A dynamic approach to churn prediction using time series classification

Si Dan Tran[12], Duy Duan Nguyen[12], Thuy Vy Nguyen[12], Thi Kim Hien Le [12*]

[1]University of Economics and Law, Ho Chi Minh City, Vietnam

[2]Vietnam National University, Ho Chi Minh City, Vietnam

[*]Corresponding author: hienltk@uel.edu.vn

ARTICLE INFO

DOI:
10.46223/HCMCOUJS…

Received:

Revised:

Accepted:

Keywords:

churn prediction, deep learning, dynamic behavior, time series classification, MINIROCKET

ABSTRACT

As customer retention is important due to the high costs of acquiring new customers, churn prediction is necessary to identify valuable customers at risk of leaving. Previous works on churn prediction models often use static data, overlooking the dynamic nature of customer behavior, which leads to less accurate predictions and less timely response of retention campaigns. To address this issue, this paper proposes a dynamic approach to churn prediction using weekly multivariate time series data to better capture changes in customer behavior. Two time series classification models, namely MINIROCKET and LSTM-SLP, were compared to three static models: Random Forest, XGBoost, and Ridge Regression. Experimental results demonstrate that the MINIROCKET model outperforms all other state-of-the-art methods, achieving the highest F1 score. Furthermore, both time series classification models, LSTM-SLP and MINIROCKET, consistently yielded better results in comparison to traditional models using static data. This superiority highlights the effectiveness of considering temporal dynamics in classification models, which previous static models fail to address. This finding indicates the potential of time series classification in enhancing churn prediction accuracy and, consequently, improving customer retention strategies.

1. Introduction

In nowaday's business environment, enterprises operating within competitive markets primarily depend on the revenue generated from their customer base. Therefore, it is necessary to focus on promoting customer retention rather than attracting new customers (Sağlam & El Montaser, 2021). Customer churn prediction has also become a major concern in the marketing and management literature in recent times. One of the most direct and effective methods to retain current customers is that companies should predict customers at risk of leaving over time and respond promptly. Recognizing the signs of potential churn, satisfying customer needs, restoring, and re-establishing loyalty are actions that are believed to help organizations reduce the costs of acquiring new customers (Mitkees, Badr, & Elseddawy, 2017).

Churn prediction is a management science problem to which machine learning approaches can be applied (Lemmens & Gupta, 2020). Based on the historical data, a machine learning model can be trained to classify the customer whether they will churn or not in the future. These research applied several common SVM, Decision Tree or Logistic Regression (Peng, Peng, & Li, 2023; Rachid et al., 2018; Tianyuan & Moro, 2021). Research by Kumar & D. (2016) showed that combining SVM with reinforcement learning algorithms increases accuracy and higher performance, which is considered potential for future prediction. (Tianyuan & Moro, 2021) performed a review of 40 related articles and demonstrated that the widely used data mining techniques are decision trees, support vector machines, and Logistic regression.

Despite making positive contributions to improving prediction models, the main limitation of these studies is the use of static data and ignoring the changing nature of customer behavior (Alboukaey, Joukhadar, & Ghneim, 2020). According to Bornemann et al. (2018), data changes over time and these changes can reveal characteristics or trends in behavior. Therefore, time series data is more suitable for predicting customer churn based on customer behaviors than building a churn prediction model. Another limitation in previous models is that they are only suitable for performing prediction monthly and use the results to propose retaining strategies in the following month. As customer behavior and the nature of time series data change rapidly, month-range predictions do not ensure the timeliness in case customers churn at the beginning of the month (Khan et al., 2015; Zhang et al., 2016). (Alboukaey et al., 2020) proposed that using a daily or weekly churn model helps design customer retention strategies at any time of the month depending on newly updated trends replacing old trends from the previous month.

In this study, we propose weekly churn prediction based on customer behavior changes presented as multivariate time series. The main contributions of the study are summarized as follows:

- Applying time series classification to predicting customer churn behavior.
- Proposing an approach to predict customer churn based on weekly customer behavior.
- Compare the performance of the proposed approach with traditional models such as Random Forest, XGBoost and Ridge Regression.

The rest of the paper is organized as follows. Section 2 presents the related work that provides selected recent collaborative studies based on the mentioned theories. Section 3 describes the proposed methodology, focusing on knowledge-based methods to improve customer churn prediction techniques. Section 4 presents the experimental results and discussion. Finally, Section 5 concludes the paper and gives directions for future research.

2. Literature review

From the perspective of traditional methods aimed at churn prediction, machine learning techniques operating on static data have made significant contributions to this field. The study by (Tianyuan & Moro, 2021) surveyed churn prediction techniques, highlighting popular methods such as Decision Trees, Support Vector Machines (SVM), and Logistic Regression. These methods utilize tabular data rather than time series data, with each training sample being an aggregated record from past data, ignoring temporal changes. This trend continues in studies like (Peng et al., 2023), which improved traditional machine learning methods and proposed the GA-XGBoost model, demonstrating high efficiency and deep insights into customer churn factors in the service industry. However, these models use static data, failing to reflect changes in customer behavior over time. (Bornemann et al., 2018) noted that temporal data can reveal patterns or groups of similar values, properties, and entities, suggesting that using static data may overlook significant customer behavior characteristics, reducing classification accuracy and chances on deeper analysis.

Recent research trends in churn prediction have shifted towards time series classification, a method which has been significantly impacting data science and various practical applications. Time series classification methods are diverse, with deep learning models showing notable effectiveness. (Karim, Majumdar, Darabi, & Chen, 2017) proposed the LSTM-FCN model, which surpassed existing advanced models in at least 43 datasets, highlighting the superiority of deep learning in time series classification. (Ismail et al., 2019) explored the performance of neural network architectures such as MLP, CNN, and ESN in time series classification, providing a finding that FCN outperformed other approaches on 36 datasets, demonstrating its effectiveness. However, these studies used limited types of input features, focusing only on behavioral or temporal data. In the field of healthcare proposed the LSTM-SLP model, combining time-varying and static features, achieving impressive performance metrics. This approach opens up new possibilities by integrating static data for time series classification rather than relying solely on time series data.

In addition to deep learning, new algorithms such as Shapelets (Liu et al., 2022) have enhanced the performance of time series classification. The proposed algorithm uses three strategies: random Shapelet selection for efficiency, embedding multiple canonical time series features for better adaptability and accuracy, and a random forest classifier to ensure generalization. Experiments on 112 UCR time series datasets showed this algorithm outperforms several other advanced time series classification methods. Another study by (Óskarsdóttir et al., 2018) proposed a novel method called time series forest to predict churn in telecommunications with weekly time series data. This research demonstrated the effectiveness of time series classification for churn prediction, however, appears to have limitations due to the experimental sampling of the training and testing datasets within the same time, which could hinder the classifier's ability to generalize to different temporal characteristics.

Churn prediction using time series classification also faces computational complexity challenges. For example, the 1-NN DTW method has quadratic time complexity concerning time series length and reduced accuracy with noise (Schäfer, 2016). The solution is a model that can predict accurately, meanwhile minimizing time-space complexity. Among the state-of-the-art methods, kernel-based classification methods, particularly MINIROCKET proposed by (Dempster, Schmidt, & Webb, 2021), are proved to be optimal for time series classification, balancing speed, and accuracy. MINIROCKET is significantly faster and more accurate than other methods with similar computational costs.

To overcome the mentioned difficulties, we propose a dynamic approach for the weekly churn prediction based on the customer’s weekly data. In this paper, the two proposed time series classification models, namely MINIROCKET and LSTM-SLP, are compared with various traditional methods using static data. This dynamic approach provides minimized time-space complexity and improved accuracy in predicting customer churn based on time series data, demonstrating the effectiveness and potential of a novel method in churn prediction.

3. Research methodology

3.1. Experimental Design

This article introduces a research methodology which has important implications in the application of time series data in identifying customer churn with weekly time steps. The model provides a comparison of time series classification methods (MINIROCKET and LSTM-SLP models) with classification methods using static features.

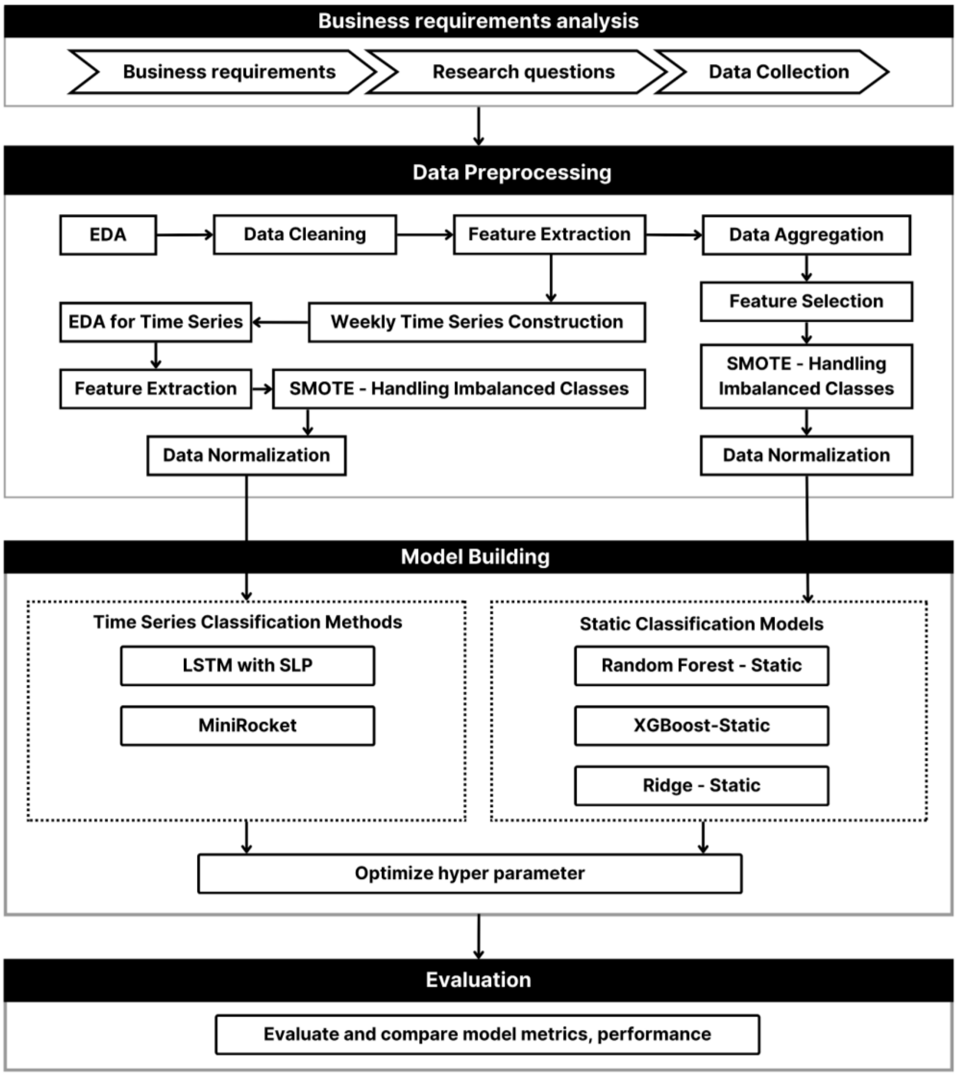

***Figure 1**. General research model (Source: Authors)*

The research model includes 4 phases as **Figure 1**. The first phase involves data collection and identifying the business issues surrounding customer churn. The second phase is the data preprocessing process, including exploratory analysis, data cleaning, feature extraction, data transformation into weekly time series, and handling data imbalance. In the third phase, the corresponding post-processed data will be used to build the machine learning model. Finally, the fourth phase involves discussing and evaluating the models to select the most suitable one.

3.2. Data Collection

The data used was introduced at the 11th ACM International Conference on Web Search and Data Mining (Y. Chen, Xie, Lin, & Chiu, 2018) churn customer data on the online subscription music platform KKBOX. Data includes 3 characteristic groups including: listening behavior of customers, demographics of customers and transaction history of customers.

3.3. Data pre-processing

The data used for training and testing the model's effectiveness were sampled over the period from October 12, 2016 to February 28, 2017, corresponding to 20-week time series data with each time step equivalent to 1 week. Time series data is aggregated and transformed weekly. After generating the time series data from the original data, the time frames are created into 16-week long time series using the "sliding window" algorithm. The time series are named "Training Window 1", "Training Window 2", "Training Window 3" and "Testing Window", presented in **Figure 2**.

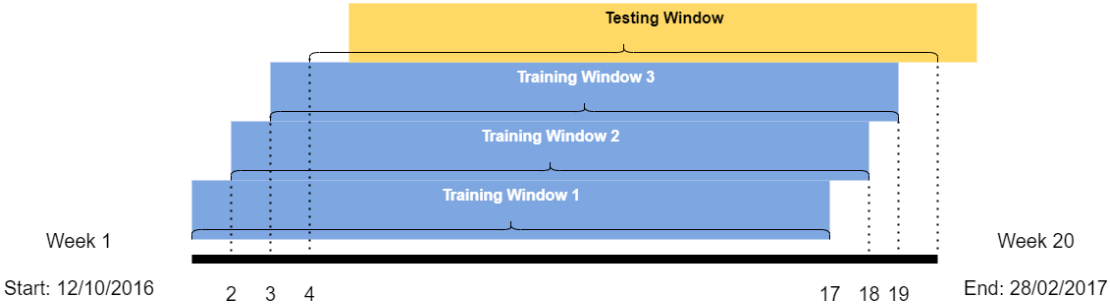

Figure 2. *Sliding window sampling (Source: Authors)*

The data provided still exists inconsistencies in customer IDs, null values and unrealistic values. Therefore, outliers are handled using the IQR method. This method determines the lower and upper bounds, then the exotic values are defined as values outside this range, this range is defined as from $Q1 - 1.5 * IQR$ to $Q3 + 1.5 * IQR$. Based on the general understanding of music listening behavior, all outliers evaluated as extraneous values should be removed without affecting the meaning of the data.

After understanding the data and identifying features that can cause noise, feature extraction is performed to extract new information and reduce the dependence between features (correlation). The new features extracted also aim to emphasize the dynamic nature of time series. The 14 features used as input features are: num_25, num_50, num_75, num_985, num_100, num_unq, total_secs, actual_amount_paid, diff_actual_plan_paid, make_cancellation, city, age_group, gender, registered_via.

Table 1*. Input features extracted from user records*

Feature	Description
num_25	Number of songs played for less than 25% of its length.
num_50	Number of songs played between 25% and 50% of its length.

num_75	Number of songs played between 50% and 75% of its length.
num_985	Number of songs played between 75% and 98.5% of its length.
num_100	Number of songs played for more than 98.5% of its length.
num_unq	Number of unique songs played.
total_secs	Total time spent listening to songs, measured in seconds.
actual_amount_paid	Actual amount of money paid by the user.
diff_actual_plan_paid	Difference between the actual amount paid and the planned amount.
make_cancellation	Indicator of whether the user made a cancellation at the time.
city	The city where the user resides.
age_group	Age group classification of the user.
gender	Gender of the user.
registered_via	The method or platform through which the user registered.

3.4. Classification based on static characteristics

Static features refer to characteristics of data that do not change over time. These can include aggregated behavioral features, transactional features or a mix of demographic and behavioral features. Examples include a customer's age, gender, or aggregated purchasing habits, amount over a specific period. While these features can be effective predictors, they often overlook the temporal nature of customer behavior, potentially reducing their ability to accurately discriminate patterns (Z. Y. Chen, Fan, & Sun, 2012). For churn prediction by static characteristics, we experiment on the static characteristics of the data sample with machine learning models including Random Forest, Ridge Regression and XGBoost. The models are optimized using the GridSearchCV method to select the most optimal parameters.

3.5. Time Series Classification

According to (Theissler et al., 2022) , a time series classification dataset $D = (X, Y)$ is a set consisting of n time series, $X = \{x_1, x_2 \dots, x_n\} \epsilon R^{n \times m \times d}$, with a corresponding vector of assigned labels (or classes) $Y = \{y_1, y_2 \dots, y_n\} \epsilon N^n$. Time series classification is the task of training a function or mapping f from input data X to a probability distribution over the classes in Y. The resulting function f takes an input time series x and returns the label y of the class to which x belongs based on what f has learned, i.e., $y = f(x)$. The notation $Y = f(X)$ is used as a shorthand for $Y = \{f(x) \mid x \epsilon X\}$.

After obtaining a dataset containing multivariate time series, the next step is to use the newly obtained training set and testing set to build the MINIROCKET model, as well as LSTM-SLP. In particular, both the models are optimized using the GridSearchCV method, including cross-validation and hyperparameter search.

3.5.1. MINIROCKET

MINIROCKET is a transformation method based on a minimal random convolution kernel (a small, fixed set of twisted nuclei) to transform time series data. The variation in the size of the convolutional kernels and the PPV (Positive predictive value) aggregation feature to calculate a unique characteristic for each resulting characteristic map (i.e. the ratio of positive values) help create a characteristic that can be used to train a classifier (Tan, Dempster, Bergmeir, & Webb, 2022). MINIROCKET. (Dempster et al., 2021) improves upon ROCKET by Dempster et al. (2020) by reducing randomness and speeding up the transformation. It achieves a balance between accuracy and optimal parameter selection.

MINIROCKET employs several key modifications to the kernel setup to achieve efficiency and accuracy:

- Length: Kernels are fixed at length 9.
- Weights: Weights are α=−1 and β=2, ensuring the kernel's sum is zero. The subset of used kernel includes:

$$[\alpha,\alpha,\alpha,\alpha,\alpha,\alpha,\beta,\beta,\beta],[\alpha,\alpha,\alpha,\alpha,\alpha,\beta,\alpha,\beta,\beta],[\alpha,\alpha,\alpha,\alpha,\beta,\alpha,\alpha,\beta,\beta],\ldots$$

- Bias: Bias values are drawn from the quantiles (e.g., $[0.25,0.5,0.75]$) of the convolution output $C = W_d \ast X$ of a randomly selected training example.
- Dilation: Fixed dilations are used, adjusted to the input length $l$input with a maximum number of 32 dilations per kernel.
- Padding: Alternating zero padding is applied for each kernel/dilation combination.
- Features: Only PPV (Proportion of Positive Values) is used, with 10,000 features. For a kernel W, input X, and bias b:

$$PPV(X \ast W - b) = \frac{1}{n}\sum[X \ast W - b > 0] \qquad (1)$$

MINIROCKET optimizes the transformation with the following strategies:

1. PPV for W and -W simultaneously for convolution C=X∗W and bias b:

$$PPV(X \ast W - b) = 1 - PPV(b - (X \ast W)) \qquad (2)$$

   This avoids redundant calculations.

2. Reusing Convolution Output: For a kernel W and dilation d, reuse $C = X \ast W_d$ for multiple bias values, reducing computational cost.
3. Avoiding Multiplications: Simplified weights $\alpha = -1$ and $\beta = 2$ reduce multiplication operations.
4. Efficient Dilation Handling: Process all kernels for each dilation almost simultaneously.

MINIROCKET refines the transformation process, reducing randomness and enhancing computational efficiency while maintaining high classification accuracy. After the transformation, it is suggested to use Ridge Regression for small-sized data to train and make predictions. For large-sized data, Logistic Regression and stochastic gradient descent is recommended. However, Logistic regression can be time-consuming on large data (Song et al., 2021), therefore Random Forest is chosen to facilitate a comprehensive comparison with the static Random Forest model.

In the experimental setup of the MINIROCKET model, the number of kernels is set to 500. Moreover, the hyperparameters of the Random Forest classifier are optimized by searching and evaluating each combination in the hyperparameter space, along with 5-fold cross-validation.

3.5.2. Bidirectional LSTM combined with Single Layer Perceptron

The idea of combining Bidirectional LSTM and Single Layer Perceptron model was presented by (Hyland et al., 2020). The basis for this approach is that the Bidirectional LSTM consists of two LSTM networks, called the forward LSTM and backward LSTM. Each LSTM has hidden states and gates to regulate the input information and hidden states. The Bidirectional LSTM will learn the relationships between time series features, while the Single Layer Perceptron will focus on the static features of the model. Single Layer Perceptron, one of the earliest and first introduced neural networks, was proposed by (Rosenblatt, 1958). The perceptron is also known as an artificial neural network with one input layer and one output layer, without any hidden layers. The model architecture includes 1 SLP layer and 2 LSTM layers, followed by a concatenation layer to combine the outputs from the LSTM and SLP layers. This combined output is then passed to another dense layer, followed by an output layer with sigmoid activation to predict whether churn will occur or not. All parameters and layers' details are listed in Table 2.

The problem of class imbalance data is handled by the method of using the focal loss function. This method reduces the influence of class imbalances by adjusting the model's attention to the minority class through the establishment of indicators. The focal loss function introduces two parameters: alpha (α) and gamma (γ). The parameter α is used to balance the importance of positive/negative examples, allowing the model to focus more on the minority class. The parameter γ reduces the relative loss for well-classified examples, putting more focus on minor samples. Additionally, the Early Stopping technique is employed during model training to halt the process when accuracy ceases to improve, thereby saving training time. In this paper, the Bidirectional LSTM combined with Single Layer Perceptron model is called LSTM-SLP as a shorthand.

***Table 2**. LSTM-SLP model layers' detail*

Layer	Kernels/Hidden Units	Output Dim	Activation
Input (Time Series)	-	16x10	-
Bidirectional LSTM Layer 1	128	16x256	-
Dropout Layer 1	-	16x256	-
Bidirectional LSTM Layer 2	64	1x128	-
Dropout Layer 2	-	1x128	-
Input (Static)	-	1x4	-
Dense Layer 1	64	1x64	ReLU

Concatenate Layer	-	1x192	-
Dense Layer 2	64	1x64	ReLU
Output Layer	1	1x1	Sigmoid

4. Result and discussion

4.1. Experimental result

For MINIROCKET, Random Forest - Static, XGBoost - Static and Ridge Regression - Static, SMOTE (Synthetic Minority Over-sampling Technique) was used to address the problem of class imbalance, with the minority class taking up only 5% in total. Also, due to class imbalance, F1 score is suggested to be the main evaluation metric (Du et al., 2023). The input features of these models are: ‘num_25’, ‘num_50’, ‘num_75’, ‘num_985’, ‘num_100’, ‘num_unq’, ‘total_secs’, ‘actual_amount_paid’, ‘diff_actual_plan_paid’, ‘make_cancellation’. Additionally, GridSearchCV is applied across all methods to fine-tune the models and optimize hyperparameters. The hyperparameters for both Random Forest in the MINIROCKET-based approach and the Random Forest - Static model are presented in Table 3.

***Table 3**. Hyperparameter tuning for Random Forest Classifier*

Search hyperparameter	Optimal hyperparameter
n_estimators: [100, 250, 500, 1000]	n_estimators: 500
max_depth: [10,15,20]	max_depth: 20

Meanwhile, the LSTM-SLP model applies the Focal Loss function method to handle class imbalance. For the input layers, 4 features are designated as static, including ‘city’, ‘age_group’, ‘gender’,’ registered_via’, while the other 10 features are considered dynamic, providing temporal insights. Additionally, it is optimized using the adam (Adaptive Moment Estimation) optimizer. This optimizer dynamically adjusts learning rates for each parameter during training, leading to improved model convergence and overall performance. The parameters used to build the LSTM-SLP model are presented in **Table 4**.

***Table 4.** Parameters for a LSTM-SLP model*

Parameter	Value
α	0.8
γ	2.0
epochs	2000
batch_size	64

Table 5 provides a comparison between Time Series Classification (TSC) methods and Static Models in the context of churn prediction. All models are evaluated using various performance metrics, including precision, accuracy, recall, F1 score, and log loss:

Accuracy: Accuracy is the average precision of the algorithms, defined as the ratio between the predicted results and the actual data. The MINIROCKET model achieved a prediction accuracy of 95%, while other models had slightly lower accuracies of 94% and 93%, respectively. This means that out of 100 predicted data points, MINIROCKET correctly predicted 95 data points compared to the actual results.

Precision: Precision is defined as the number of correctly predicted positive observations divided by the total predicted positive observations. The MINIROCKET model achieved a macro-average precision of 82%, indicating that the model correctly predicted, on average, 82 out of 100 data points in the positive class.

Recall: Recall measures the proportion of actual positive cases that were correctly identified. Both MINIROCKET and LSTM-SLP demonstrated high recall values, at 0.77 and 0.8, respectively. This indicates that they effectively identify positive cases.

F1_score: The F1 score is the harmonic mean of precision and recall, providing a comprehensive assessment of the model's performance. MINIROCKET had the highest F1_score of 0.79, indicating that it balances precision and recall effectively.

Log Loss: Log loss measures the degree of uncertainty in the model’s probability predictions for the classes. MINIROCKET had the lowest log loss at 1.74, compared to other models. This indicates that it predicts class probabilities accurately and consistently.

Overall, both time series classification models MINIROCKET and LSTM-SLP outperform the static models across all metrics. The time series classification models demonstrate superior accuracy, precision, recall, and F1 scores, highlighting their effectiveness in handling time series data and making accurate predictions, without ignoring the dynamic nature of data. The MINIROCKET model shows outstanding performance, excelling in both prediction accuracy and probability estimation, as reflected by its low log loss. These results emphasize the robustness and reliability of the MINIROCKET-based approach in predicting churn, making them preferable over static models for such tasks, while having low time complexity.

Table 5. *Model evaluation results*

Model	Precision	Accuracy	Recall	F1_score	Log loss
MINIROCKET	0.82	0.95	0.77	0.79	1.74
LSTM-SLP	0.74	0.94	0.8	0.76	1.95
Random Forest - Static	0.7	0.93	0.77	0.73	2.5
XGBoost - Static	0.69	0.93	0.76	0.72	2.54
Ridge Regression - Static	0.77	0.94	0.51	0.5	2.04

4.2. Discussion

The discussion should highlight the special and outstanding points of the research; explain research results, the impacts; compared with previous studies.

Experimental results on multiple models show that the MINIROCKET method is the most effective with an F1 score of 0.79, followed by LSTM-SLP with an F1 score of 0.76. These results demonstrate that the two proposed time series classification models outperform state-of-the-art methods which use static data. In terms of studies that also propose dynamic methods for churn prediction, (Alboukaey et al., 2020) suggests a daily churn prediction approach with three RFM-based, CNN-based, and LSTM-based models. However, the best method, namely the RFM-based model, only achieved an F1 score of 0.571 with the LSTM-based model, which is lower than the MINIROCKET model under similar testing conditions.

The proposed MINIROCKET and LSTM-SLP method has shown significant improvements over previous research. First, the research team proposes using weekly time series data. Using weekly time series data not only reduces computational resources compared to daily time series (by 7 times) but also maintains the accuracy of the prediction model. Additionally, the sampling method for the experiment was conducted more realistically than in (Óskarsdóttir et al., 2018). In this proposed approach, experimental data is selected based on 4 weekly time windows, with the last window as the testing set. This sampling method is more realistic compared to using both training set and testing set within the same time period as in previous research. Another important issue is evaluating the cost of false predictions. The presented matrix shows the existence of incorrect predictions, and clearly identifying their costs can help adjust the model to reduce the risks and costs of incorrect predictions.

As mentioned, the approach used is based on weekly time series data. However, this method emphasizes flexibility, meaning it can be applied at any time to support customer retention campaigns. Therefore, it is recommended to use this time series classification approach regularly, specifically that predictive analyses should be conducted weekly or no longer than an interval of a month. This is because as customer behavior tends to change weekly within a month and in the lead-up to a churn decision, they begin to behave differently. (Alboukaey et al., 2020) provided an example of this issue as follows: suppose $x\_month = (600,600,600)$ is a vector representing a customer's monthly expenditure over the last three months. According to this representation, this customer is a non-churner with stable behavior. However, it could also represent a churner with stable expenditure for the first 8 weeks (i.e., the first two months) followed by a peak and then a drop at the beginning of the last month, for example, $x\_weekly = (20,20,20,20,180,160,120,140,0,0,0,0)$. Therefore, if predictions are conducted monthly and data is presented monthly, customer retention campaigns may not react in time to such a change in customer behaviors. To optimize the customer retention rate, it is essential to conduct churn predictions frequently, no longer than a month apart.

5. Conclusions & recommendations

In conclusion, this research proposes a new approach based on MINIROCKET and LSTM-SLP model to predict customer churn based on weekly time series data, focusing on changes in customer behavior. The results proved the effectiveness of the proposed approach compared to some traditional methods such as SVM, Decision Tree or Logistic Regression. This proposed approach contributes to improving prediction performance, which helps businesses develop and apply in the practical context to build suitable strategies for retaining the customers.

In the future, the research will promote research directions related to dynamic behaviors in churn prediction and developing the explainable MINIROCKET-SHAP time series classification model. The model is expected to overcome the weakness in previous works, which is that the predicted results cannot be explained, and extensions proposed from the predicted results. This is a crucial step towards creating a comprehensive model and an effective tool for churn management.

ACKNOWLEDGEMENTS

We would like to express our sincere thanks to all organizations and individuals who supported and contributed to this research work. Your advice not only includes providing content, form and financial conditions but is also extremely valuable support. We appreciate reasonableness and understanding of construction-related aspects, including formulations and conflicts of interest. Thank you very much for your support and dedication, which has helped us complete this process successfully.

References

Alboukaey, N., Joukhadar, A., & Ghneim, N. (2020). Dynamic behavior based churn prediction in mobile telecom. *Expert Systems with Applications*, *162*. doi:10.1016/j.eswa.2020.113779

Bornemann, L., Bleifuß, T., Kalashnikov, D., Naumann, F., & Srivastava, D. (2018). Data Change Exploration Using Time Series Clustering. *Datenbank-Spektrum*, *18*(2). doi:10.1007/s13222-018-0285-x

Chen, Y., Xie, X., Lin, S. De, & Chiu, A. (2018). *WSDM Cup 2018: Music Recommendation and churn prediction*. In *WSDM 2018 - Proceedings of the 11th ACM International Conference on Web Search and Data Mining* (Vol. 2018-Febuary). doi:10.1145/3159652.3160605

Chen, Z. Y., Fan, Z. P., & Sun, M. (2012). A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research*, *223*(2). doi:10.1016/j.ejor.2012.06.040

Dempster, A., Schmidt, D. F., & Webb, G. I. (2021). *MiniRocket: A Very Fast (Almost) Deterministic Transform for Time Series Classification*. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. doi:10.1145/3447548.3467231

Du, J., Zhang, X., Liu, P., Vong, C. M., & Wang, T. (2023). An Adaptive Deep Metric Learning Loss Function for Class-Imbalance Learning via Intraclass Diversity and Interclass Distillation. *IEEE Transactions on Neural Networks and Learning Systems*. doi:10.1109/TNNLS.2023.3286484

Hyland, S. L., Faltys, M., Hüser, M., Lyu, X., Gumbsch, T., Esteban, C., … Merz, T. M. (2020). Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature Medicine*, *26*(3). doi:10.1038/s41591-020-0789-4

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, *33*(4). doi:10.1007/s10618-019-00619-1

Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2017). LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access*, *6*. doi:10.1109/ACCESS.2017.2779939

Khan, M. R., Manoj, J., Singh, A., & Blumenstock, J. (2015). *Behavioral Modeling for Churn Prediction: Early Indicators and Accurate Predictors of Custom Defection and Loyalty*. In *Proceedings - 2015 IEEE International Congress on Big Data, BigData Congress 2015*. doi:10.1109/BigDataCongress.2015.107

Kumar, S., & D., C. (2016). A Survey on Customer Churn Prediction using Machine Learning Techniques. *International Journal of Computer Applications*, *154*(10). doi:10.5120/ijca2016912237

Lemmens, A., & Gupta, S. (2020). Managing churn to maximize profits. *Marketing Science*, *39*(5). doi:10.1287/mksc.2020.1229

Liu, H. Y., Gao, Z. Z., Wang, Z. H., & Deng, Y. H. (2022). Time Series Classification with Shapelet and Canonical Features. *Applied Sciences (Switzerland)*, *12*(17). doi:10.3390/app12178685

Mitkees, I. M. M., Badr, S. M., & Elseddawy, A. I. B. (2017). *Customer churn prediction model using data mining techniques*. In *ICENCO 2017 - 13th International Computer Engineering Conference: Boundless Smart Societies* (Vol. 2018-January). doi:10.1109/ICENCO.2017.8289798

Óskarsdóttir, M., Van Calster, T., Baesens, B., Lemahieu, W., & Vanthienen, J. (2018). Time series for early churn detection: Using similarity based classification for dynamic networks. *Expert Systems with Applications*, *106*. doi:10.1016/j.eswa.2018.04.003

Peng, K., Peng, Y., & Li, W. (2023). Research on customer churn prediction and model interpretability analysis. *PLoS ONE*, *18*(12 December). doi:10.1371/journal.pone.0289724

Rachid, A. D., Abdellah, A., Belaid, B., & Rachid, L. (2018). Clustering prediction techniques in defining and predicting customers defection: The case of e-commerce context. *International Journal of Electrical and Computer Engineering*, *8*(4). doi:10.11591/ijece.v8i4.pp2367-2383

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6). doi:10.1037/h0042519

Sağlam, M., & El MONTASER, S. (2021). *THE EFFECT OF CUSTOMER RELATIONSHIP MARKETING IN CUSTOMER RETENTION AND CUSTOMER ACQUISITION*. *International Journal of Commerce and Finance* (Vol. 7).

Schäfer, P. (2016). Scalable time series classification. *Data Mining and Knowledge Discovery*, *30*(5). doi:10.1007/s10618-015-0441-y

Tan, C. W., Dempster, A., Bergmeir, C., & Webb, G. I. (2022). MultiRocket: multiple pooling operators and transformations for fast and effective time series classification. *Data Mining and Knowledge Discovery*, *36*(5). doi:10.1007/s10618-022-00844-1

Theissler, A., Spinnato, F., Schlegel, U., & Guidotti, R. (2022). Explainable AI for Time Series Classification: A Review, Taxonomy and Research Directions. *IEEE Access*, *10*. doi:10.1109/ACCESS.2022.3207765

Tianyuan, Z., & Moro, S. (2021). *Research Trends in Customer Churn Prediction: A Data Mining Approach*. In *Advances in Intelligent Systems and Computing* (Vol. 1365 AIST). doi:10.1007/978-3-030-72657-7_22

Zhang, J., Fu, J., Zhang, C., Ke, X., & Hu, Z. (2016). *Not too late to identify potential churners: Early churn prediction in telecommunication industry*. In *Proceedings - 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, BDCAT 2016*. doi:10.1145/3006299.3006324