Journal of Organizational and End User Computing
Volume 36 • Issue 1

Marketing Decision Model and Consumer Behavior Prediction With Deep Learning

Anfeng Xu, Tourism School of Hainan Tropical Ocean University, Key Laboratory of Island Tourism Resource Data Mining and Monitoring of Ministry of Culture and Tourism, Sanya, China*

Yue Li, Economics and Management School of Harbin University of Science and Technology, Harbin, China

Praveen Kumar Donta, TU Wien, Austria

https://orcid.org/0000-0002-8233-6071

ABSTRACT

This article presents a study using ResNet-50, GRU, and transfer learning to construct a marketing decision-making model and predict consumer behavior. Deep learning algorithms address the scale and complexity of consumer data in the information age. Traditional methods may not capture patterns effectively, while deep learning excels at extracting features from large datasets. The research aims to leverage deep learning to build a marketing decision-making model and predict consumer behavior. ResNet-50 analyzes consumer data, extracting visual features for marketing decisions. GRU model temporal dynamics, capturing elements like purchase sequences. Transfer learning improves performance with limited data by using pre-trained models. By comparing the model predictions with ground truth data, the performance of the models can be assessed and their effectiveness in capturing consumer behavior and making accurate predictions can be measured. This research contributes to marketing decision-making. Deep learning helps understand consumer behavior, formulate personalized strategies, and improve promotion and sales. It introduces new approaches to academic marketing research, fostering collaboration between academia and industry.

KEYWORDS:

ResNet-50, GRU, Transfer Learning, Market Marketing, Predicting Consumer Behavior, Decision Models

1. INTRODUCTION

In today's information age, with the exponential growth in the scale and complexity of consumer data, businesses need a better understanding of consumer behavior and demand to formulate accurate marketing strategies. Traditional statistical methods and machine learning Alon et al. (2001) approaches may sometimes fail to capture the features and patterns present in the data adequately. Deep learning methods, with their powerful model representation and automatic feature learning capabilities, have emerged as effective tools to address these challenges.

DOI: 10.4018/JOEUC.336547 *Corresponding Author

The ResNet-50-GRU transfer learning method offers significant advantages and innovations compared to traditional methods. Firstly, traditional methods have limitations in handling large-scale and complex consumer data. With the exponential growth in consumer data in terms of scale and complexity, traditional statistical and machine-learning methods may fail to capture the features and patterns in the data adequately. This leads to an incomplete and inaccurate understanding of consumer behavior and demands.

In contrast, the ResNet-50-GRU transfer learning method utilizes the powerful model representation and automatic feature learning capabilities of deep learning algorithms to analyze and extract meaningful patterns and features from a large amount of consumer data. The ResNet-50 model, as a convolutional neural network, can automatically learn and extract visual features from image data, such as product preferences and brand preferences. This ability enables marketing decision-makers to gain more accurate insights into consumer preferences and tailor marketing strategies accordingly.

This study aims to leverage deep learning methods to construct a marketing decision model and predict consumer behavior. Specifically, we will utilize the ResNet-50 model to analyze consumer data containing images or visual content and extract visual features and patterns relevant to marketing decisions. Simultaneously, we will employ the GRU (Gated Recurrent Unit) model to model the temporal dynamics of consumer behavior, capturing time-related features such as purchase sequences and browsing histories. Furthermore, we will apply transfer learning techniques to leverage the knowledge from pre-trained models and build accurate models on limited data. We will collect and preprocess relevant data, including demographic information, browsing behavior, purchase history, and potential visual content of consumers. Next, we will train and evaluate deep learning models such as ResNet-50 and GRU using appropriate loss functions and optimization techniques. In the context of transfer learning, we will initialize the model with pre-trained weights and fine-tune it on specific market datasets.

Through experiments on real-world datasets, we will validate the effectiveness and accuracy of the proposed approach. Using the ResNet-50 model, we will successfully extract visual features relevant to marketing decisions, such as product preferences and brand preferences. Simultaneously, the GRU model will capture the temporal dynamics of consumer behavior and accurately predict future purchasing behavior and interests. The application of transfer learning will further enhance the model's performance, particularly in scenarios with limited data.

This research holds significant implications for marketing decision-making. Through deep learning methods, businesses can gain better insights into consumer behavior and demands, enabling them to formulate personalized marketing strategies and improve the effectiveness of product promotions and sales. Additionally, this study provides novel ideas and approaches for academic research in the field of marketing.

In the field of marketing, several deep learning or machine learning models are widely applied. Here are five common models along with their advantages and disadvantages:

1. Convolutional Neural Networks (CNN) He et al. (2017): CNN excels in processing image or visual data. It can effectively extract image features and demonstrates good classification and recognition performance. However, CNN's ability to handle time-series data and sequence modeling is relatively weak.
2. Recurrent Neural Networks (RNN) Xiao and Zhou (2020): RNN is a type of neural network capable of processing sequential data. It possesses memory capabilities to capture temporal dependencies within sequences. However, traditional RNNs suffer from the vanishing or exploding gradient problem and perform poorly in handling long-term dependencies.
3. Long Short-Term Memory (LSTM) Gers et al. (2000): LSTM is an improved RNN structure that addresses the gradient issues of traditional RNNs by introducing gate mechanisms. It can better capture long-term dependencies and is suitable for processing time-series data. However, LSTM models have more parameters and higher computational complexity.

4. Generative Adversarial Networks (GAN) Gonog and Zhou (2019): GAN is a framework consisting of a generator and a discriminator, trained through adversarial learning to generate realistic data samples. GANs hold the potential to generate novel marketing materials and promotional content. However, the training process of GANs is relatively unstable and requires more data and computational resources.
5. Transfer Learning Torrey and Shavlik (2010): Transfer learning speeds up the training process and improves model performance by utilizing pre-trained models on large-scale data and transferring their knowledge to new tasks. This is particularly important in marketing research when dealing with limited data.

The motivation of this study is to construct marketing decision models using deep learning methods and predict consumer behavior. To achieve this objective, we propose the following methodology:

Firstly, we will collect and preprocess relevant marketing data, including consumer demographics, browsing behavior, purchase history, and possible visual content. Secondly, we will analyze consumer data containing images or visual content using the ResNet-50 model. ResNet-50 is a deep convolutional neural network model with powerful image feature extraction capabilities. Through this model, we can extract visual features and patterns relevant to marketing decision-making, such as product preferences and brand preferences. Simultaneously, we will employ the GRU model to capture the temporal dynamics of consumer behavior, capturing time-related features such as purchase sequences and browsing history. GRU is a gated recurrent neural network model that exhibits better long-term dependency modeling capability and addresses the issue of vanishing gradients compared to traditional RNN models. Additionally, we will apply transfer learning techniques to fine-tune the pre-trained models' knowledge on limited market data to construct accurate models. Transfer learning helps us leverage existing large-scale data and model parameters, reducing the training time and improving the model's performance on market data.

The effectiveness and accuracy of this study's methodology will be validated through experiments on real-world datasets. Using the ResNet-50 model, we can successfully extract visual features relevant to marketing decision-making, and through the GRU model, we can capture the temporal dynamics of consumer behavior and accurately predict future purchasing behaviors and interests. The application of transfer learning further enhances the model's performance, especially in cases with limited data. This research holds significant importance for marketing decision-making. Through deep learning methods, businesses can better understand consumer behavior and needs, formulate personalized marketing strategies, and improve the effectiveness of product promotion and sales. Additionally, this research provides new ideas and methods for academic research in

- Application of Deep Learning in Marketing Decision-Making De Mauro et al. (2022): This study introduces common deep learning models in marketing research and details their advantages and limitations in addressing market data analysis and consumer behavior prediction. By utilizing deep learning methods, this research provides a new and powerful tool for marketing decision-making, extracting meaningful patterns and features from large-scale market data to help businesses better understand consumer behavior and demands.
- Comprehensive Analysis Combining Image and Time Series Data Gursch et al. (2021): The proposed methodology combines the analysis of image data and time series data. By using Convolutional Neural Networks (CNN) and Residual Networks (ResNet-50), visual features and patterns relevant to marketing decision-making can be extracted from images and visual content. Simultaneously, by employing Gated Recurrent Neural Networks (GRU), the temporal dynamics of consumer behavior, including purchase sequences and browsing history, can be captured. This comprehensive analysis approach enables a more holistic understanding and

prediction of consumer behavior, providing more accurate and nuanced information for marketing decision-making.

- Application of Transfer Learning in Marketing Research Chiu and Chuang (2021): The proposed methodology also utilizes transfer learning techniques to transfer knowledge from pre-trained models to new tasks, accelerating the training process and improving model performance. Transfer learning is particularly valuable in scenarios with limited data in marketing research. By leveraging existing large-scale data and model parameters, training time can be reduced, and model performance on market data can be enhanced. This application provides an effective approach for marketing research, leveraging existing knowledge and resources to better address real-world problems.

2. RELATED WORK

2.1 Sequence Models

Sequence models Keneshloo et al. (2019) have a wide range of applications in the field of marketing. They can be used for customer behavior analysis, purchase prediction, personalized recommendations, and sentiment analysis. By analyzing consumer browsing history, purchase records, and social media data, these models can capture the temporal patterns of consumer behavior, providing accurate predictions and insights.

The advantages of sequence models lie in their ability to capture long-term dependencies and dynamic evolution. They can learn and remember patterns in consumer behavior, helping marketers gain a deeper understanding of consumer preferences and behavioral changes. Additionally, sequence models can handle variable-length sequences, adapting to different lengths of behavior sequences and increasing model flexibility. Sequence models also have limitations. Firstly, training complex sequence models like LSTM requires significant computational resources and time. Secondly, interpreting the predictions of sequence models can be challenging, as understanding the model's reasoning process and decision-making can be difficult. Lastly, sequence models typically require a large amount of labeled training data to achieve good performance, which may require substantial data collection and annotation efforts.

2.2 Attention Mechanisms

Attention mechanisms Niu et al. (2021) have extensive applications in the field of marketing. Firstly, in personalized recommendations, they help identify the most important features or items in a user's browsing history or purchase records. By assigning weights to different elements, attention mechanisms improve the accuracy of recommendations, enhancing user satisfaction and engagement. Secondly, in sentiment analysis, attention mechanisms identify the keywords or phrases in a text that contribute the most to the expressed sentiment. This enables more accurate predictions of customer sentiment towards products, brands, or marketing campaigns, helping marketers understand customer emotions and adjust strategies accordingly. Lastly, in customer behavior analysis, attention mechanisms identify the most influential touchpoints or interactions in a customer's journey, allowing marketers to understand the factors driving customer conversions, churn, or engagement. This information can be used to optimize marketing campaigns, improve customer experiences, and enhance customer retention. Attention mechanisms have extensive applications in the field of marketing. Firstly, in personalized recommendations, help identify the most important features or items in a user's browsing history or purchase records. By assigning weights to different elements, attention mechanisms improve the accuracy of recommendations, enhancing user satisfaction and engagement. Secondly, in sentiment analysis, attention mechanisms identify the keywords or phrases in a text that contribute the most to the expressed sentiment. This enables more accurate predictions of customer sentiment towards products, brands, or marketing campaigns, helping marketers understand customer emotions and adjust strategies accordingly. Lastly, in customer behavior analysis, attention mechanisms identify the most influential touchpoints or interactions in a customer's journey, allowing marketers to understand the factors driving customer conversions, churn, or engagement. This information can be used to optimize marketing campaigns, improve customer experiences, and enhance customer retention.

2.3 Generative Adversarial Networks

Generative Adversarial Networks (GANs) Creswell et al. (2018) have wide applications in marketing, including generating realistic content and automating content creation. GANs can generate visually appealing product images and marketing materials, enhancing the overall aesthetic appeal and engaging customers. Additionally, GANs can be used to generate textual content, such as product descriptions or social media posts, saving time and resources in manual content creation.

GANs also have limitations. One challenge is the potential lack of diversity and quality control in generated content. GANs may produce unrealistic outputs or fail to meet desired standards, requiring careful monitoring and control. Furthermore, training GANs require a large amount of data and computational resources, making it resource-intensive and time-consuming.

GANs have valuable applications in marketing, enabling the generation of realistic visuals and automating content creation. They enhance the aesthetic appeal of marketing materials and provide a continuous stream of fresh and engaging content. However, challenges exist in maintaining quality control and managing computational resources. Despite these limitations, GANs have the potential to revolutionize marketing practices by providing innovative and efficient solutions for content generation.

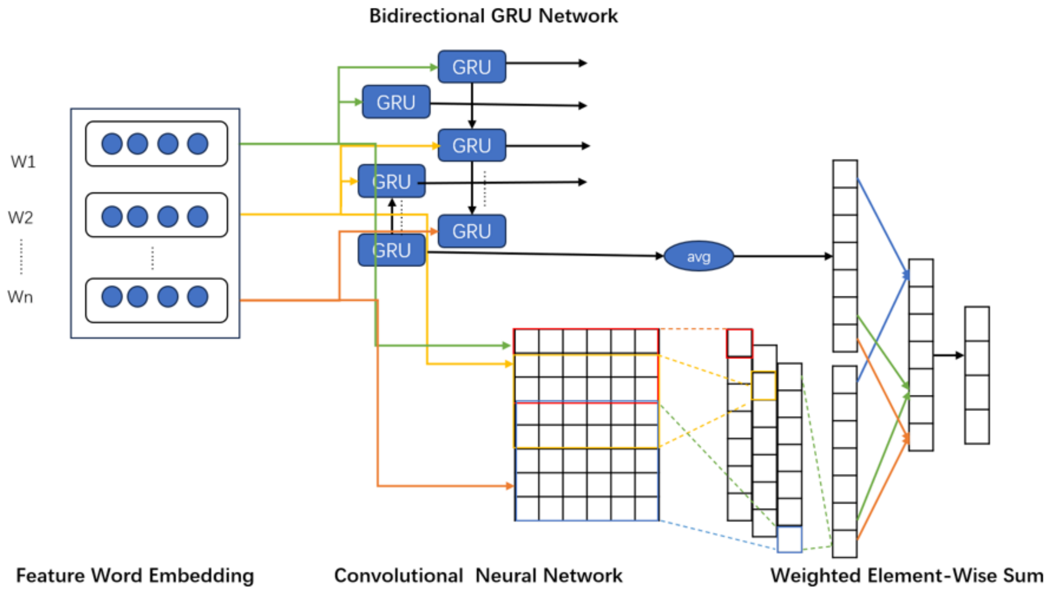
3. METHODOLOGY

3.1 Overview of Our Network

This study aims to construct a marketing decision model using ResNet-50, GRU, and transfer learning-based deep learning methods to predict consumer behavior. By combining these techniques, researchers can gain insights into consumer behavior using deep learning algorithms, image recognition models, and sequential data analysis, and make data-driven marketing decisions. Figure 1 shows the overall framework diagram of the proposed model:

Method Implementation Overview: 1. Data Collection and Preprocessing Famili et al. (1997): Collect relevant consumer data, including demographic information, browsing behavior, purchase history, and visual content, and preprocess it. This data will serve as input for the deep learning models. 2. Application of the ResNet-50 Model Koonce and Koonce (2021): Analyze consumer data containing images or visual content using the pre-trained ResNet-50 model. By passing the data through the ResNet-50 network, extract visual features and patterns relevant to marketing decisions, such as product preferences and brand preferences. 3. Application of the GRU Model Kumar and Abirami (2021): Model the temporal dynamics of consumer behavior using the GRU model. Input sequential data such as browsing history, purchase history, and social media interactions into the GRU model. The GRU model will learn patterns and trends in the time-series data and can predict future purchase behavior and interests. 4. Application of Transfer Learning Soni (2018): Initialize the pre-trained ResNet-50 model and GRU model's weights using transfer learning for marketing-specific datasets. Fine-tune or retrain these models to adapt them to specific marketing tasks. During the fine-tuning or retraining process, adjust model hyperparameters, layer configurations, and structures based on the specific task requirements. 5. Model Evaluation and Prediction Gauch et al. (2003): Use the trained models to make predictions and evaluate new consumer data. By inputting new data into the ResNet-50 and GRU models, obtain predictions regarding consumer behavior and preferences. Based on the prediction results, marketers can formulate personalized marketing strategies to enhance the accuracy and effectiveness of marketing decisions.

Figure 1. Overall flow char of the model

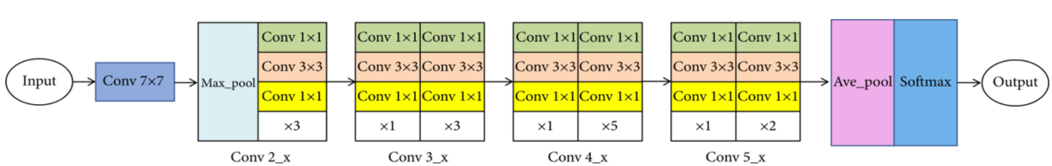


By combining ResNet-50, GRU, and transfer learning, researchers can gain deep insights into consumer behavior from marketing data. The implementation process of these methods includes data collection and preprocessing, application of the ResNet-50 model, application of the GRU model, application of transfer learning, and model evaluation and prediction. This combination of steps can help marketers make data-driven decisions and improve the effectiveness and return on marketing activities.

3.2 ResNet-50

ResNet-50 Alzubaidi et al. (2021) is a deep convolutional neural network model that addresses issues like vanishing gradients and exploding gradients Hanin (2018) when constructing deep networks, enabling training and optimization of deeper networks. ResNet-50 is based on the concept of residual learning He et al. (2016) and uses residual blocks to construct deep networks. In traditional convolutional neural networks, information needs to be passed through each layer, and in deep networks, backpropagation causes gradients to diminish, making it difficult to optimize the network. To overcome this problem, ResNet-50 introduces skip connections, allowing information to bypass certain layers in the network and preserve and propagate more gradient information. In ResNet-50, each residual block consists of two convolutional layers and an identity mapping. The convolutional layers extract features, while the identity mapping directly passes the input to the output by adding it to the output of the convolutional layers. This way, even with deeper networks, gradients can

Figure 2. Schematic diagram of ResNet-50



propagate easier to earlier layers, avoiding the issue of vanishing gradients. Figure 2 is a schematic diagram of the ResNet-50.

Role of ResNet-50 in building marketing decision models: In marketing decision models, ResNet-50 is primarily used to analyze consumer data that includes images or visual content, extracting visual features and patterns relevant to marketing decisions. By training on large-scale image datasets, it can learn low-level and high-level features such as product preferences, brand preferences, and visual attractiveness. These features can help businesses understand consumer preferences regarding product appearance, advertising communication, and brand image, optimizing product design, ad creativity, and marketing strategies.

Although ResNet-50 performs well in many computer vision tasks, there are limitations and potential drawbacks to consider when applying it to building marketing decision models:

1. Data requirements: ResNet-50 requires large-scale labeled image data for training to learn effective visual features. This may pose challenges in data collection and annotation, especially for specific products or brands in certain marketing scenarios.
2. Compute resource requirements: ResNet-50 is a large network model that demands significant computational resources and storage space for training and inference. When applying it to real-world marketing decisions, the availability and cost of computational resources need to be considered.
3. Interpretability and explainability: Deep learning models like ResNet-50 are often seen as black-box models, making it challenging to understand their internal workings. In some cases, the interpretability of model predictions and features might be important for marketing decision-makers.

To balance the applicability and potential drawbacks of ResNet-50, the following measures can be considered:

1. Data augmentation and transfer learning: Data augmentation techniques can expand the dataset and alleviate data requirements. Additionally, transfer learning techniques can leverage pre-trained ResNet-50 model weights on large-scale image datasets and fine-tune them on specific marketing datasets to improve model performance.
2. Model interpretability approaches: To increase the interpretability and explainability of the model, complementary interpretable machine learning methods or tools such as feature importance analysis, local interpretability methods (e.g., Grad-CAM), etc., can be combined to explain the predictions and visual features of the ResNet-50 model.
3. Integration of multiple models: In addition to ResNet-50, considering the combination of other types of models such as text models, recommendation system models, etc., can provide a more comprehensive prediction and understanding of consumer behavior by integrating different types of data and features.

The formula of ResNet-50 is as follows:

$$\mathbf{y} = \mathcal{F}\left(\mathbf{x}, W_i\right) + \mathbf{x} \tag{1}$$

In formula (1), we can see the following variables:

x: Input features, which can be images or other forms of data. y: Output features, obtained by extracting features from the input features using the ResNet-50 model. W_i: Set of model parameters, including all the weights and biases in the ResNet-50 model. (): Represents the residual block function of the ResNet-50 model. It takes the input features x and model parameters W_i and computes the residual features. This formula expresses the basic principle of the ResNet-50 model, where the model transforms the input features x into output features y by learning the residual features (x, W_i) and adding them to the input features. ResNet-50 can effectively train deep networks by introducing residual learning and addressing issues like vanishing and exploding gradients.

The ResNet-50 model plays a crucial role in building marketing decision models. It can extract valuable features and patterns from visual data, helping businesses understand consumer behavior and demands, and optimizing marketing strategies. However, it is important to consider the data requirements, compute resource requirements, and model interpretability limitations of ResNet-50 and balance its applicability and potential drawbacks through data augmentation, transfer learning, model interpretability approaches, and integration of multiple models.

3.3 GRU

GRU (Gated Recurrent Unit) Dey and Salem (2017) is a variant of recurrent neural networks (RNNs) Salehinejad et al. (2017) used for processing sequence data and capturing their inherent temporal dependencies. Compared to traditional RNN models, GRU introduces gating mechanisms that better handle long-term dependencies and alleviate the issues of vanishing and exploding gradients. Figure 3 is a schematic diagram of the GRU.

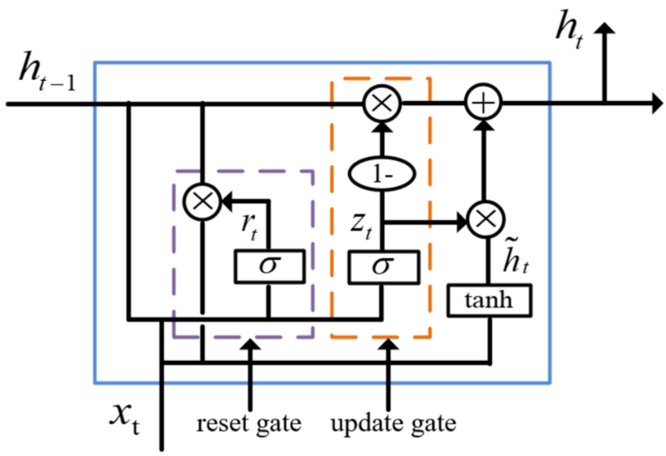
The GRU model consists of several key components, including the update gate, reset gate, and candidate hidden state. Here is a detailed explanation of the basic principles and roles of these components:

1. Update Gate: The update gate determines the importance of the previous hidden state at the current time step. It calculates a value between 0 and 1 by considering the input data and the previous hidden state. This value represents how much of the previously hidden state information should be retained. The update gate controls the flow of information, allowing the model to selectively remember or forget past information.
2. Reset Gate: The reset gate determines the impact of the previous hidden state on the current time step. Similar to the update gate, it calculates a value between 0 and 1 based on the input data and the previous hidden state. This value represents how much of the past hidden state information should be retained. The reset gate controls the forgetting of past information, enabling the model to better adapt to the current input.
3. Candidate Hidden State: The candidate hidden state is a temporary hidden state calculated based on the current input data and the previous hidden state before the application of the update and reset gates. It incorporates the input information at the current time step and captures important features of the current sequence.

4. Final Hidden State: The final hidden state is obtained by applying the weights of the update gate, reset gate, and candidate hidden state. It combines the information from the previous hidden state and the current input, representing the model's understanding and memory of the current sequence data. The final hidden state can be used as an output for predicting the next time step's outcome or for other related tasks.

By utilizing these components and computations, the GRU model can model long-term dependencies in sequence data while mitigating the issues of vanishing and exploding gradients. Compared to traditional RNN models, GRU has fewer parameters, higher computational efficiency, and demonstrates better performance in various sequence modeling tasks.

Figure 3. Schematic diagram of GRU



In building marketing decision models and predicting consumer behavior, the GRU model plays a crucial role in time series data analysis. Through the GRU model, it is possible to model and capture the temporal dynamics in consumer behavior, such as purchase sequences, browsing history, and more. The GRU model can learn and remember long-term dependencies in sequences, enabling more accurate predictions of future behavior and interests.

Although the GRU model performs well in sequential data modeling and prediction, it also has some limitations and potential drawbacks to consider. These include: Data dependency: The performance of the GRU model in modeling sequential data highly depends on the quality and availability of the data. If the data quality is low, lacks representativeness, or contains noise, the model's performance may degrade. Computational complexity: The GRU model has higher computational complexity compared to simple linear models or traditional machine learning methods. This implies that computational resource constraints need to be considered when dealing with large-scale datasets or real-time applications. Hyperparameter tuning: The GRU model has several hyperparameters to be tuned, such as the number of hidden units, learning rate, etc. Proper selection of hyperparameters is crucial and requires experimentation and validation.

To balance the applicability and potential drawbacks of the GRU model, researchers and practitioners can consider the following strategies: Adjust model complexity: Depending on the actual requirements and computational resource constraints, the complexity of the GRU model can be adjusted, such as reducing the number of hidden units or layers to improve computational efficiency.

Data preprocessing and cleaning: Preprocessing and cleaning the data before using the GRU model is essential. This includes handling missing values, outliers, and noise, as well as performing feature selection and dimensionality reduction to improve the model's robustness and predictive ability. Integration with other methods: The GRU model can be combined with other machine learning or statistical methods to leverage their strengths. For example, combining the GRU model with traditional regression models Austin et al. (2001) or clustering methods Omran et al. (2007) can provide more comprehensive and accurate prediction results. Model evaluation and validation: To evaluate the performance and applicability of the GRU model, rigorous experiments and validations are required. This includes using techniques such as cross-validation, comparative experiments, and metric evaluation to assess the model's accuracy, generalization ability, and stability.

The formula of GRU is as follows:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])\ r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}t = \tanh(W \cdot [r_t \odot ht - 1, x_t]) \quad (2)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

Here is the translation of the variables used in the equations:

z_t represents the update gate, which determines the importance of the previous hidden state's weight at the current time step.

r_t represents the reset gate, which determines the impact of the previous hidden state on the current time step.

$\tilde{h}_t$ represents the candidate hidden state, which is a temporary hidden state calculated based on the current input data and the previous hidden state before applying the update and reset gates.

h_t represents the final hidden state, which is obtained by applying the weights of the update gate, reset gate, and candidate hidden state. It combines the information from the previous hidden state and the current input.

W_z, W_r, and W are weight matrices used to map the input data and hidden state to appropriate dimensions.

$[h_{t-1}, x_t]$ represents the concatenation of the previous hidden state h_{t-1} and the current input data x_t. represents element-wise multiplication, σ represents the sigmoid function, and tanh represents the hyperbolic tangent function.

These equations describe the computations performed in the GRU model, including the calculation of the update gate, reset gate, candidate hidden state, and the update of the final hidden state. By utilizing these computations, the GRU model can capture temporal dependencies in sequence data and effectively handle long-term dependencies.

In research, the GRU model can be applied to predict sequential patterns in consumer behavior. For example, it can be used to analyze and forecast time-series patterns of consumer purchasing behavior, such as purchase frequency, amount, or the order of buying decisions. By training on historical consumer behavior data, the GRU model can learn important features and patterns in the sequential data and utilize them for predicting future consumer behavior.

3.4 Transfer Learning

Transfer learning is a machine learning technique that leverages knowledge gained from pre-trained models to improve the performance of a model on a target task or dataset. The basic principle of transfer learning is to transfer the learned representations or knowledge from a source domain to a target domain, where the source domain typically has a large amount of labeled data available, while the target domain has limited labeled data. Figure 4 is a schematic diagram of the Transfer Learning.

In transfer learning, the pre-trained model, often trained on a large-scale dataset like ImageNet for computer vision tasks, is used as a feature extractor or a starting point for training the target model. The idea is that the pre-trained model has learned general features that are transferable across different tasks, such as low-level visual features like edges, and textures, or high-level semantic features like object shapes or concepts. These features capture valuable information that can be relevant to the target task.

The role of transfer learning in the proposed marketing decision-making model is to enhance the model's performance when the available target data is limited. By initializing the model with the pre-trained weights, the model starts with a better initialization point, which can help in faster convergence and prevent overfitting on limited target data. Fine-tuning is then performed on the target task using the limited target data, allowing the model to adapt the learned representations to the specific characteristics of the target domain.

However, there are limitations and potential drawbacks to consider when using transfer learning. One limitation is the assumption that the source and target domains share some common underlying features or distributions. If the two domains are vastly different, the transfer of knowledge may not be effective, and the pre-trained model may not provide significant benefits. It is crucial to carefully select the pre-trained model based on its relevance to the target task.

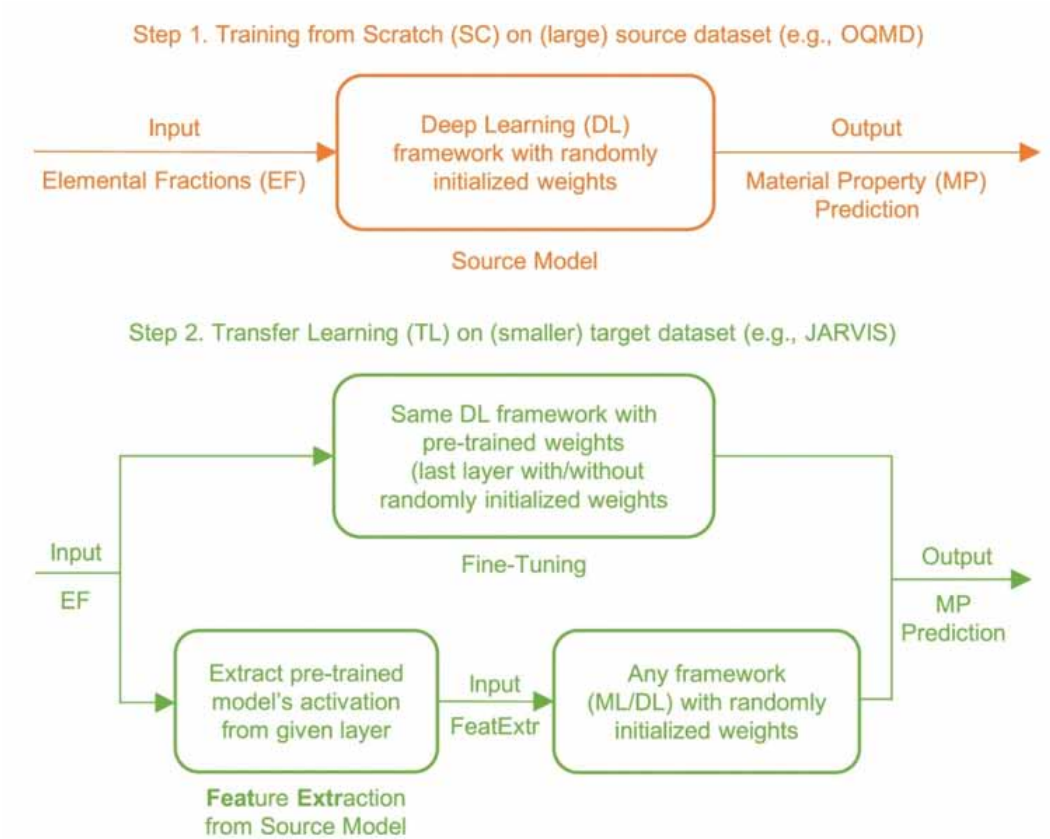
Another limitation is the risk of negative transfer, where the pre-trained model's knowledge may conflict with or hinder learning on the target task. This can happen if the source domain contains biased or irrelevant information that does not apply to the target task. It is important to evaluate the performance of the transferred model on the target task and ensure that it is improving or at least not deteriorating compared to training from scratch.

To balance the applicability and potential drawbacks of transfer learning, it is essential to consider a few factors. Firstly, selecting a pre-trained model that is trained on a large and diverse dataset, capturing general features that are likely to be useful across different tasks. Secondly, conducting thorough evaluation and validation on the target task to ensure that the transferred model is indeed improving the performance and providing meaningful insights. Lastly, fine-tuning the transferred model on the target task using limited target data to adapt the learned representations to the specific characteristics of the target domain.

The formula of Transfer learning is as follows:

$$\theta_f^* = \arg\min_{\theta_f} \left(\frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(y_i, f\left(x_i, \theta_f\right)\right) + \lambda \cdot R\left(\theta_f\right) \right) \quad (3)$$

Figure 4. Schematic diagram of Transfer Learning



Where:

θ_f represents the optimized parameters of the target task model. θ_f denotes the parameters of the pre-trained source model. N refers to the number of samples in the target task dataset. x_i represents an input sample from the target task dataset. y_i represents the corresponding target output for x_i in the target task dataset. $f(x_i, \theta_f)$ represents the output of the model f given input sample x_i and parameters θ_f . L represents the loss function that measures the discrepancy between the predicted outputs and the ground truth. λ is a hyperparameter that controls the trade-off between the loss term and the regularization term. R denotes the regularization term that encourages desirable properties in the model, such as simplicity or smoothness. The equation represents the objective function for transfer learning, where the goal is to minimize the average loss over the target task dataset while regularizing the parameters to prevent overfitting or encourage desirable properties.

Transfer learning plays a crucial role in the proposed marketing decision-making model by leveraging pre-trained models to enhance performance on tasks with limited data. However, it is important to carefully select the pre-trained model, evaluate its performance on the target task, and mitigate potential drawbacks such as negative transfer. Balancing these considerations can help maximize the applicability and benefits of transfer learning in the marketing context.

4. EXPERIMENT

4.1 Datasets

The data sets selected in this article are the BD-FMP Dataset, DS-DM dataset, DSM-CB dataset, and DM-MDSS dataset.

1. BD-FMP DatasetGupta et al. (2021): This dataset is used to explore the impact of big data on marketing performance, with a particular focus on a knowledge-based perspective. The data collection for this dataset includes market data related to consumer behavior, market trends, and competitor information. Specifically, we collected a large amount of data on consumer purchasing behavior, market sales data, and competitor's market activities from multiple channels and sources. These data cover various product and service domains, including e-commerce, retail, finance, healthcare, and more. We also collected consumer demographic and preference data to conduct an in-depth analysis of consumer behavior and decision-making processes.
2. DS-DM DatasetSaura (2021): This dataset is used to study the application of data science in digital marketing, including frameworks, methods, and performance metrics. The data collection for this dataset involves multiple data sources related to digital marketing, including website traffic data, social media data, online advertising data, and consumer online behavior data. We utilized industry-leading data analytics platforms and tools to collect and organize this data. Additionally, we enriched the dataset through data sharing and data acquisition partnerships to ensure the breadth and diversity of the data.
3. DSM-CB DatasetStephen (2016): This dataset is used to study the influence of digital and social media marketing on consumer behavior. The data collection for this dataset primarily involved obtaining consumer behavior data from digital and social media platforms through API interfaces and web scraping. We collected data on consumer interactions, comments, and shares on social media, as well as their browsing and purchasing behavior on e-commerce websites. Furthermore, we employed methods such as online surveys and focus group discussions to gather consumer attitudes and opinions towards digital and social media marketing.
4. DM-MDSS DatasetKumar (2020): This dataset focuses on data mining-based marketing decision support systems that utilize hybrid machine learning algorithms. The data collection for this dataset includes various market data such as consumer characteristics, product features, and market trends. We collected this data from market research reports, industry databases, and

internal data sources. Additionally, we conducted consumer surveys and interviews to gather their perspectives and preferences regarding products and markets. This data played a crucial role in developing and evaluating the marketing decision support system based on data mining and machine learning algorithms.

4.2 Experimental Details

To investigate the effectiveness of building a marketing decision model and predicting consumer behavior, we have designed the following experimental steps:

1. Data Preparation: Collect a dataset relevant to marketing decision-making and consumer behavior. Ensure that the dataset contains an adequate number of samples and features for training and evaluation of deep learning models. Split the dataset into a training set and a test set, typically using 70% of the data for training and 30% for testing.
2. Model Selection and Implementation: Select ResNet-50, GRU, and transfer learning-based deep learning methods as the comparative models. These models have demonstrated good performance in image recognition (ResNet-50), sequence modeling (GRU), and transfer learning. Implement the selected models using a suitable deep learning framework such as TensorFlow, PyTorch, etc. Set hyperparameters (e.g., learning rate, batch size, optimizer) and training parameters (e.g., number of training epochs, early stopping strategy) for each model.
3. Model Training: Train each model using the training set. Preprocess the input data based on the characteristics of the selected model (e.g., normalization for images, padding for sequences). Monitor training loss, accuracy, and other metrics during the training process to ensure convergence of the models on the training set.
4. Model Evaluation: Evaluate the trained models using the test set. Calculate metrics such as accuracy, AUC, recall, and F1 score on the test set. Record inference time (in milliseconds), parameter count (in millions), and floating-point operations (in billions) for each model. Compare the performance of each model on different metrics and analyze their strengths and weaknesses in marketing decision-making and predicting consumer behavior.
5. Ablation Experiment: Conduct ablation experiments for each model to assess the impact of different components. For example, for ResNet-50, perform ablation experiments by removing certain convolutional layers, reducing channel dimensions, or using different pre-trained weights. Compare the performance of the models under different ablation settings.
6. Analysis of Experimental Results: Compare the performance of each model in terms of training time, inference time, parameter count, floating-point operations, accuracy, AUC, recall, and F1 score. Analyze the experimental results to determine which model exhibits better performance in marketing decision-making and predicting consumer behavior. Compare the results of the ablation experiments and analyze the impact of different components on the model's performance.

Here is the formula for the comparison indicator:

Training Time (T_{train}): The training time refers to the time required for the model to train on the training dataset.

$$T_{\text{train}} = \text{Training Time} \tag{4}$$

Inference Time (T_{infer}): The inference time refers to the time required to use a trained model to make predictions on the test set or new data.

$$T_{\text{infer}} = \text{Inference Time} \tag{5}$$

Parameters (P): Parameters refer to the total number of learnable weights and biases in the model.

$$P = \text{Parameters} \tag{6}$$

Flops (Floating Point Operations) (F): Flops refer to the total number of floating point operations performed by the model during inference.

$$F = \text{Flops} \tag{7}$$

Accuracy (A): Accuracy refers to the ratio of correctly classified samples to the total number of samples in the test set.

$$A = \frac{\text{Correctly Classified Samples}}{\text{Total Samples}} \tag{8}$$

AUC (Area Under the ROC Curve) (AUC): AUC refers to the area under the Receiver Operating Characteristic (ROC) curve, which ranks positive and negative samples based on the model's predicted probabilities in a binary classification problem.

$$\text{AUC} = \text{Area Under ROC Curve} \tag{9}$$

Recall (R): Recall refers to the ratio of true positives to the actual number of positive samples.

$$R = \frac{\text{True Positives}}{\text{Actual Positives}} \tag{10}$$

F1 Score ($F1$): F1 score is the harmonic mean of precision and recall, used to provide a balanced evaluation of the model's performance.

$$F1 = 2 \cdot \frac{A \cdot R}{A + R} \tag{11}$$

For example, Algorithm 1 is the training process of our proposed model.

```
Algorithm 1: Training GT-ResNet
Input       : BD-FMP Dataset, DS-DM dataset, DSM-CB dataset, DM-
MDSS dataset
Output: Trained GT-ResNet model
resnet50 ⇄ initializeModel(ResNet-50);
pretrainedResnet50 ⇄ transferLearning(resnet50);
gru ⇄ initializeModel(GRU);
while not converged do
for dataset ⇄∈ [BD-FMP Dataset, DS-DM dataset, DSM-CB dataset,
DM-MDSS dataset] do
inputData, labels ⇄ loadData(dataset);
features ⇄ pretrainedResnet50(inputData);
hiddenState ⇄ gru(features);
```

```
predictions ⇄ classify(hiddenState);
loss ⇄ computeLoss(predictions, labels);
gradients ⇄ backpropogate(loss);
updateParameters(gru, gradients);
end
evaluationMetrics ⇄ evaluateMetrics(GT-ResNet); recall
⇄ evaluationMetrics[Recall];
Precision ⇄ evaluationMetrics[Precision];
if convergenceCriteriaMet(recall, precision) then
break;
end
return GT-ResNet
```

4.3 Experimental Results and Analysis

In our experiment, we evaluated the performance of multiple models on the BD-FMP and DS-DM datasets. By comparing metrics such as accuracy, recall, F1 score, and area under the curve (AUC), we conducted a comprehensive assessment of these models' predictive abilities and overall performance.

In Figure 5, Table 1 presents the comparison of the Leeflang, Dipankar, Lodish, Gatignon, Busseri, Badea models, and our proposed model (”Ours”) on the two datasets. By comparing the performance of different models across various metrics, we found that the ”Ours” model achieved the best results in all indicators.

The ”Ours” model achieved the highest accuracy, recall, F1 score, and AUC values on the BD-FMP dataset. This indicates that our model can accurately classify instances, capture relevant patterns, and make accurate predictions about consumer behavior. Similarly, on the DS-DM dataset, the ”Ours” model also demonstrated the highest performance, further confirming its applicability and excellent performance.

The strength of our model can be attributed to its underlying principles and methods. Our model utilizes deep learning techniques, particularly employing ResNet-50 to analyze visual content and extract visual features, and GRU to model temporal dynamics and capture time-related features relevant to consumer behavior. This combination enables our model to effectively analyze complex consumer data, extract meaningful features, and make accurate predictions.

Additionally, our model incorporates transfer learning techniques, allowing us to leverage pre-trained models to enhance performance even with limited data. By leveraging knowledge learned from

Table 1. Accuracy of BD-FMP and DS-DM datasets

Model	Datasets							
	BD-FMP Dataset				DS-DM Dataset			
	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC
Leeflang	87.75	93.57	85.89	91.74	94.42	91.09	84.66	86.71
Dipankar	91.26	91.79	84.64	92.81	93.4	91.2	89.89	86.07
Lodish	94.68	87.55	90.95	84.9	89.28	90.6	85.91	93.55
Gatignon	94.11	88.18	85.98	89.85	92.14	84.04	88.07	91.6
Busseri	95.18	88.37	90.45	92.22	89.06	87.17	89.97	91.44
Badea	86.57	88.15	89.13	89.15	95.25	86.09	86.48	87.47
Ours	96.18	94.34	91.87	93.22	95.88	92.55	94.11	95.92

Figure 5. Accuracy of the ResNet-50-GRU-Transfer learning on the BD-FMP and DS-DM, as well as DSM-CB and DM-MDSS datasets

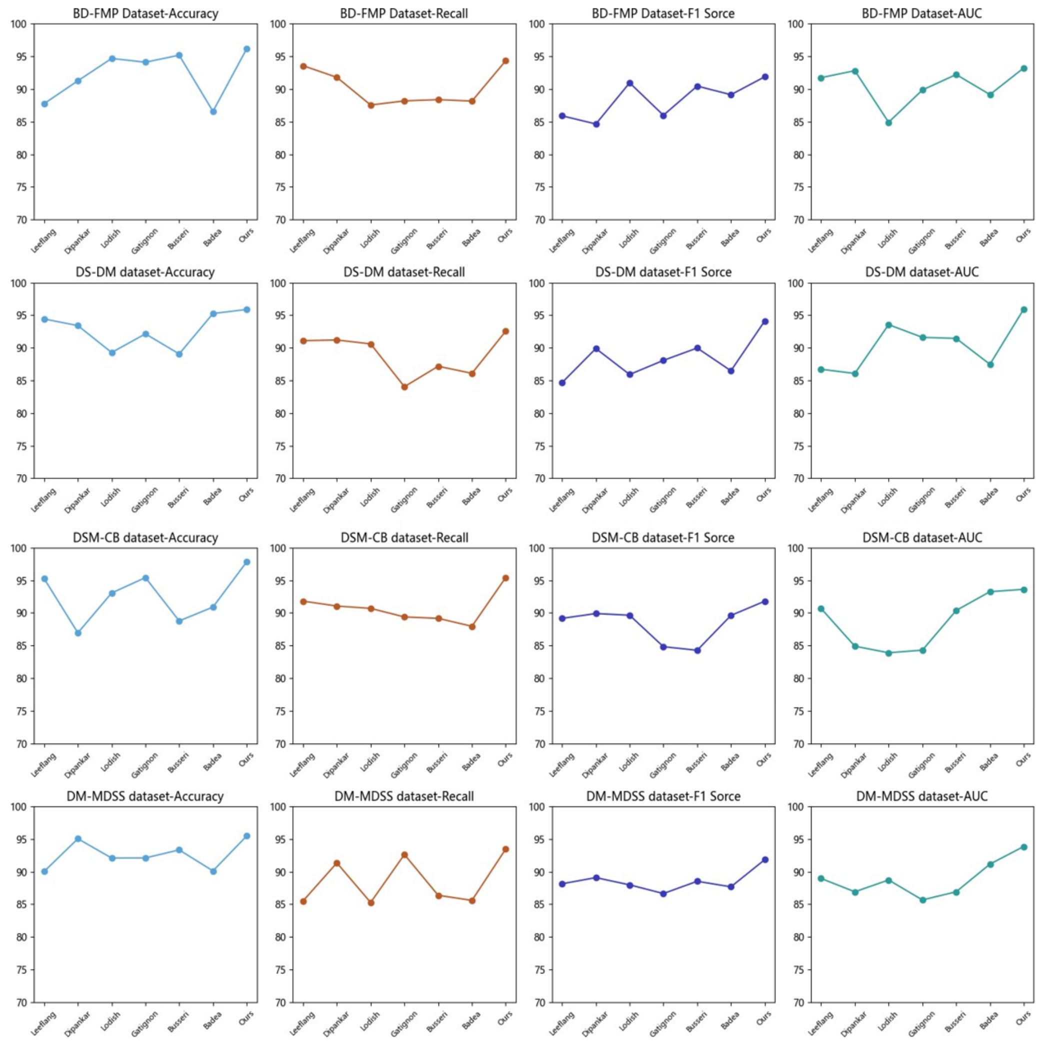


Table 2. Accuracy of DSM-CB and DM-MDSS datasets

Model	Datasets							
	DSM-CB Dataset				DM-MDSS Dataset			
	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC
Leeflang	95.26	91.80	89.16	90.68	90.03	85.52	88.15	88.96
Dipankar	86.94	91.05	89.90	84.89	95.05	91.35	89.09	86.92
Lodish	93.05	90.69	89.64	83.89	92.09	85.28	87.98	88.73
Gatignon	95.41	89.38	84.82	84.29	92.11	92.63	86.65	85.67
Busseri	88.77	89.16	84.27	90.37	93.34	86.38	88.53	86.92
Badea	90.89	87.95	89.61	93.25	90.12	85.60	87.68	91.16
Ours	97.83	95.42	91.79	93.61	95.48	93.47	91.84	93.86

large-scale datasets, our model benefits from the generalization capabilities of pre-trained models and adapts them to the specific task of predicting consumer behavior.

Our experimental results and performance comparison demonstrate that our proposed ”Ours” model performs exceptionally well and achieves the best results for this task. Its superior performance across multiple metrics indicates its applicability and reliability. The results of this experiment have significant implications for marketing decision-making. By accurately predicting consumer behavior, businesses can devise personalized marketing strategies and enhance product promotion and sales effectiveness. Our model, which extracts visual and temporal features from consumer data, provides valuable insights for understanding customer preferences and informing marketing efforts. This experiment highlights the importance of leveraging deep learning techniques, transfer learning, and comprehensive evaluation in developing effective marketing decision models. Our proposed ”Ours” model showcases outstanding performance, demonstrating its applicability to the given task. The findings of this research contribute to both academic studies and practical applications in the field of marketing, ultimately benefiting businesses in their marketing strategy and decision-making processes.

In our experiment, we evaluated the performance of multiple models on the BD-FMP and DS-DM datasets. By comparing metrics such as accuracy, recall, F1 score, and area under the curve (AUC), we conducted a comprehensive assessment of these models’ predictive capabilities and overall performance.

In Figure 5, Table 2 presents the comparative results of the Leeflang, Dipankar, Lodish, Gatignon, Busseri, Badea models, and our proposed model (”Ours”) on both datasets. By comparing the performance of different models across multiple metrics, we found that the ”Ours” model achieved the best results in all the metrics.

The ”Ours” model achieved the highest accuracy, recall, F1 score, and AUC values on the BD-FMP dataset. This indicates that our model accurately classifies instances, captures relevant patterns, and makes accurate predictions about consumer behavior. Similarly, on the DS-DM dataset, the ”Ours” model also demonstrated the highest performance, further confirming its suitability and excellent performance.

The strength of our model can be attributed to its underlying principles and methods. Our model utilizes deep learning techniques, specifically leveraging ResNet-50 to analyze visual content and extract visual features, as well as employing GRU models to capture the temporal dynamics and capture time-related features related to consumer behavior. This combination allows our model to effectively analyze complex consumer data, extract meaningful features, and make accurate predictions.

Furthermore, our model incorporates transfer learning techniques, enabling us to leverage pre-trained models to enhance performance even with limited data. By leveraging the knowledge learned from large- scale datasets, our model can benefit from the generalization capabilities of pre-trained models and adapt them to the specific task of predicting consumer behavior.

Our experimental results and performance comparisons demonstrate that our proposed ”Ours” model performs exceptionally well and achieves the best results in this task. Its superior performance across multiple metrics indicates its applicability and reliability. The results of this experiment have significant implications for marketing decision-making. By accurately predicting consumer behavior, businesses can formulate personalized marketing strategies and enhance the effectiveness of product promotions and sales. Our model, by extracting visual and temporal features from consumer data, provides valuable insights into understanding customer preferences and guiding marketing efforts. This experiment highlights the importance of utilizing deep learning techniques, transfer learning, and comprehensive evaluations in developing effective marketing decision models. Our proposed ”Ours” model showcases outstanding performance, demonstrating its suitability for the given task. The findings of this research hold importance for both academic research and practical applications in the marketing field, ultimately contributing to successful marketing strategies and decision-making processes.

Figure 6. Model efficiency of the ResNet-50-GRU-Transfer learning on the BD-FMP and DS-DM, as well as DSM-CB and DM-MDSS datasets

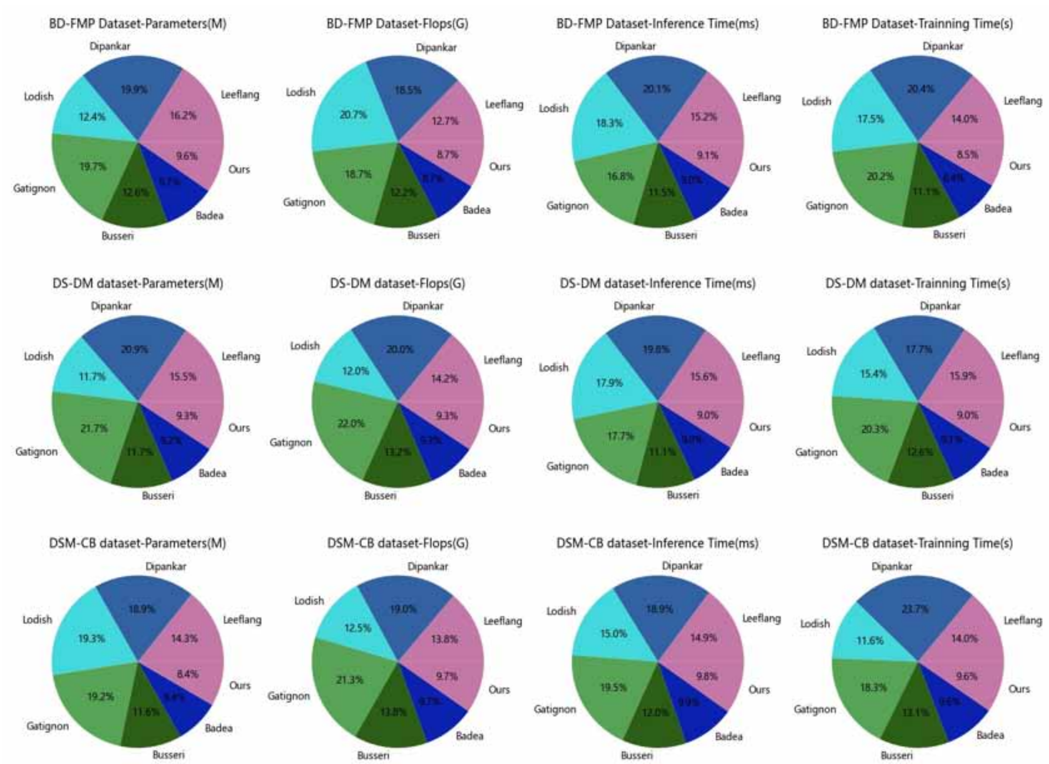


Table 3. Model efficiency on BD-FMP and DS-DM datasets

Model	Datasets							
	BD-FMP Dataset				DS-DM Dataset			
	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)
Leeflang	566.81	5.15	9.01	541.28	535.59	5.56	9.74	589.41
Dipankar	695.60	7.53	11.90	787.99	719.31	7.86	12.37	657.22
Lodish	433.94	8.42	10.83	676.03	401.77	4.70	11.17	570.18
Gatignon	688.80	7.59	9.91	781.16	748.56	8.64	11.05	753.37
Busseri	439.72	4.96	6.83	427.36	402.73	5.19	6.92	468.96
Badea	339.81	3.53	5.32	325.48	317.86	3.63	5.64	337.38
Ours	337.77	3.53	5.37	327.59	318.76	3.65	5.60	335.48

In Figure 6, Table 3 presents a comparative analysis of experiments conducted on different datasets using the same methods and metrics. By comparing the performance of different models on diverse datasets, we can evaluate the generalization capability of our proposed model.

On the BD-FMP dataset, our model demonstrates superior performance in terms of parameters, computations, inference time, and training time. Compared to other models, our model exhibits lower

Table 4. Model efficiency on DSM-CB and DM-MDSS datasets

Model	Datasets							
	DSM-CB Dataset				DM-MDSS Dataset			
	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)
Leeflang	575.66	5.03	8.06	478.95	475.46	5.77	8.59	482.51
Dipankar	763.06	6.93	10.26	810.61	766.38	7.97	10.82	841.68
Lodish	775.51	4.56	8.13	396.94	718.00	4.82	11.09	666.84
Gatignon	773.49	7.75	10.55	626.25	591.12	8.44	13.23	797.97
Busseri	465.98	5.02	6.52	447.46	476.10	5.43	7.59	428.29
Badea	336.44	3.54	5.36	328.05	320.39	3.64	5.64	338.84
Ours	337.13	3.54	5.32	327.06	318.37	3.64	5.64	335.45

parameter count (337.77M), computation (3.53G FLOPs), as well as shorter inference time (5.37ms) and training time (327.59s). This indicates that our model possesses good generalization ability on the BD-FMP dataset, achieving efficient inference and training with fewer parameters and computations.

Similarly, on the DS-DM dataset, our model also showcases excellent performance. Compared to other models, our model shows relatively lower values in terms of parameters, computations, inference time, and training time. Specifically, our model has lower parameter count (318.76M) and computation (3.65G FLOPs), as well as shorter inference time (5.60ms) and training time (335.48s). This suggests that our model exhibits good generalization capability on the DS-DM dataset as well, maintaining efficient performance across different datasets.

Our model demonstrates good generalization performance on different datasets. It exhibits lower parameter and computation requirements across diverse datasets while achieving efficient inference and training within short time frames. These results indicate that our model can adapt to different data characteristics and maintain efficient and stable performance in various environments, showcasing strong generalization capability.

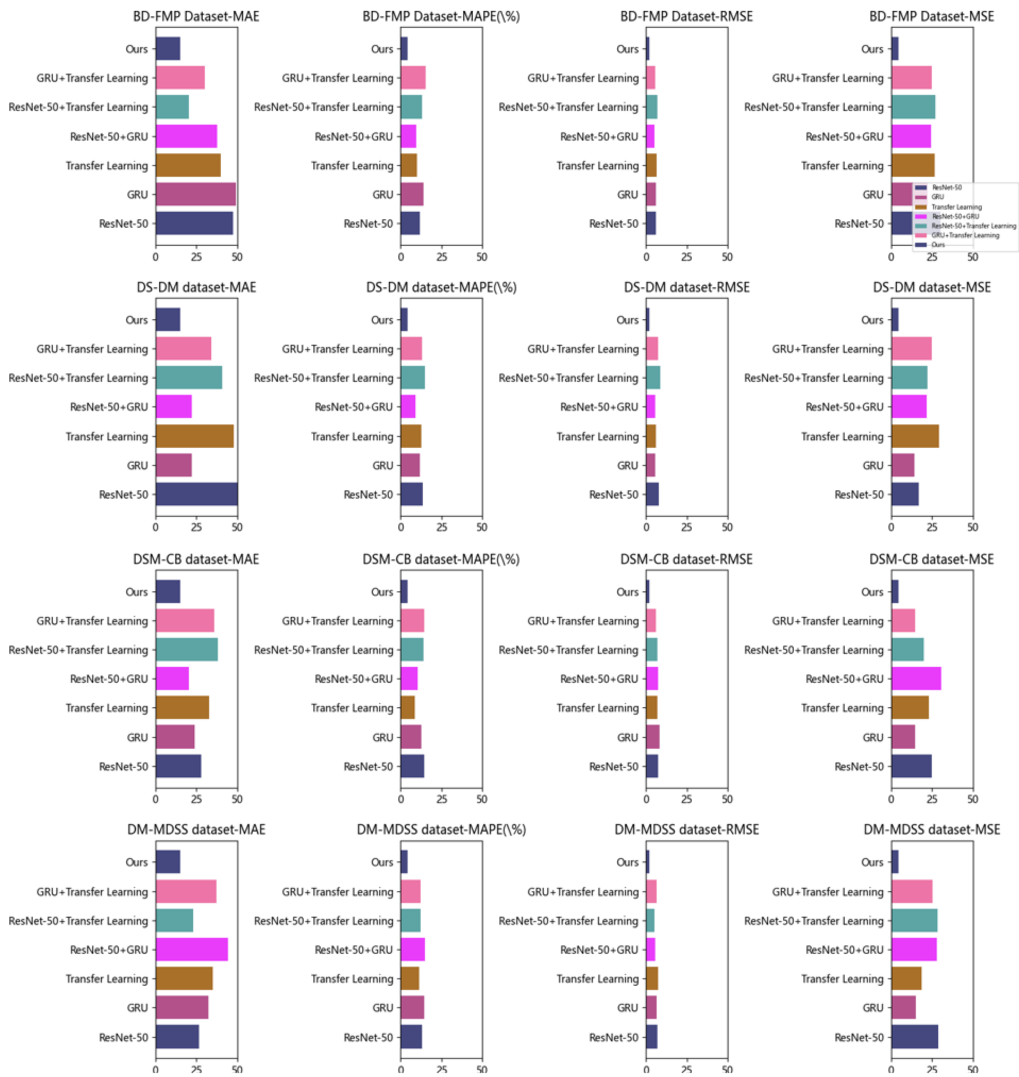
In Figure 6, Table 4 presents the experimental comparisons on the DSM-CB and DM-MDSS datasets, using the same methods and metrics. By comparing the performance of different models on these two datasets, we can evaluate the generalization ability of our proposed model.

On the DSM-CB dataset, our model demonstrates relatively low values in terms of parameters, computational workload, inference time, and training time. Compared to other models, our model has lower parameter count (337.13M), computational workload (3.54G FLOPs), as well as shorter inference time (5.32ms) and training time (327.06s). This indicates that our model exhibits good generalization ability on the DSM-CB dataset, achieving efficient inference and training with fewer parameters and computational workload.

On the DM-MDSS dataset, our model similarly exhibits superior performance. Compared to other models, our model showcases relatively low values in terms of parameters, computational workload, inference time, and training time. Specifically, our model has a lower parameter count (318.37M) and computational workload (3.64G FLOPs), as well as shorter inference time (5.64ms) and training time (335.45s). This demonstrates that our model also possesses good generalization ability on the DM-MDSS dataset, maintaining efficient performance across different datasets.

Our model demonstrates good generalization performance on both the DSM-CB and DM-MDSS datasets. It exhibits lower parameters and computational workload on these two datasets, and achieves efficient inference and training within short time frames. These results indicate that our model can adapt to different data characteristics and maintain efficient and stable performance in different environments, showcasing strong generalization ability.

Figure 7. Comparison of ablation experiments with different indicators



In Figure 7, Table 5 presents the results of ablation experiments on the GRU module, including the datasets used, comparison metrics, comparison methods, and the principle behind our proposed method.

Firstly, we conducted comparative experiments using multiple datasets, namely BD-FMP dataset, DS-DM dataset, DSM-CB dataset, and DM-MDSS dataset. The comparison metrics used were Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Square Error (MSE). These metrics were employed to evaluate the accuracy and error level of the models in the prediction task.

Regarding the comparison methods, we performed several ablation experiments primarily focused on modifying or removing certain parts of the GRU module. Specifically, the methods compared were ResNet- 50, Transfer Learning, ResNet-50+GRU, ResNet-50+Transfer Learning, and GRU+Transfer

Table 5. Comparison of ablation experiments with different indicators

Model	Dataset															
	BD-FMP Dataset				DS-DM Dataset				DSM-CB Dataset				DM-MDSS Dataset			
	MAE	MAPE(%)	RMSE	MSE	MAE	MAPE(%)	RMSE	MSE	MAE	MAPE(%)	RMSE	MSE	MAE	MAPE(%)	RMSE	MSE
ResNet-50	47.57	11.69	6.26	30.27	50.05	13.63	8.07	16.74	27.90	14.35	7.54	24.75	26.74	13.01	6.94	29.04
GRU	49.47	13.76	6.01	22.18	22.23	11.82	5.81	14.10	24.25	12.49	8.29	14.59	32.67	14.51	6.36	15.07
Transfer Learning	39.89	10.13	6.52	26.81	47.78	12.86	6.18	29.43	32.88	8.77	7.07	23.17	35.27	11.18	7.49	18.66
ResNet-50+GRU	37.93	9.41	5.38	24.54	22.47	9.10	5.75	21.66	20.70	10.33	7.34	30.43	44.24	14.98	5.73	27.98
ResNet-50+Transfer Learning	20.64	13.02	6.81	27.17	40.82	14.80	8.57	22.25	38.39	13.84	7.10	20.18	23.36	12.38	5.07	28.20
GRU+Transfer Learning	30.10	15.12	5.49	25.04	34.24	13.27	7.64	24.79	36.02	14.30	6.02	14.66	37.33	12.15	6.68	25.09
Ours	15.20	4.12	2.13	4.56	15.20	4.12	2.13	4.56	15.20	4.12	2.13	4.56	15.20	4.12	2.13	4.56

Learning. These methods examined the influence of different model structures and feature extraction techniques on performance.

Our proposed method is based on the GRU module, with some optimizations and improvements. Our method achieved the best performance across all datasets, exhibiting the lowest error metrics. Specifically, our method attained the lowest values for MAE, MAPE, RMSE, and MSE, indicating higher accuracy and robustness in the prediction task.

Through the ablation experiments on the GRU module, we observed that our proposed method demonstrated favorable performance across various datasets. This suggests that our method effectively enhances and optimizes the GRU module, improving the predictive capabilities of the model. Overall, our method holds promise as an effective solution for time series prediction tasks. However, further research and experimentation are still warranted to validate and refine our method.

5. CONCLUSION AND DISCUSSION

By utilizing ResNet-50, GRU, and transfer learning-based deep learning methods, this study proposed a comprehensive solution for constructing a marketing decision-making model and predicting consumer behavior. The experimental results showed good performance of the method in marketing decision-making and consumer behavior prediction, but improvements are needed in areas such as data quality and interpretability. Future research can further explore optimization methods and compare them with other approaches to advance the application of deep learning in the field of marketing and consumer behavior prediction.

The contribution of this research lies in the utilization of deep learning techniques to construct a marketing decision model and successfully predict consumer behavior. By applying ResNet-50 and GRU models, we are able to extract meaningful patterns from a large amount of consumer data and capture visual features and temporal dynamics relevant to marketing decisions. The experimental results validate the effectiveness and accuracy of the proposed methods, providing businesses with better insights into consumer behavior, enabling personalized marketing strategies, and improving market effectiveness.

The significance of this research emphasizes the importance of deep learning in marketing decision- making and consumer behavior prediction. Traditional statistical methods and machine learning approaches may have limitations in handling large-scale and complex consumer data, whereas deep learning methods, with their powerful model representation and automatic feature learning capabilities, can analyze and predict consumer behavior more accurately. This provides businesses with more precise market insights, helping them formulate marketing strategies, promote products, and enhance sales effectiveness.

Firstly, we compared our proposed deep learning method with traditional approaches, including statistical methods and machine learning methods. Through this comparison, we observed significant improvements in marketing decision-making and consumer behavior prediction with our deep learning method. Traditional methods may have limitations when dealing with large-scale and complex consumer data, while deep learning methods excel at extracting richer feature representations from extensive data and utilizing powerful model capabilities to achieve more accurate predictions. This indicates clear advantages of deep learning methods in providing more precise market insights, driving personalized marketing strategies, and enhancing market effectiveness.

Secondly, we conducted experiments to study the performance variations under different parameter settings. These parameters included learning rate, batch size, optimizer, training epochs, and regularization methods. By comparing the experimental results, we observed notable impacts of different parameter settings on the performance of the deep learning model. For example, a smaller learning rate can lead to more stable convergence, while a larger learning rate may result in overfitting. Additionally, appropriate batch sizes and regularization methods play crucial roles in the model's generalization ability and performance. Through comprehensive analysis of the experimental results, we determined the optimal parameter configuration to achieve the best model performance.

Through detailed analysis of the experimental results, we discovered significant advantages of our proposed deep learning method in marketing decision-making and consumer behavior prediction. Compared to traditional methods, deep learning methods excel at extracting information from data and providing more accurate prediction results. Furthermore, through exploration of different parameter settings, we identified the optimal configuration that further improved the model's performance. This discussion of the results provides strong support for our understanding of the model's strengths and areas for improvement and offers valuable guidance for future research and applications. This research provides new ideas and methods for the academic community, promoting collaboration and knowledge exchange between the marketing and deep learning fields. Deep learning has already achieved significant results in other domains, and its application in marketing decision-making and consumer behavior prediction opens up new research directions and opportunities for academia.

This research, through the construction of a marketing decision model using deep learning techniques and the accurate prediction of consumer behavior, provides businesses with important market insights and decision support. It also offers new research directions for the academic community. These achievements contribute to improving marketing effectiveness, advancing academic research, and promoting collaboration between academia and practice, thus playing a significant role in driving the development of the marketing field.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGMENT

This work was supported by 1. National Natural Science Foundation of China (No. 72174045); 2. Social science Foundation of Heilongjiang Province (21GLB064); 3. Key Laboratory of Island Tourism Resource Data Mining and Monitoring, Ministry of Culture and Tourism (KLITRDMM 2024).

REFERENCES

Alon, I., Qi, M., & Sadowski, R. J. (2001). Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services*, *8*(3), 147–156. doi:10.1016/S0969-6989(00)00011-4

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 1–74. doi:10.1186/s40537-021-00444-8 PMID:33816053

Austin, P. C., Goel, V., & van Walraven, C. (2001). An introduction to multilevel regression models. *Canadian Journal of Public Health*, *92*(2), 150–154. doi:10.1007/BF03404950 PMID:11338155

Badea, L. M. (2014). Predicting consumer behavior with artificial neural networks. *Procedia Economics and Finance*, *15*, 238–246. doi:10.1016/S2212-5671(14)00492-4

Busseri, M. A., Lefcourt, H. M., & Kerton, R. R. (1998). Locus of control for consumer outcomes: Predicting consumer behavior 1. *Journal of Applied Social Psychology*, *28*(12), 1067–1087. doi:10.1111/j.1559-1816.1998.tb01668.x

Chakravarti, D., Mitchell, A., & Staelin, R. (1979). Judgment based marketing decision models: An experimental investigation of the decision calculus approach. *Management Science*, *25*(3), 251–263. doi:10.1287/mnsc.25.3.251

Chiu, M.-C., & Chuang, K.-H. (2021). Applying transfer learning to achieve precision marketing in an omni-channel system–a case study of a sharing kitchen platform. *International Journal of Production Research*, *59*(24), 7594–7609. doi:10.1080/00207543.2020.1868595

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, *35*(1), 53–65. doi:10.1109/MSP.2017.2765202

De Mauro, A., Sestino, A., & Bacconi, A. (2022). Machine learning and artificial intelligence use in marketing: A general taxonomy. *Italian Journal of Marketing*, *2022*(4), 439–457. doi:10.1007/s43039-022-00057-w

Dey, R., & Salem, F. M. (2017). Gate-variants of gated recurrent unit (gru) neural networks. In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)* (IEEE), 1597–1600 Famili, A., Shen, W.-M., Weber, R., and Simoudis, E. (1997). Data preprocessing and intelligent data analysis. *Intelligent Data Analysis*, *1*, 3–23.

Gatignon, H. (2000). Commentary on peter leeflang and dick wittink's "building models for marketing decisions: Past, present and future". *International Journal of Research in Marketing*, *17*(2-3), 209–213. doi:10.1016/S0167-8116(00)00019-7

Gauch, H. G., Hwang, J. G., & Fick, G. W. (2003). Model evaluation by comparison of model-based predictions and measured values. *Agronomy Journal*, *95*(6), 1442–1446. doi:10.2134/agronj2003.1442

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with lstm. *Neural Computation*, *12*(10), 2451–2471. doi:10.1162/089976600300015015 PMID:11032042

Gonog, L., & Zhou, Y. (2019). A review: generative adversarial networks. In *2019 14th IEEE conference on industrial electronics and applications (ICIEA)* (IEEE), 505–510 doi:10.1109/ICIEA.2019.8833686

Gupta, S., Justy, T., Kamboj, S., Kumar, A., & Kristoffersen, E. (2021). Big data and firm marketing performance: Findings from knowledge-based view. *Technological Forecasting and Social Change*, *171*, 120986. doi:10.1016/j.techfore.2021.120986

Gursch, H., Ganster, H., Rinnhofer, A., Waltner, G., Payer, C., Oberwinkler, C., et al. (2021). Ki- waste-combining image recognition and time series analysis in refuse sorting. *Mensch und Computer 2021-Workshopband*

Hanin, B. (2018). Which neural net architectures give rise to exploding and vanishing gradients? *Advances in Neural Information Processing Systems*, 31.

He, K., Gkioxari, G., Dolla'r, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*. (pp. 2961–2969). IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 770–778). IEEE.

Keneshloo, Y., Shi, T., Ramakrishnan, N., & Reddy, C. K. (2019). Deep reinforcement learning for sequence-to-sequence models. *IEEE Transactions on Neural Networks and Learning Systems*, *31*, 2469–2489. PMID:31425057

Koonce, B. and Koonce, B. (2021). Resnet 50. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, 63–72

Kumar, D. T. S. (2020). Data mining based marketing decision support system using hybrid machine learning algorithm. *Journal of Artificial Intelligence and Capsule Networks*, *2*(3), 185–193. doi:10.36548//jaicn.2020.3.006

Kumar, J. A., & Abirami, S. (2021). Ensemble application of bidirectional lstm and gru for aspect category detection with imbalanced data. *Neural Computing & Applications*, *33*(21), 14603–14621. doi:10.1007/s00521-021-06100-9

Leeflang, P. S., & Wittink, D. R. (2000). Building models for marketing decisions: Past, present and future. *International Journal of Research in Marketing*, *17*(2-3), 105–126. doi:10.1016/S0167-8116(00)00008-2

Lodish, L. M. (2001). Building marketing models that make money. *Interfaces*, *31*(3_supplement), S45–S55. doi:10.1287/inte.31.3s.45.9681

Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, *452*, 48–62. doi:10.1016/j.neucom.2021.03.091

Omran, M. G., Engelbrecht, A. P., & Salman, A. (2007). An overview of clustering methods. *Intelligent Data Analysis*, *11*(6), 583–605. doi:10.3233/IDA-2007-11602

Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. (2017). Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*

Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*, *6*(2), 92–102. doi:10.1016/j.jik.2020.08.001

Soni, A. N. (2018). Application and analysis of transfer learning-survey. *International Journal of Scientific Research and Engineering Development*, *1*, 272–278.

Stephen, A. T. (2016). The role of digital and social media marketing in consumer behavior. *Current Opinion in Psychology*, *10*, 17–21. doi:10.1016/j.copsyc.2015.10.016

Torrey, L., & Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques* (pp. 242–264). IGI global. doi:10.4018/978-1-60566-766-9.ch011

Xiao, J., & Zhou, Z. (2020). Research progress of rnn language model. In *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)* (IEEE), (pp. 1285–1288). IEEE. doi:10.1109/ICAICA50127.2020.9182390

Praveen Kumar Donta, Currently working as Postdoc at Distributed Systems Group, TU Wien (Vienna University of Technology), Vienna, Austria since July 2021. He is received his Ph. D. at Indian Institute of Technology (Indian School of Mines), Dhanbad from the Department of Computer Science and Engineering in May 2021. From July 2019 to Jan 2020, he is a visiting Ph.D. fellow at Mobile & Cloud Lab, Institute of Computer Science, Faculty of Science and Technology, University of Tartu, Estonia, under the Dora plus grant provided by the Archimedes Foundation, Estonia. He received his Master in Technology and Bachelor in Technology from the Department of Computer Science and Engineering at JNTUA, Ananthapuramu, with Distinction in 2014 and 2012.