Predicting the Next Transaction on Anonymized Payment Datasets with Deep Learning Models

Claudia Francesca Suarez Mariscal, Renata Galante, Weverton Cordeiro

[1]Institute of Informatics (INF) Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

{cfsmarisal, galante, weverton.cordeiro}@inf.ufrgs.br

Abstract. *Predicting customer behavior has long been a critical area of exploration for many companies, who often analyze purchase history to uncover behavioral trends and enhance their services. However, analyzing large amounts of personal customer data while maintaining compliance with data protection regulations (GDPR or LGPD) is challenging. In this paper, we propose three models that tackle the complexities of recognizing purchasing patterns for diverse applications in anonymized data. First, we evaluate architectures leveraging DL models for predicting subsequent purchase transactions using a dataset that safeguards confidential customer data while adhering to data protection regulations. The suggested models rely on Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) to discern behaviors within a dataset devoid of personal information, allowing for comparison with other models pursuing the same goal. Then, we optimize each model's parameters, with findings indicating that the GRU-based model demonstrates superior generalization capabilities.*

1. Introduction

The analysis of one's personal data (e.g., items visited, movies viewed, musics listened, products purchased) is pivotal for recommendation systems in many online platforms [Nery et al. 2021, Suarez Mariscal et al. 2023]. In the context of e-commerce, purchase history uncovers customer behavior, highlights trending products, and supports customer segmentation useful for strategic and informed decision-making. With the consolidation of regulations devised to protect users' personal data, the challenge lies in effectively using data science to derive insights into customer preferences and buying patterns, while prioritizing personal data protection [Wachter et al. 2017]. In Europe, the General Data Protection Regulation (GDPR) advances privacy initiatives to foster user trust in data sharing. Similarly, Brazil's General Data Protection Law (LGPD) aims to enhance data protection by establishing legal security and standardizing regulations in line with international benchmarks [Pinheiro 2020]. Following these guidelines helps ensure data privacy and regulate its use [Wieringa et al. 2021].

The main task of analyzing customer behavior is to infer patterns in the data. For example, purchasing behaviors are often correlated, and acquiring particular items may meet specific scenarios. This analysis must be carried out without requiring access to strictly personal information (such as geographic coordinates that allow user tracking). Consequently, models should be able to run any analysis without accessing this type of data [Vasupula 2022]. In summary, data scientists should not only focus on the results but also set the boundaries of their analysis [Martens 2022].

Proposals-based probabilistic models [Yuan et al. 2017, Wen et al. 2018] can extract relationships between observations to calculate the likelihood of specific events occurring. However, these models need frequent updates to reflect changing probabilities. Additionally, other studies leveraging Machine Learning (ML) [Fleder and Shah 2020, Tabianan 2022, Wang et al. 2023] are trained and tested on historical data, enabling automatic learning and recognition of behavioral patterns [Li et al. 2019]. They can also identify crucial features and comprehend their influence on customer behavior. Nevertheless, due to the variability of purchasing behavior, a more in-depth analysis requires a thorough examination considering past actions. Some research employing Deep Learning (DL) [Huang et al. 2019, Zhu et al. 2020, Zhou et al. 2018, Sarkar and De Bruyn 2021] has shown promising outcomes in analyzing sequence-based data, capturing temporal dependencies in customer behavior.

This paper proposes creating three models to handle the complexities of identifying purchasing patterns. The key concept is using non-private user information to predict transactions through deep learning models. The study aimed to determine if using this dataset could produce the most effective results. The experiments detailed specific configurations for each architecture, highlighting that the LSTM and GRU architectures were suitable for this analysis. Results demonstrated that these architectures are optimal compared to other machine learning models for this type of analysis.

The rest of this paper is structured as follows. Section 2 describes related works to our Models. In Section 3, the dataset is explained. In Section 4, the methodology is presented, while in Section 5, the configuration of each architecture is described. Finally, Section 6 presents the conclusions and future works.

2. Related Work

The literature las long been concerned with privacy-preserving data analysis [Neto et al. 2018, Campagna et al. 2020]. Next, we cover investigations focused on predictive analysis in e-commerce, organizing the discussion into three categories: probabilistic models, Machine Learning, and Deep Learning.

Probabilistic models - An approach focused on basket customization was addressed using matrix factorization and Markov Chains [Rendle et al. 2010]. Meanwhile, PRED incorporates geographic and temporal information through a non-parametric Bayesian model [Yuan et al. 2017]. Another approach suggests including temporal, geographic, payment, and category information to create a probabilistic model. These methods emphasize inferring relationships to determine model parameters, analyzing periodicity with Fast Fourier Transform (FFT) [Wen et al. 2018]. Besides, Shopper [Ruiz et al. 2020], a sequential probabilistic model, links items through latent attributes, representing baskets with item pairs. Meanwhile, PRIMA++, a probabilistic framework, can learn user preferences from limited data by considering attributes with similar tradeoffs and information on competitive items [Li et al. 2021].

Machine Learning (ML) - A low-cost iterative algorithm was proposed as a regression problem to determine the set of purchased items and their quantities based on the payment amount [Fleder and Shah 2020]. Another method involved extracting 274 features from a customer's monthly transactions, which were then processed using Logistic Lasso and Gradient Tree Boosting to predict whether the customer made a purchase or not [Martínez et al. 2020]. To group customers with similar characteristics in small datasets, a smoothing and standardization procedure should be adopted to mitigate the impact of this limitation on the analysis [Tabianan 2022]. Another strategy involves forecasting consumer behavior during the pandemic by incorporating a feature indicating if a purchase is COVID-19-related, utilizing correlation analysis to pinpoint critical factors influencing online buying patterns [Safara 2022]. Additionally, XGBoost, an ensemble model, is used to predict purchasing patterns through an LDTD user value model and, enabling the differentiation of user types based on their account history [Wang et al. 2023].

Deep Learning - To identify potential customers interested in banking products, a model utilizes a dynamic window to detect time-based sequential dependencies among specific transactions utilizing Random Forest and Deep Neural Networks [Ładyżyński et al. 2019]. Also, the recurrent neural networks could establish groups of customers according to similar behaviors, and later, this same network discovered possible dependencies between purchases from different categories [Huang et al. 2019]. Meanwhile, LSTM or GRU can reveal patterns in purchasing behavior, understanding the impact of stationary behaviors or micro-behaviors [Zhou et al. 2018]. An alternative approach employs the LSTM to process raw data without necessitating feature engineering, adjusting hyperparameters only, yielding favorable outcomes when the dataset is complete [Sarkar and De Bruyn 2021]. A hybrid DL model (EE-CNN) merges entity embeddings in a convolutional neural network to predict the next item for purchase, suggesting that the purchase location feature can enhance results [Zhu et al. 2020].

Comparison. While all three model types show promise in predictive analysis for E-commerce, probabilistic models rely heavily on data quantity for probability calculations. On the other hand, ML and DL models require comprehensive datasets for proper training; although more intricate, they yield high-quality results. Additionally, certain proposals [Wen et al. 2018, Huang et al. 2019, Martínez et al. 2020, Sarkar and De Bruyn 2021] emphasize the significance of feature engineering to extract additional insights from base features and enhance analysis. Regarding the datasets used, most proposals utilize confidential datasets, limiting their use for further study. Despite this limitation, safeguarding personal data allows for selecting public datasets to proceed with proposals. Moreover, PRED [Yuan et al. 2017] and STPC-PGM [Wen et al. 2018] are viewed as more straightforward approaches for predicting the next purchase transaction and are thus chosen as baselines. Consequently, baselines are implemented based on information from the articles, and the Online Retail Dataset is also used.

3. Dataset

A public dataset, the Online Retail Dataset, was selected for predictive analysis in E-commerce. This dataset contains all the transactions between 01/12/2010 and 09/12/2011 for an online store based in the United Kingdom that offers unique gifts for all occasions. The Online Retail Dataset, cited as [1], consists of 541,909 instances and 8 attributes. You can find this dataset in the UCI Machine Learning Repository at https://archive.ics.uci.edu/ml/datasets/online+retail.

- *Data Collection and Data Grouping.* The data was collected from a transnational online store in the United Kingdom, but the focus of predicting purchasing behavior needs to rely on users' history. Thus, the dataset was grouped according to the history size per user. Each of these groups is described in Table 1.
- *Data Annotation.* To employ RNN-type architectures, the data needed to be modeled sequentially, where the target of a transaction will be the next transaction. Likewise, for a history of size n ($n > 2$), the target will be the $transaction_n$ after the last $transaction_{n-1}$ from this history.
- *Data Dictionary.* The features description that includes Online Retail Dataset can be found at UCI Machine Learning Repository.

Table 1. Groups established from the online retail dataset

Group	Description
Group A	Comprised of the history of users who have less than 100 transactions.
Group B	Comprised of the history of users who have more than 100 transactions.
Group C	Composed of the purchase history of all customers.

4. Methodology

The first focus of the proposal was forecasting the next purchase transaction. This section presents the methodology developed to build three models based on RNN architecture: (1) a simple RNN architecture; (2) an LSTM architecture; and (3) a GRU architecture. The dataset employed is the Online Retail Dataset, and the methodology presented in Figure 1 contains the fundamental steps of data science: data collection, data pre-processing, data transformation, data exploration, model building, and model evaluation (section 5).

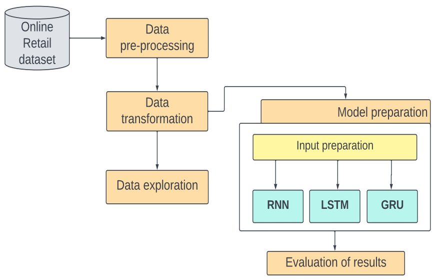


Figure 1. Methodology to predict the next transaction

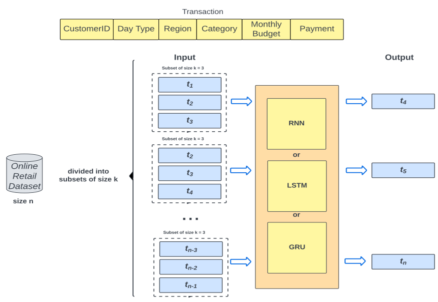


Figure 2. Model Training

4.1. Preprocessing Stages

Data Preprocessing. The dataset had inconsistencies requiring correction before. Hence, the following were excluded: duplicate instances, instances with negative payment values (returns), and instances with missing values in the CustomerID attribute. As a result, the dataset decreased from 541,909 to 392,732 instances. In the *Data Transformation* stage, six features were established based on the attributes presented in section 3), Table 2 contains the description of each of them.

Data Exploration. At this stage, the data was explored and analyzed to discover implicit information associated with the purchase history. Thus, it was possible to identify

Table 2. Description of features achieved after data transformation

Attribute	Description
CustomerID	Maintains the customer identification (ID).
Day type	Determines when purchases were made; if they were made between Monday and Friday, they are considered on the days of the week (0), and if they were made between Saturday and Sunday, they are considered on the weekend (1).
Region	Holds the name of the country in which the purchase is registered.
Category	Corresponds to the purchase category codes. To group them, it was necessary to keep the first four strings of the code from StockCode.
Monthly Budget	Corresponds to the level of consumption per month according to the payment.
Payment	Calculated by multiplying the unit price by the quantity of each product.

some weaknesses that could harm the analysis. Below is a summary of the most relevant aspects identified during data exploration.

- The dataset comprises 4,339 customers, and most have less than 100 transactions. In addition, the largest historical figure was 7,676. This imbalance resulted in the data set being divided into three groups.
- The dataset contains 37 regions, but the majority is concentrated in five Regions. This aspect could affect the customer's mobility pattern identification since most only buy in one region.
- No purchases were recorded on Saturdays, which generated a disproportion between weekdays and weekends.

4.2. Models Construction

Our proposal evaluates the RNN, LSTM, and GRU architectures for predicting the next transaction by analyzing recorded transaction history. It presents a solution to predict the next transaction using sequential methods. Given Tu, a set of user u transactions, each transaction contains $t_i = \{id_i, t_i, c_i, r_i, mb_i, p_i\}$: Customer_ID ($id$), day type ($t$), region ($r$), product category ($c$), monthly budget ($mb$), and payment ($p$). With a set Tu_m of size m where $m > 2$, each Tu must be divided into subsets of size k called "blocks" to predict $transaction_{k+1}$. Two key points in this proposal are: (1) By analyzing the six features in Table 2, relationships were identified to predict the next transaction in history; each feature helped gather information and identify common relationships and irregular actions. Also, (2) the proposal utilizes the parameter k, representing the block size for extracting information, enabling model training with histories of varying sizes.

Input Preparation. Encoding non-numeric attributes is essential to working with neural networks. There are various methods to encode categorical variables, such as: (a) Ordinal Encoding, assigning each instance an integer, and (b) One Hot Encoding, assigning each instance a binary vector. We opted for Ordinal Encoding due to its simplicity in implementation. Therefore, the inputs and outputs were modeled differently, as indicated in section 3. The objective of the model was to predict the next transaction in each block of size k ($k > 2$); that is, the target of a block of size k will be the transaction $k + 1$, as explained in Figure 2. The model needs to be trained with all blocks of size k that can be assembled from a history of size n; so that the model is fed with multiple sequences.

Model Implementation. After preparing the inputs, we implemented architectures based on recurrent networks using the TensorFlow package. For the RNN, we used the BasicRNNCell; for LSTM, we used the BasicLSTMCell; and for GRU, we used the GRUCell, as recommended by previous modeling. Detailed configuration of each network is described in section 5.

5. Experiments and Results

The main objective is to predict the next purchase transaction. We have implemented three types of architectures based on recurrent neural networks and the configuration of parameters to improve results. We performed five experiments, such as:

1. Evaluation of the impact of block size (h).
2. Assessment of the impact of the number of layers.
3. Assessment of the impact of the number of neurons.
4. Evaluation of the number of epochs or iterations.
5. Comparison of architecture performance with baselines

5.1. Experimental Setup

We used 70% data for the test, 10% for validation, and 20% for testing. The initial configuration use the following parameters: $number\_of\_layers = 2$, $number\_of\_neurons = 256$, $activation\_function = ReLU$, $epochs = 100$. Some static parameters were $AdamOptimizer$ and $batch\_size = 250$, so they must be configured for computational cost reasons. The initial three experiments aimed to determine the network configuration, utilizing *Group A* and *Group B*. In contrast, *Group C* was employed for experiments 4 and 5 to assess outcomes with the entire dataset. Accuracy was computed in all trials, while precision, recall, and F1-score were solely evaluated in the final one, as both were defined as benchmark metrics for comparing baselines.

Metrics - In Equation 1, we can see the calculation of the total metric, where α represents the accuracy of *CustomerID*, β represents the accuracy of the type of day, δ represents the region accuracy, γ represents the category accuracy, ϵ represents the accuracy of *monthly budget*, and λ represents the payment accuracy. These values were determined based on the importance level of each *feature* identified by the authors during data exploration. They may be adjusted in future iterations based on additional insights. For example, the *feature* region was assigned the lowest weight because most customers recorded their purchases in a single region. Also, accuracy, precision, recall, and F1-score are the standard metrics for comparing *baselines*.

$$Total_{metric} = 0.15 \cdot \alpha + 0.20 \cdot \beta + 0.20 \cdot +0.10 \cdot \delta + 0.20 \cdot \epsilon + 0.15 \cdot \lambda \qquad (1)$$

5.2. Experiment 1 - Evaluation of the impact of block size (h)

This experiment was based on running the RNN, LSTM, and GRU architectures for *Group A* and *Group B* with the initial network configuration. In this case, performance was evaluated by defining two different values of the block size: $h = \{7.30\}$, which corresponds to the number of days in a week or a month. From this, the accuracy for each feature is obtained, in addition to the total accuracy through Equation 1.

The models that obtained the best results were the RNN and LSTM architectures and are shown in Table 3. Furthermore, a smaller block size ($h = 7$) seems to be suitable for the complete data set, as in the case of *Group B*, for which the RNN model (of lower complexity) was sufficient for analysis. On the other hand, *Group A*, which did not have complete histories, obtained good results with $h = 7$ employing an LSTM architecture. In the case of feature Region, considering $h = 7$, 8.8% accuracy was achieved with *Group A*. In *Group B*, the accuracy obtained was 94.4%. Therefore, the prediction of attributes is more balanced in this last group. Regarding block size, $h = 7$ yields better outcomes for *Group A* and *Group B* across the three architectures. Additionally, the purchased quantity was not distributed between weekdays and weekends, suggesting that a larger block size negatively impacts the results.

Table 3. Accuracy of block size evaluation (h)

		RNN	**LSTM**	**GRU**
$h = 7$	Group A	60.3%	59.6%	57.3%
	Group B	**70.3%**	62.5%	66.5%
$h = 30$	Group A	57.3%	**58.4%**	42.7%
	Group B	50.7%	48.3%	58.0%

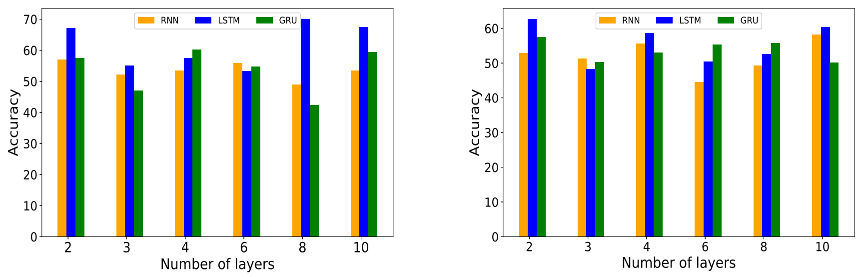
5.3. Experiment 2 - Assessment of the impact of the number of layers

From the results obtained in subsection 5.2 the value of $h = 7$ was defined. RNN, LSTM, and GRU architectures were executed for *Group A* and *Group B* with values of $num\_layers = \{2, 4, 6, 8, 10\}$. In addition, the accuracy for each feature was obtained, as well as the total accuracy using Equation 1.

In Figure 3(a), it is observed that the best accuracies were obtained by the LSTM architecture for *Group A* with 8 and 10 layers. The LSTM network with 8 layers achieved almost 70% accuracy, but the feature Region achieved only 15.2% accuracy. Furthermore, LSTM obtained an accuracy of 60.3% with 10 layers and 49.3% for the Region attribute, which indicates better performance in the calculation of this feature. In the case of RNN and GRU architectures, good results were not obtained compared to LSTM. In summary, training with a greater number of layers has a positive impact when dealing with the feature Region, although the dataset has smaller histories (*Group A*).

In Figure 3(b), the best accuracy for *Group B* was obtained by the LSTM architecture with 62.6% with 10 layers, however the prediction of features was not equitable because feature Region was unable to exceed 15.7% accuracy. Meanwhile, the GRU architecture only achieved an accuracy of 58.8% accuracy with two layers, which was not enough to surpass the results obtained by *Group A*.

Regarding the number of layers, it is observed that a greater number of layers can improve the results; however, once again, the feature Region recorded a low accuracy because it does not have variety in the data. Even so, the LSTM architecture improved the results for *Group A* and *Group B*.



(a) (b)

Figure 3. Model Accuracy with different number of layers for each group

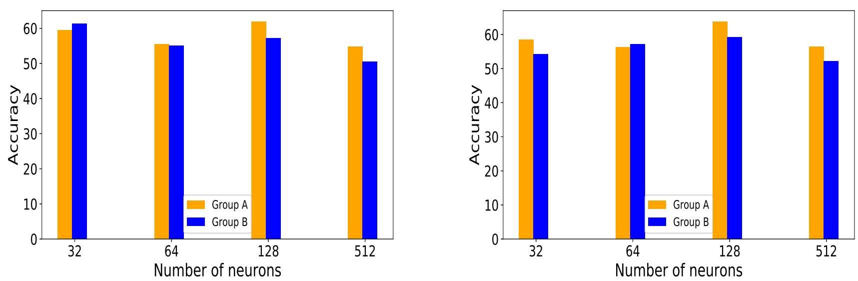
5.4. Experiment 3 - Assessment of the impact of the number of neurons

From the results obtained in subsections 5.2 and 5.3, the values of $h = 7$ and $num\_layers = 10$ were defined. Subsequently, just LSTM and GRU architectures are executed for *Group A* and *Group B*, as they were the architectures that obtained the best results in previous experiments. Then, the values of $num\_neurons = \{32, 54, 128, 512\}$ were defined, and the accuracy was obtained for each feature, in addition to the total accuracy through Equation 1.

In Figure 4(a), for *Group A*, the LSTM architecture obtained good accuracy with $num\_neurons = 128$, however, the feature Region had a low accuracy of 11%. Meanwhile, for *Group B*, the LSTM architecture obtained better results with $num\_neurons = 32$, in addition to balanced accuracies for each feature.

In Figure 4(b), for *Group A* and *Group B*, the GRU architecture achieved greater accuracy with $num\_neurons = 128$. *Group A* recorded low accuracy for feature Region; however *Group B* obtained balanced accuracies for each feature.

Regarding the number of neurons, the $num\_neurons = 128$ improved the results, but the LSTM architecture always obtained superior results. The number of neurons is related to the complexity of the model, as it favors the ability to learn patterns in the model. Also, $num\_neurons = 128$ allows the LSTM architecture to achieve balanced results in features about *Group B*.



(a) (b)

Figure 4. Model Accuracy with different numbers of neurons for each group

5.5. Experiment 4 - Evaluation of the number of epochs or iterations

This experiment was performed with the parameters obtained previously, that is, $h = 7$, $num\_layers = 10$ and $num\_neurons = 128$. This experiment only considered the LSTM and GRU architectures because they had the best results from previous experiments. Furthermore, only *Group C* was used because it was necessary to know the results about complete and incomplete histories. Then, the value of $num\_épocas = \{100, 200, 300, 400, 500\}$ was defined, and then the accuracy was obtained for each feature, in addition to the total accuracy through Equation 1.

In Table 4, LSTM achieved its highest accuracy with 500 iterations, but the GRU achieved its highest accuracy only with 200 iterations. Even so, the highest accuracy between the two (60.9%) was obtained by the LSTM architecture. In this way, it was verified that LSTM managed to improve the results with a greater number of epochs than the GRU architecture. On the other hand, the improvements in results when employing *Group C* were not very significant, which could mean that incomplete data affects the analysis.

Table 4. Accuracy of models with different epoch values

	LSTM	GRU
100 epochs	55.8%	56.9%
200 epochs	57.8%	**59.1%**
300 epochs	57.9%	54.7%
400 epochs	56.1%	54.4%
500 epochs	**60.9%**	50.8%

5.6. Experiment 5 - Comparison with Baselines

This experiment considered the execution of the LSTM and GRU architectures for*Group C* in terms of accuracy and precision. These results are considered as the final results that will be compared with the baselines.

The results for each feature are described in Table 5. LSTM obtained the best values for four features (CustomerID, category, region, and monthly budget) and GRU for just two features (Day type and Payment). Unusually, the LSTM architecture achieved high accuracy in predicting the feature Region, but it could be overfitting. In summary, the results so far validate that the LSTM architecture is promising for analyzing purchasing behavior in the Online Retail Dataset, although the dataset does not have histories with the complete information necessary for the analysis.

Table 5. Accuracy for each feature using the final configuration

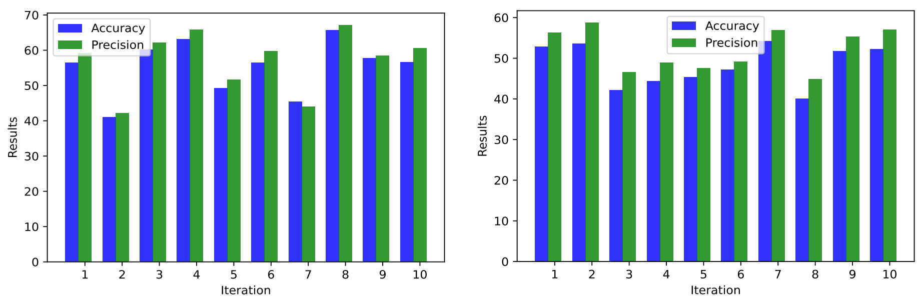
	CustomerID	**Day Type**	**Category**	**Region**	**Monthly Budget**	**Payment**
LSTM	**33.7%**	80.1%	**59.7%**	**95.5%**	**46.2%**	60.8%
GRU	23.6%	**83.3%**	59.4%	88.2%	45.0%	**61.0%**

In Table 6, LSTM achieved 60.9% accuracy, and GRU achieved 59.1%, with LSTM showing superiority. However, GRU achieved 66.7% accuracy compared to LSTM's 63.2%, achieving the best results in recall and F1-score. Despite LSTM's higher accuracy, the model could not be generalized to the test data, possibly due to overfitting. On the other hand, GRU showed better generalization during testing, accurately predicting at least 376,859 transactions. While both models yielded acceptable results, neither surpassed 80% accuracy in predictions, though they notably outperformed the selected baselines. In conclusion, although LSTM obtained the best accuracy, GRU demonstrated superior generalization in forecasting the subsequent transaction in the purchase history using the Online Retail Dataset.

Table 6. Results compared to baselines

	Accuracy	Precision	Recall	F1-score
STPC-PGM	10.0%	12.5%¨	10.4%	15.3%
Naive Bayes	43.2%	23.6%	22.5%	35.1%
LSTM	**60.9%**	**63.2%**	**55.7%**	**57.2%**
GRU	**59.1%**	**66.7%**	**60.4%**	**62.4%**

Results Validation. At this stage, the hold out method was used to execute ten iterations of the model with the final configuration of the previously selected parameters. This method consists of dividing the total dataset into two mutually exclusive subsets[Yadav and Shukla 2016], one for training (parameter estimation) and the other for testing (validation), in which the data is randomly selected. In Figure 5(a), it is observed that for three iterations, the results did not exceed 50%, this may be indicative that the necessary information was not recorded in the training data for the model to learn correctly. In the case of iteration 8, the training data achieved a better execution of the model, and therefore, the best results were obtained. Furthermore, it is observed that the precision metric almost always obtained better results compared to accuracy. Similarly, in Figure 5(b), results didn't exceed 50%, suggesting insufficient training data information for proper learning. For iteration 7, the training data led to improved model execution and, consequently, better outcomes. In conclusion, model results are closely tied to training data quality, necessitating adequate generalization capacity. Exploring alternative training methods like stratification or cross-validation could enhance outcomes.



(a) (b)

Figure 5. Models Validation

Discussion. Predicting customer behavior is a subject that ML techniques can address, but DL models appear more promising for this task. Additionally, we stress the importance of handling large datasets containing sensitive information, making complete data access challenging due to confidentiality constraints. Therefore, we opted to utilize a public dataset, ensuring data privacy. Our experimentation involved two recurrent neural network structures (LSTM and GRU) due to their potential to analyze long-term customer behavior trends. Our findings lead us to the following conclusions:

1. *Challenges and limitations.* LSTM and GRU were trained using a hold-out strategy, leading to observed overfitting issues. As future work, resampling methods and regularization techniques will be explored to optimize the learning process.
2. *Scalability and versatility.* Both models are suitable for datasets similar to the Retail Online Dataset. Better results are obtained with richer customer histories, indicating that the models are scalable for large datasets. Future research could explore the adaptability of these models for other sequential problems.
3. *Privacy and ethics.* The proposed models adhere to GDPR and LGPD principles by using the Online Retail Dataset, which contains standard customer information. While some analyses utilize precise geographic coordinates, which our dataset lacks, it's important to prioritize ethics. The results demonstrate that meaningful analyses can still be conducted.

6. Final Considerations

This paper highlights the importance of predicting customer behavior using three models (RNN, LSTM, and GRU) for sequential behavior modeling. The results demonstrate the effectiveness of these models in working with a dataset that does not contain private customer information, unlike previous works that rely on confidential data. The LSTM model achieved 60.9% accuracy and 63.2% precision, while the GRU model achieved 59.1% accuracy and 66.7% precision, showing improvements over baseline results. Both models proved scalable when applied to different datasets, requiring only the described steps. However, LSTM exhibited challenges in generalizing test data, resulting in lower accuracy compared to GRU and possibly being influenced by inconsistencies in the dataset. Future work could involve employing additional metrics to evaluate results, addressing issues such as overfitting, and exploring alternative training methods (e.g., Bagging or Boosting) to enhance predictive performance. Additionally, the use of hyperparameter optimization libraries could be considered.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This study was also financed in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq).

References

Campagna, D. P., da Silva, A. S., and Braganholo, V. (2020). Achieving gdpr compliance through provenance: An extended model. In *Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 13–24. SBC.

Fleder, M. and Shah, D. (2020). I know what you bought at chipotle for $9.81 by solving a linear inverse problem. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 4(3):1–17.

Huang, C., Wu, X., Zhang, X., Zhang, C., Zhao, J., Yin, D., and Chawla, N. V. (2019). Online purchase prediction via multi-scale modeling of behavior dynamics. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2613–2622.

Ładyżyński, P., Żbikowski, K., and Gawrysiak, P. (2019). Direct marketing campaigns in retail banking with the use of deep learning and random forests. *Expert Systems with Applications*, 134:28–35.

Li, J., Pan, S., Huang, L., et al. (2019). A machine learning based method for customer behavior prediction. *Tehnički vjesnik*, 26(6):1670–1676.

Li, Q., Chen, Z., and Zhao, H. V. (2021). Prima++: A probabilistic framework for user choice modelling with small data. *IEEE Transactions on Signal Processing*, 69:1140–1153.

Martens, D. (2022). *Data science ethics: Concepts, techniques, and cautionary tales*. Oxford University Press.

Martínez, A., Schmuck, C., Pereverzyev Jr, S., Pirker, C., and Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*, 281(3):588–596.

Nery, C., Galante, R., and Cordeiro, W. (2021). FIP-SHA - finding individual profiles through shared accounts. In Strauss, C., Kotsis, G., Tjoa, A. M., and Khalil, I., editors, *Database and Expert Systems Applications - 32nd International Conference, DEXA 2021, Virtual Event, September 27-30, 2021, Proceedings, Part II*, volume 12924 of *Lecture Notes in Computer Science*, pages 115–126. Springer.

Neto, E. R., Mendonça, A. L., Brito, F. T., and Machado, J. C. (2018). Privlbs: uma abordagem para preservação de privacidade de dados em serviços baseados em localização. In *Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 109–120. SBC.

Pinheiro, P. P. (2020). *Proteção de dados pessoais: Comentários à lei n. 13.709/2018-lgpd*. Saraiva Educação SA.

Rendle, S., Freudenthaler, C., and Schmidt-Thieme, L. (2010). Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 811–820.

Ruiz, F. J., Athey, S., and Blei, D. M. (2020). Shopper: A probabilistic model of consumer choice with substitutes and complements.

Safara, F. (2022). A computational model to predict consumer behaviour during covid-19 pandemic. *Computational Economics*, 59(4):1525–1538.

Sarkar, M. and De Bruyn, A. (2021). Lstm response models for direct marketing analytics: Replacing feature engineering with deep learning. *Journal of Interactive Marketing*, 53(1):80–95.

Suarez Mariscal, C., de Lima, B. S. M., Galante, R., and Cordeiro, W. (2023). Assessing explainable recommendations from knowledge graph-based in an international streaming platform. In *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web*, WebMedia '23, page 213–220, New York, NY, USA. Association for Computing Machinery.

Tabianan, Kayalvily e Velu, S. e. R. V. (2022). K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability*, 14(12):7243.

Vasupula, NarsingRao e Munnangi, V. e. D. S. (2022). Modern privacy risks and protection strategies in data analytics. In *Soft Computing and Signal Processing: Proceedings of 3rd ICSCSP 2020, Volume 2*, pages 81–89. Springer.

Wachter, S., Mittelstadt, B., and Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the gdpr. *Harv. JL & Tech.*, 31:841.

Wang, W., Xiong, W., Wang, J., Tao, L., Li, S., Yi, Y., Zou, X., and Li, C. (2023). A user purchase behavior prediction method based on xgboost. *Electronics*, 12(9):2047.

Wen, Y.-T., Yeh, P.-W., Tsai, T.-H., Peng, W.-C., and Shuai, H.-H. (2018). Customer purchase behavior prediction from payment datasets. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 628–636.

Wieringa, J., Kannan, P., Ma, X., Reutterer, T., Risselada, H., and Skiera, B. (2021). Data analytics in a privacy-concerned world. *Journal of Business Research*, 122:915–925.

Yadav, S. and Shukla, S. (2016). Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In *2016 IEEE 6th International conference on advanced computing (IACC)*, pages 78–83. IEEE.

Yuan, Q., Zhang, W., Zhang, C., Geng, X., Cong, G., and Han, J. (2017). Pred: Periodic region detection for mobility modeling of social media users. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 263–272.

Zhou, M., Ding, Z., Tang, J., and Yin, D. (2018). Micro behaviors: A new perspective in e-commerce recommender systems. In *Proceedings of the eleventh ACM international conference on web search and data mining*, pages 727–735.

Zhu, B., Tang, W., Mao, X., and Yang, W. (2020). Location-based hybrid deep learning model for purchase prediction. In *2020 5th International Conference on Computational Intelligence and Applications (ICCIA)*, pages 161–165. IEEE.