. Chowdhury, A. Mahmud, K. Nur, and H. Haque, "Predicting behavior trends among students based on personality traits," 2020.

[21] M. Subongkod, W. Sinlapasawet, C. Lalaeng, and B. Hongsakul, "The relationship between organizational politics and organizational citizenship behavior in the management perspective: Evidence in thailand," *Procedia Computer Science*, vol. 237, p. 819–826, 2024. [Online]. Available: http://dx.doi.org/10.1016/j.procs.2024.05.170

[22] K. Nur, M. Morenza-Cinos, A. Carreras, and R. Pous, "Projection of rfid-obtained product information on a retail stores indoor panoramas," *IEEE Intelligent Systems*, vol. 30, no. 6, pp. 30–37, 2015.

[23] A. M. Choudhury and K. Nur, "A machine learning approach to identify potential customer based on purchase behavior," in *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. IEEE, 2019, pp. 242–247.

[24] D. Chen, "Online Retail," UCI Machine Learning Repository, 2015, DOI: https://doi.org/10.24432/C5BW33.

[25] R. A. M. Rudro, M. F. A. Al Sohan, S. K. Chaity, and R. I. Reya, "Enhancing ddos attack detection using machine learning: A framework with feature selection and comparative analysis of algorithms," *Turkish Journal of Computer and Mathematics Education Vol*, vol. 14, no. 3, pp. 1185–1192, 2023.

[26] H. Xiong, L. Liu, and Y. Lu, "Artificial reef detection and recognition based on faster-rcnn," in *2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, vol. 2, 2021, pp. 1181–1184.

[27] B. J. T and D. Misbha, "Detection of attacks using attention-based conv-lstm and bi-lstm in industrial internet of things," in *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, 2022, pp. 402–407.

[28] R. A. M. Rudro and M. F. A. Al Sohan, "Utilization of machine learning strategies in the investigation of suspected credit card fraud," *Int. J. Advanced Networking and Applications*, vol. 15, no. 02, pp. 5869–5874, 2023.

[29] B. Jonathan, Z. Rahim, A. Barzani, and W. Oktavega, "Evaluation of mean absolute error in collaborative filtering for sparsity users and items on female daily network," in *2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 2019, pp. 41–44.

[30] I. A. Sabilla, M. Meirisdiana, D. Sunaryono, and M. Husni, "Best ratio size of image in steganography using portable document format with evaluation rmse, psnr, and ssim," in *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, 2021, pp. 289–294.

[31] N. M. Shailee, A. Alam, T. Ahmed, R. A. Mamun Rudro, and K. Nur, "Software bug prediction using machine learning on jm1 dataset," in *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS)*, 2024, pp. 01–06.