# A novel artificial intelligent-based approach for real time prediction of telecom customer's coming interaction

**Reyad Hussien[1], Mohamed Mahgoub[1], Shahenda Youssef[1], Ashraqat Torky[1], Nermin K. Negied[2,3]**
[1]School of Electrical and Computer Engineering, University of Ottawa, Ottawa, Canada
[2]School of Communication and Information Engineering, Zewail City of Science and Technology, Giza, Egypt
[3]School of Engineering and Applied Science, Nile University, Giza, Egypt

**Article Info**

***Article history:***

Received Oct 25, 2023
Revised Nov 10, 2023
Accepted Nov 14, 2023

***Keywords:***

Churn prediction
Convolutional neural networks
Gated recurrent units
Machine learning
Next interaction prediction
Statistical analysis
Time series analysis

**ABSTRACT**

Predicting customer's behavior is one of the great challenges and obstacles for business nowadays. Companies take advantage of identifying these future behaviors to optimize business outcomes and create more powerful marketing strategies. This work presents a novel real-time framework that can predict the customer's next interaction and the time of that interaction (when that interaction takes place). Furthermore, an extensive data exploratory analysis is performed to gain more insights from the data to identify the important features. Transactional data and static profile data are integrated to feed a deep learning model which is implemented using two methodologies: time-series approach and statistical approach. It is found that the time-series approach gives the best performance and fulfills all the requirements. The experiments show that the proposed framework introduces a good overall performance in comparison to existing approaches based on standard metrics like accuracy and mean absolute error (MAE) values. What makes the proposed work novel and special is that it is the first approach that addresses the telecom customer's next future interaction not just churn prediction like the other approaches in literature.



***Corresponding Author:***

Nermin K. Negied
School of Engineering and Applied Science, Nile University
Giza, Egypt
Email: nnegied@nu.edu.eg

## 1. INTRODUCTION

Predicting a customer's next interaction in real-time for companies, especially in telecommunications field, has significant impact on their revenue and sales. This will not only reduce the huge amount of money paid for advertisement but will also improve the customer's satisfaction by offering better alluring deals to their customers at the time of their next interaction, which could be recharge or bundle renewal, or cash transaction. The capability of predicting a customer's intended actions would help the companies to detect any problem with the system early if there is a huge gap between the predictions and the actual values of the recharge amounts.

Moreover, this can facilitate the discovery of what is trendy, interesting, and most popular for the different customers' segments. Furthermore, this can provide an interactive dashboard with peaks and troughs of recharges for the company sales department to analyze and extract useful insights from it. Forecasting customer interaction could be through their historical call logs, short message service (SMS) logs, billing logs, and bundle renewal logs. Moreover, this can facilitate gathering their reviews and ratings of customer services. Most related work done is specifically aiming at predicting customer churn or downgrade based on aggregating the historical data and feeding them into machine learning models. However, this research differs in two ways. First, the problem definition itself as we focus on predicting the next customer action rather than whether the customer would churn or not. Second, the way we deal with the problem. We did not use traditional machine learning models, we applied two different methodologies and in each one of them, we tried several different models. The rest of this paper is organized as follows: section 2 introduces a literature review for past related work. In section 3 describes the dataset used in this work. In section 4 demonstrates the data preparation steps conducted to get data ready to be fed to our model. A detailed explanation of the proposed approach is in section 5. Section 6 demonstrates the experimental work and discusses the results. Section 7 summarizes the evaluation of the proposed approach, and finally section 8 concludes the paper.

## 2. LITERATURE REVIEW

Through this research, we found that all state-of-the-art research related to customer of telecom companies next interaction focused on predicting customer churn. Many machine learning classifiers, regression models, and data mining techniques were used to predict customers churn to maximize revenue and decrease losses. This section reviews the most recent work done in this area. Tan *et al.* [1] proposed a blended approach in which they tested their approach on two public datasets for open online courses to predict students drop out and one private dataset provided by Adobe Creative Cloud and their target was to predict churn or dropping out over a target period. They incorporated three different types of users' data namely: i) the activity log, which is customer interaction with the given services; ii) dynamic data, which is like the activity log, yet it contains derivatives to customer settings like automatic renewal or cancellation and some personal profile data like birthday, and lastly; and iii) customer group and subscription date using their approach, they were able to capture the evolving intended action of users, their approach has proven better results over more classical approaches models such as logistic regression (LR), random forest (RF), Naive Bayes (NB), support vector machine (SVM), deep neural network (NNs), convolutional neural networks (CNNs) and long short-term memory (LSTM). A similar task to predicting the next interaction is predicting the next purchase.

Jong [2] aimed at predicting the customer's next purchase by feeding transactional data into machine learning algorithm. The algorithms applied during the research are multiverses, simple NN, LSTM, and LSTM with uncertainty cutoff (LUC). The predictions gained were then used to add a personalized banner in an e-mail with products the customer is more likely to buy. The results showed that this method did not make the customers buy the mentioned specific products. However, it increased the click rate in e-mails as well as improving the overall revenue. The limitation of this methodology was that only one test can be executed at a time. Another remarkable research done by Coullandreau [3] showed the use of sequential models to prevent customer churn and determine optimal marketing strategies, where the author used a dataset from Sparkify, an online mobile app that offers networking services to its users. The purpose of this work was to detect customer churn, the authors tested the data on three models: LR, gradient boosted tree (GBT), and RF and they reported that the GBT was the best best with an F1-score of 86.36%.

A year later Lu *et al.* [4] proposed a deep learning network to predict the stock price based on the data of daily stock prices from July 1, 1991, to August 31, 2020. Figure 1 shows the architecture of the proposed method. In their work, a CNN layer was used to extract the useful time series features of the previous 10 days. LSTM layer was used to predict the stock price with the extracted feature data. The paper then offered a detailed comparison between the proposed method and other forecasting methods used MLP, CNN, RNN, LSTM, and CNN-RNN models. The authors proved that the CNN-LSTM method is more suitable for forecasting stock prices with the highest prediction accuracy. Lastly, they presented a comparison of multiple models. Figure 2 shows the results of the different experiments. It can be observed that the best experiment with the lowest mean absolute error (MAE) was achieved by the CNN-LSTM method. Zhao *et al.* [5] used data mining algorithms to analyze the causes of customer churn and provide the telecommunication companies with information that can enhance the relationship between companies and customers to avoid churns. The authors applied LR over a high value customer operation in telecommunication industry in China. In the same year Zamil *et al.* [6] used LR again to detect the probability of customer churn and reached accuracy of 80%. Srinivasan *et al.* [7] used the popular machine learning algorithms like LR and RF to accurately identify the customers who are most likely to churn. The author worked on real customer records obtained from a telecommunication company. Later than that, Nirmal [8] used the minimax probability calculation to predict customer churn to maximize profits, and the author assumed that the approach achieved prediction accuracy of 97%. Afterwards, Suguna [9] used a hybrid model to tackle the customer churn problem. The author combined LR and decision tree (DT) to analyze customer churn statistics and the ability to prevent churns accomplished an accuracy of 80.93%, the dataset they used was downloaded from IBM sample knowledge sets for client retention programs.

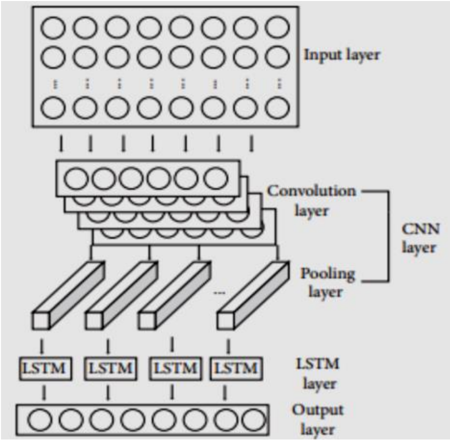


Figure 1. CNN-LSTM model architecture for stock price prediction [4]

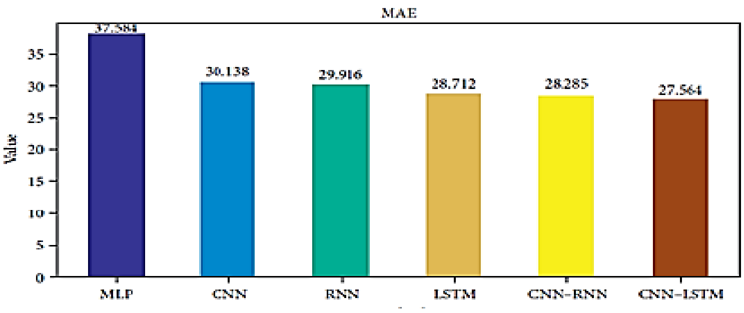


Figure 2. Stock price prediction MAE for different models [4]

Hybrid models were present again in the work of Awang *et al.* [10]. The authors performed a comparison between ensemble method combining 6 different learning algorithms and the prediction results obtained by single classifiers. They concluded that ensemble way achieves much better results compared to using single classifier. A different example from industry is the recent work done by Fareniuk *et al.* [11]. The author used LSTM to predict the next steps for current customers, assigning a probability to each step that would predict which helps in expecting customer churn and consequently prevent it. Fareniuk *et al.* [11] is 2022 used different data science and machine elarning models like k-nearest neighbor (KNN), NNs, and RF to classify the customers to loyal and likely to be churn for telecom companies using Ukrainian telecom company dataset and they achieved classification accuracy rate of 90%. Nguyen *et al.* [12] in the same year, used the kernel SVM to predict customer's churn for telecom companies and they achieved a classification accuracy rate of 98.9%. Shrestha and Shakya [13] in the same year also used XGBoost to address the same problem of telecom customer's churn prediction and they achieved an accuracy rate of 96%. Pebrianti *et al.* [14] used and compared eight different machine learning models including SVM, RF, LR, and XGBoost, and they confirmed that XGBoost obtained the best accuracy rate that reached 94%.

At the end of 2022, Peng [15] improved the XGBoost by integrating it with SHAP framework to analyze the important features affecting telecom customer churn. Saha *el al.* [16], used and tested different classification models including XGBoost, DT, RF, NNs, and CNN, and they confirmed that CNN outperformed all other models in their study where it achieved an accuracy rate of 99%. Bharambe *et al.* [17] also used XGBoost to address the same problem but they mentioned nothing factual about the accuracy of their classification model. Pandithurai *et al.* [18] in the same year, used the ensemble voting method after they combined four well-known classification algorithm which were SVM, RF, DT, and LR to get the classification rate of telecom customer churn, and the ir methodology succeeded to obtain a classification accuracy rate of 84%. Again in 2023, Zatonatska *et al.* [19] used and compared four different classification algorithms which were SVM, RF, LR, and XGBoost. Authors confirmed that they achieved an accuracy rate of 80% in telecom customer churn prediction. Babatunde *et al.* [20] used a hybrid approach which combined analysis of variance (ANOVA) with SVM to predict customer's churn. The authors confirmed that they successfully predicted future churns with an accuracy rate that reached 95%.

## 3. DATASET

Vodafone Egypt is the main data provider for this work. The Egyptian telecom company encrypted the data to keep the privacy of the clients and sent us the datafiles to work on this research. A detailed description of the data and the analysis performed on it are described clearly in the following subsections.

### 3.1. Dataset description

The dataset consists of multiple components called streams. A stream is a type of interaction done by customers and each stream contains the timestamp of the interaction along with features exclusive to this type of interaction. The received streams are internet, calls, SMS, recharges, and other services streams. A summary covering the number of records and number of customers for each stream is depicted in Table 1. Customers were traced among the streams and the full dataset to be used is constructed from multiple features contributed by different streams as follows:

- Recharge stream, contains recharge transactions and their respective features, and the used features are:
  i) Mobile number: the masked mobile number of customers.
  ii) Total recharge value: the value of the recharge.
  iii) Call recharge amount: the value of the recharge after removing any fees such as bundle renewals or subscribed services fees.
  iv) Recharge type: either micro recharge card which means a small amount, or main which means large amount.
  v) CI grouping: the category of the customer.
  vi) Card group: the category of the recharge card.
- Internet, calls, SMS, and other services streams share the same used feature which are mobile number, balance before interaction, balance after interaction and lastly the duration volume, where the duration volume has a different meaning for each stream, as follows:
  i) Internet: it is the number of megabytes per second for the session.
  ii) Calls, it is the number of minutes for each call.
  iii) SMS, it is the number of characters for each SMS.
  iv) Other services, it is the number of minutes for each service call.

Table 1. Summary of the received data

| | Internet stream | Call stream | SMS stream | Recharge stream | Other service |
|---|---|---|---|---|---|
| # of records | 627,418 | 776,500 | 771,895 | 12,888 | 59,748 |
| Unique customers | 1,332 | 1,777 | 1,938 | 1,535 | 952 |
| Common customers | | | 952 | | |
| Total unique customers | | | 2,680 | | |
| # of features used | 3 | 3 | 3 | 5 | 3 |

The dataset at the simplest form can be perceived as two parts; static data that is the same among all customers, and transnational data which resembles customer interactions such as recharges, calls, and internet usage. The data is then fed into a pipeline composed of three steps, namely: i) feature engineering, ii) data preprocessing, and iii) data transformation as shown in Figure 3. A further discussion can be found in section 4. Detailed explanations for the time-series approach and the statistical approach are discussed in subsections 5.1. and 5.2. respectively.

### 3.2. Exploratory analysis

Treating all customers as one cluster causes an increase in the number of anomalies and non-useful insights as shown in Figure 3. For that reason, the data was divided according to customers' categories provided by Vodafone as the Ci-grouping attribute [21]. Figure 4 demonstrates the number of recharges of all customer types.

One of the challenges faced us during this work was the data imbalance. The customers' categories were extremely imbalanced as postpaid business, control business, youth consumer, and prepaid business are severely under-sampled. On the other hand, customers categories like control consumers and prepaid consumers are oversampled as shown in Figure 5. Luckily, all recharge data have constant distributions that show insightful patterns and can provide useful information for the models, The recharge categories included in our dataset and analyzed in this work were recharge distribution for control business shown in Figure 6, prepaid business shown in Figure 7, postpaid business shown in Figure 8, and youth consumer shown in Figure 9.

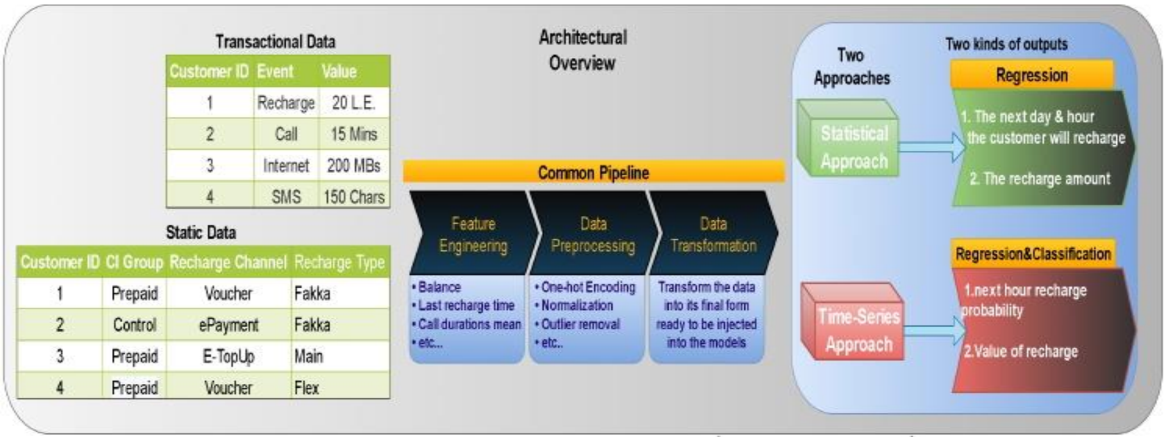


Figure 3. A full overview of the implemented solutions

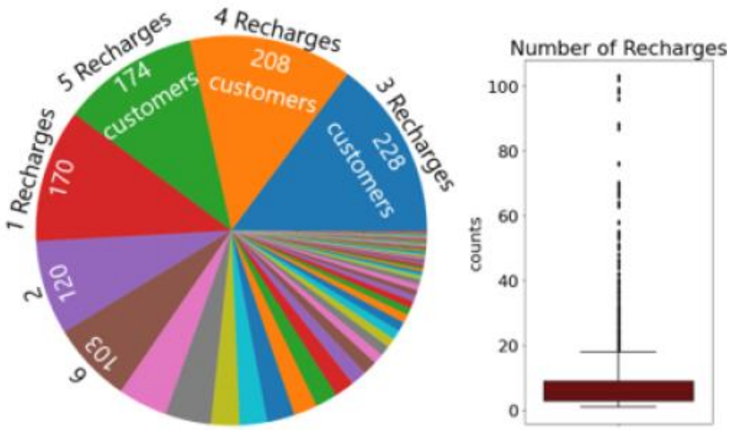


Figure 4. Number of recharges of all customers

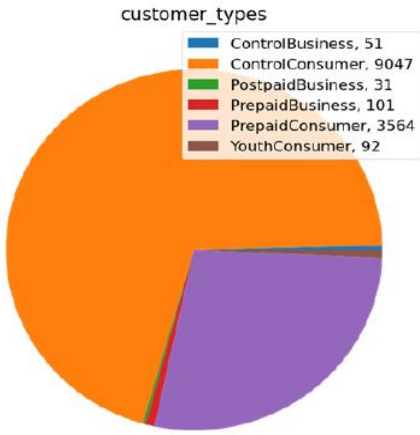


Figure 5. Customers category distribution

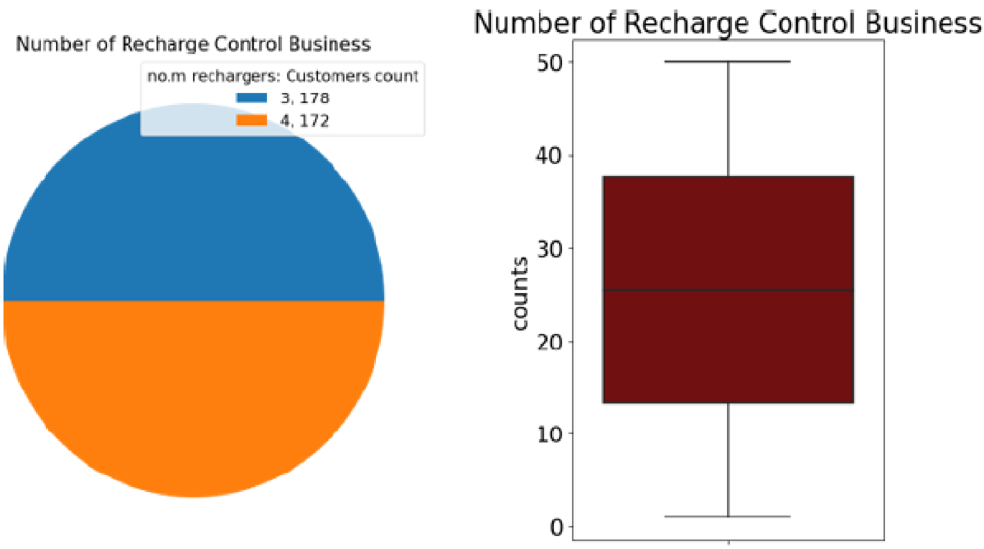


Figure 6. Number of recharges of control business customers

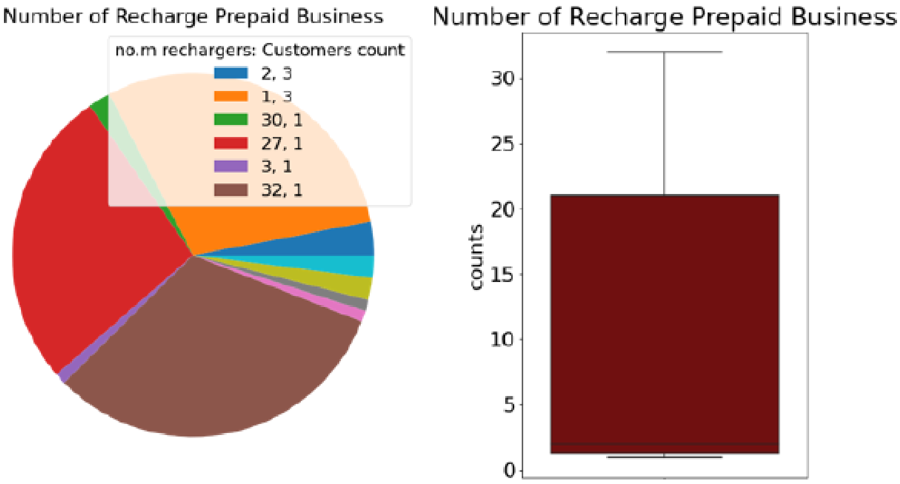


Figure 7. Number of recharges of prepaid business customers

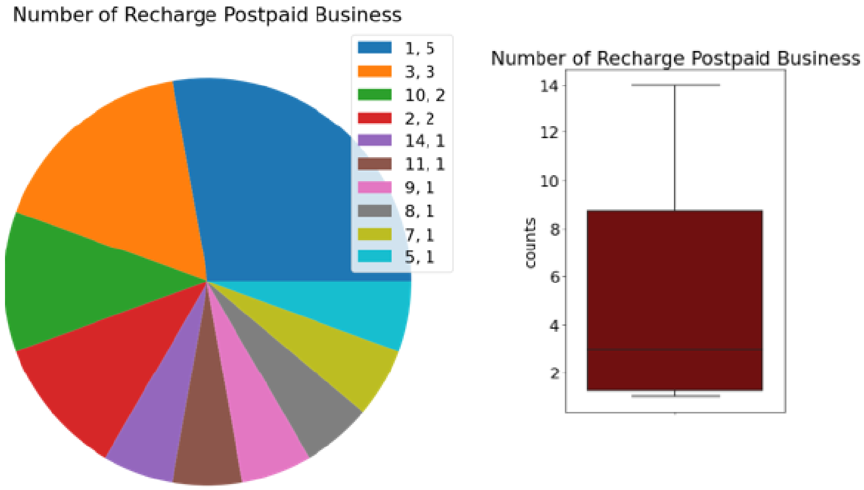


Figure 8. Number of recharges of postpaid business customers

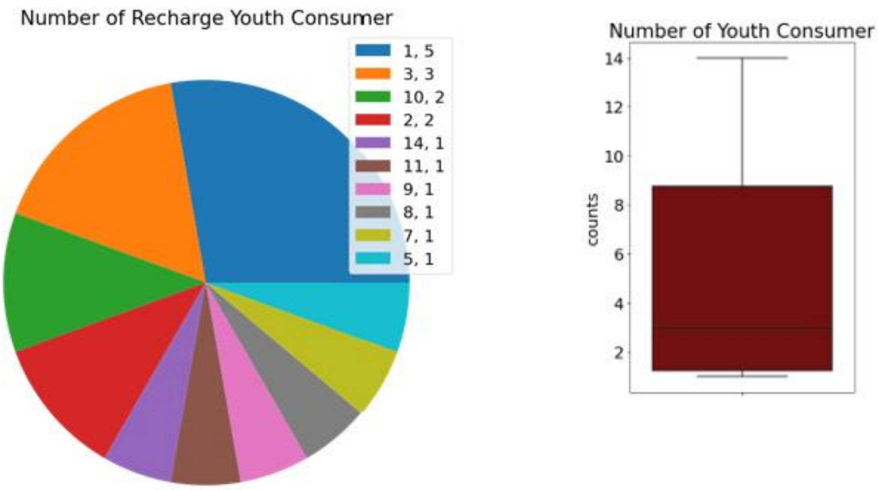


Figure 9. Number of recharges of youth consumer customers

On the contrary, the number of recharges for control consumers is shown in Figure 10. Figure 11 shows prepaid consumers. Prepaid consumers data has a great number of anomalies that were misleading and did not provide any useful insights.

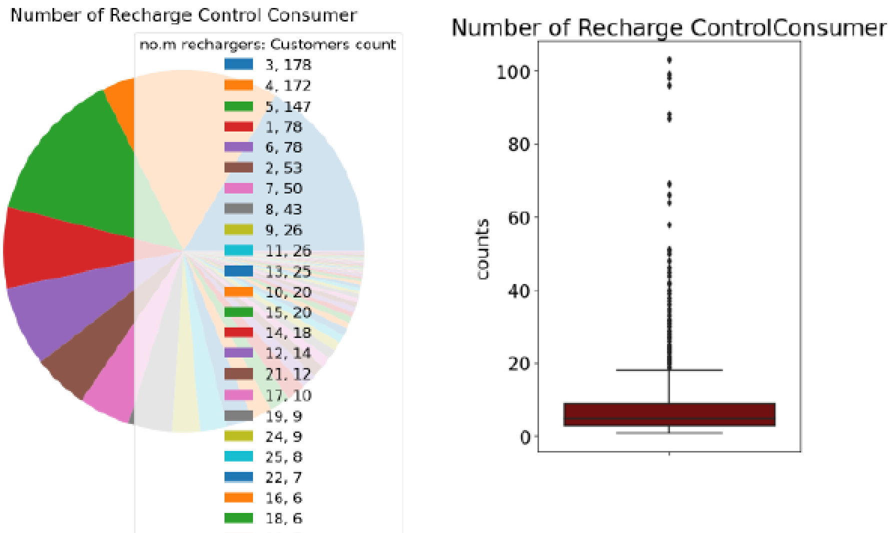


Figure 10. Number of recharges of control consumer customers

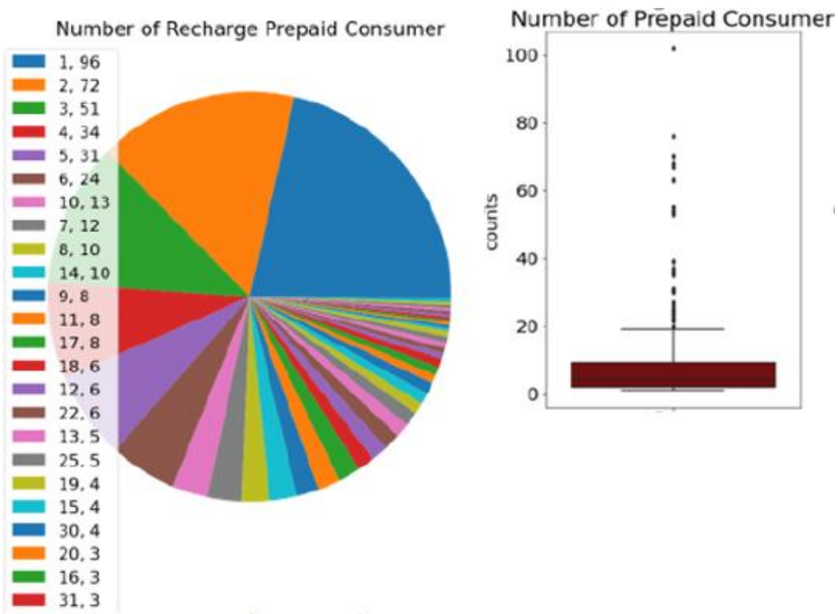


Figure 11. Number of recharges of prepaid consumer customers

By performing data analysis on call usage and internet usage, we noticed that the call usage is lower than internet usage by a great margin for all customers. The highest internet average usage among the customer's categories goes for control consumer with 902 megabytes, while the greatest call usage among the customer's categories goes for control business with 85.34 minutes as shown in Figure 12. Figure 13 also shows internet and calls usage distribution for all customers categories which shows that there is nearly zero median usage for youth consumers, postpaid business, and control business, the fact that indicates that customers in these categories do not have normal usage as others and need to be clustered. Finally, Figure 14 shows the expected relationship between the balance and the customers that recharged their accounts. Naturally, all customers recharge their accounts when it reaches zero or negative balance.

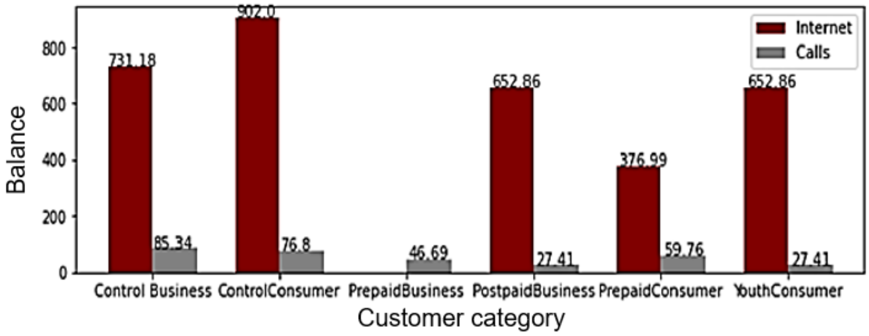


Figure 12. Average calls and internet for customers categories

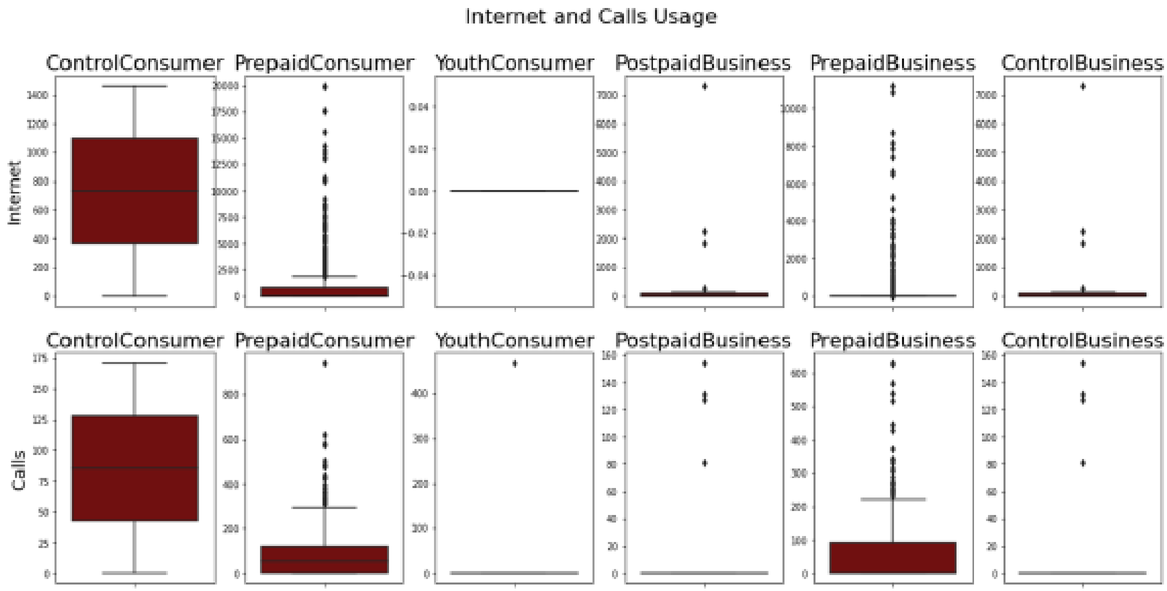


Figure 13. Distribution of calls and internet for customers categories

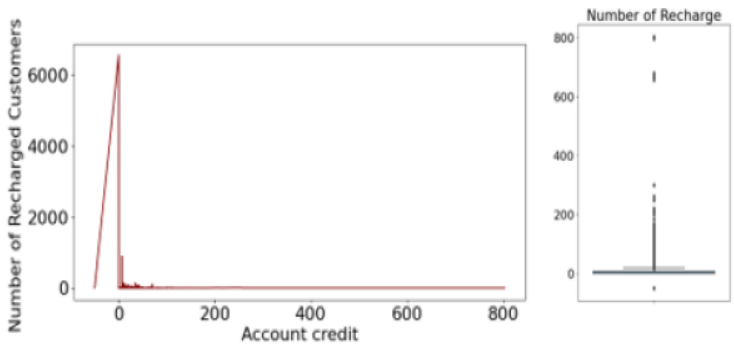


Figure 14. Relationship between customers balance and recharge

## 4. DATA PREPARATION

The dataset received from vodafone in its original form, can be perceived as two parts: static data that is the same among all customers, and transnational data which resembles customer interactions such as recharges, calls, and internet usage. The three steps pipeline implemented in this work to prepare data before feeding it to our model were: feature engineering, data preprocessing, and data transformation. Below is a detailed explanation of each step.

### 4.1. Feature engineering

The first data preparation step was adding new features that were not available in the original dataset. The added features included the following:

- Data balancing: which is crucial in predicting whether a customer is going to recharge or not. This feature is obtained by tracing customer interactions across the internet, calls, SMS, and other services streams and recording the before and after balance from each interaction.
- Last recharge time: given a certain time step, last recharge time is calculated as the number of hours, days, and months that passed since the last recharge has been made, this was obtained from the recharges stream.
- Number of frequently called mobile numbers for each customer: this was obtained from the calls stream.

### 4.2. Data preprocessing

Some classical data preprocessing techniques were used in this step like z-normalization [22], log transformation [23], outlier removal [24], one-hot encoding [25], and label encoding [26] to unlock the potential of the predictive models used in this work as shown in Figure 15. Another different technique was used for the time-series approach which may be called LogPlus and was introduced as a trainable layer for the deep learning model. The LogPlus was calculated as follows:

$$\text{LogPlus}\ (b; x) = \text{logSoftmax}\ (b)\ \text{Softmax}\ (x) \tag{1}$$

where $b$ is a trainable parameter and x is the input and SoftMax [27] is utilized to remove negative values. LogPlus proved to be useful in all experimental work as it increased both accuracy and training speed by giving the models the ability to determine the log transformation intensity.

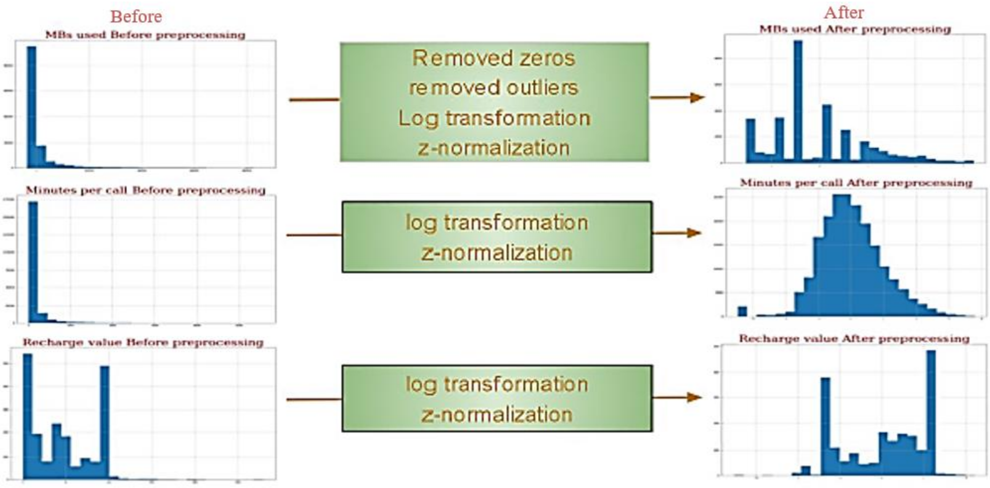


Figure 15. Before and after preprocessing histograms for three examples, as can be noticed the produced histograms are much more well distributed and gaussian like

### 4.3. Data transformation

The final step in data preparation is to convert all the data into a format that can be fed into the proposed models. For the statistical approach, the features are summarized into a set of numerical measures such as mean call duration, and number of recharges over the last month. For the time-series approach, each customer is perceived as a multi-variate time series, and the transformation is done by discretizing each variable by a time unit (i.e., hour, day, and week) and events happening within the same time step are aggregated into one value by any suitable method (i.e., mean, median, max, and min). It is worth mentioning that the same time series could be discretized by two different time intervals and hence two different representations. Afterwards, we take a few time samples for each customer, where the exact values of the variables are considered as labels and a time window preceding the sample is taken as the input features as depicted in Figure 16, where the red blocks resemble ones, and the grey blocks resemble zeros. Then, two snapshots are taken (labeled by the green and blue arrows) and their respective time windows (the green and blue Xs). Two main labels for this approach were adopted: the recharge event and the recharge value.

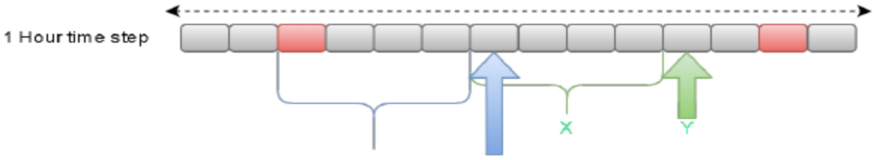


Figure 16. A uni-variate binary time-series for one customer is discretized by hour time unit

## 5. METHOD

Two approaches of solutions were developed in this work to predict customer next interaction. The approach is statistical i.e., it is based on summarizing features into a set of numerical measures [28]. The second approach treats the features as a more like time-series fashion with multiple variables [27]. The whole process is depicted in Figure 4.

### 5.1. Time series approach

In this approach, two deep learning models were used to tackle the prediction problem. The first one is built using gated recurrent units (GRUs) [29]–[32] and the second one is bulit using GRUs combined to CNNs [33] to produce a novel hybrid predictor. The architectures setup and the results of both approaches are demonstrated in the following subsections.

#### 5.1.1. GRU model

Here the GRU network is used as GRUs proved to be a very powerful predictor in the recent few years [34], [35]. The design of the main architecture of the GRU network proposed in this work was as follows: One input layer, one optimization layer, one layer of GRUs, and a fully connected dense layer. An overview of the proposed GRU model architecture is shown in Figure 17. Two inputs are fed to the model: i) time-series features and profile features. The time-series features input is followed by an optimization layer, a layer of 30 GRU, and a dense layer of 25 neurons and two outputs to predict the next hour recharge probability and the recharge amount; and ii) the profile features are fed to a fully connected dense layer of 10 neurons.

#### 5.1.2. CNN+GRU model

The idea of building such a hybrid model is to combine the advantages of CNNs which are characterized by their ability to extract useful knowledge from time-series data and GRU networks that are effective for capturing long sequence pattern information. The main structure is CNN and GRU, including input layer, one-dimensional convolution layer, a max-pooling, GRU layer, and a fully connected dense layer. An overview of the proposed CNN+GRU model architecture is shown in Figure 18. This model also has two inputs: time series features and profile features. The time-series features input is followed by an optimization layer, two one-dimensional convolution layers to extract features from the time series input, a max-pooling layer to reduce the high feature dimension, then a layer of 50 GRUs, and finally a dense layer of 30 neurons to predict the next hour recharge probability and the recharge amount by analyzing previous customer patterns. The profile features are fed to fully connected dense layers of 20 neurons.

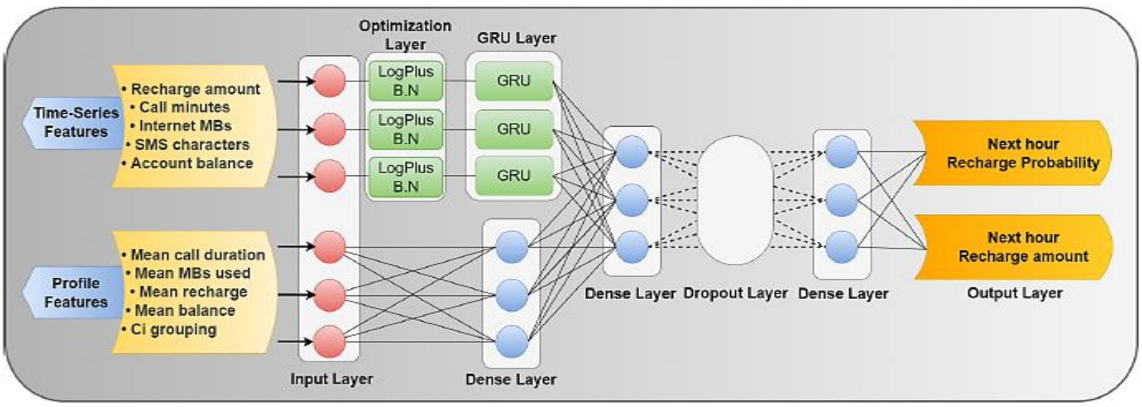


Figure 17. An overview of the proposed GRU model

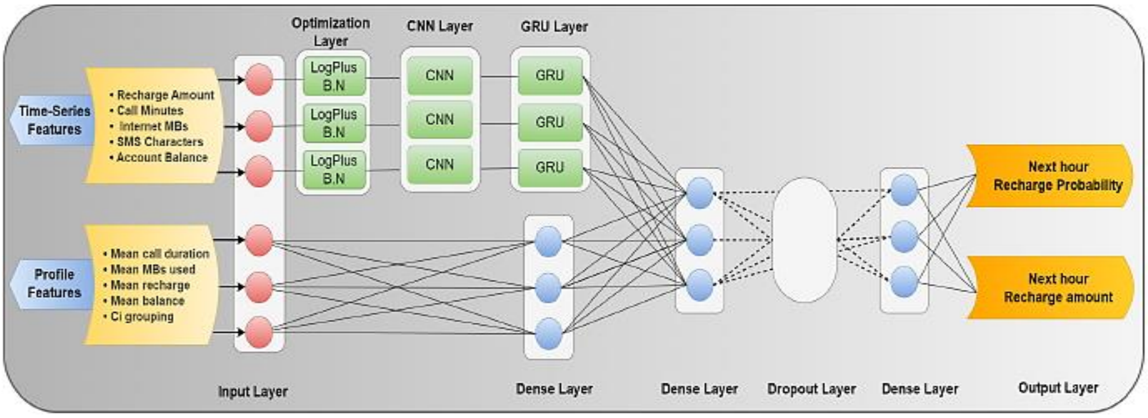


Figure 18. An overview of the proposed CNN+GRU model

### 5.2. Statistical Approach

In this approach, statistical features are used as inputs to a machine learning algorithm instead of using stream of data as in the first approach. Those statistical features are obtained either by aggregating the values for a certain period (3 months) and getting their average like average call minutes per time step, or by just using the last 5 values like last 5 recharge times (hour/day/month). After that, the machine learning algorithm takes these statistical features as inputs and generates four outputs: amount of recharge, hour of recharge, day of recharge, and month of recharge. Figure 19 shows the full input/output architecture for this approach.

To get the best result, six different machine learning algorithms were utilized in this work and their performances were compared. The algorithms used were: i) DT, ii) RF, iii) SVM, iv) XGBoost, v) AdaBoost, and vi) NN. out of the six algorithms used, RF showed the capability of obtaining the best results.

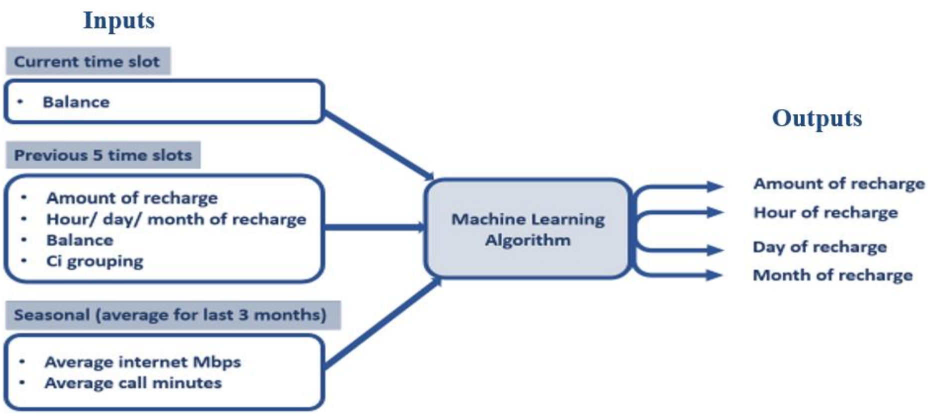


Figure 19. Statistical approach diagram

## 6. RESULTS AND DISCUSSION

Google Colaboratory (Colab) was used as a testing environment to train and test the proposed approaches. The proposed approaches proved to achieve results that allow to tell that this work presents the first successful system for next interaction pridction of telecom customer. A detailed explanation of settings and findings can be found in the following subsections.

### 6.1. Time-series approach

The dataset was divided into a training set, validation set, and testing set. To tune our model in the experiments, we have worked over a set of parameter values for each attribute:

– For activation function: Relu, LeakyRelu, and tanh.
– For the number of dense neurons: 5, 10, 20, and 30.
– For the number of GRU units: 20, 30, 50, and 100.

The GRU achieved an F1-score of 85%, the confusion matrix of the GRU model can be seen in Figure 20. The results of the GRU+CNN hybrid approach is shown in Figure 21, which shows the confusion matrix of the GRU+CNN hybrid model with an F1-score of 83%. Table 2 represents a comparison between the two time series approach models using precision, recall, F1-score, accuracy, and MAE. The results indicate that the performance of the GRU model is better than the performance of the hybrid approach CNN+GRU.

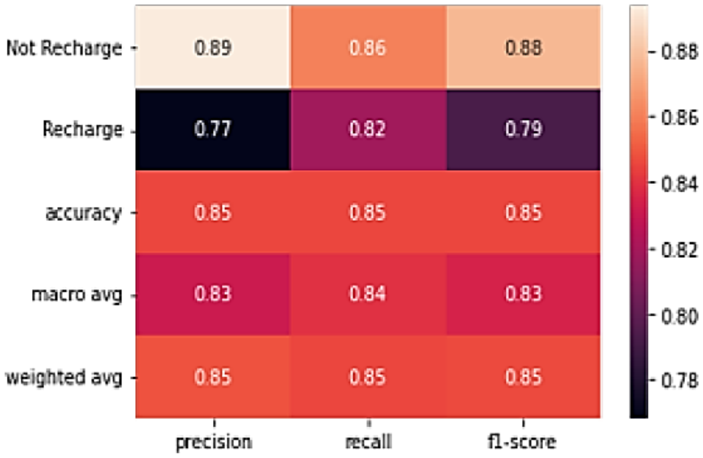


Figure 20. Confusion matrix of GRU results

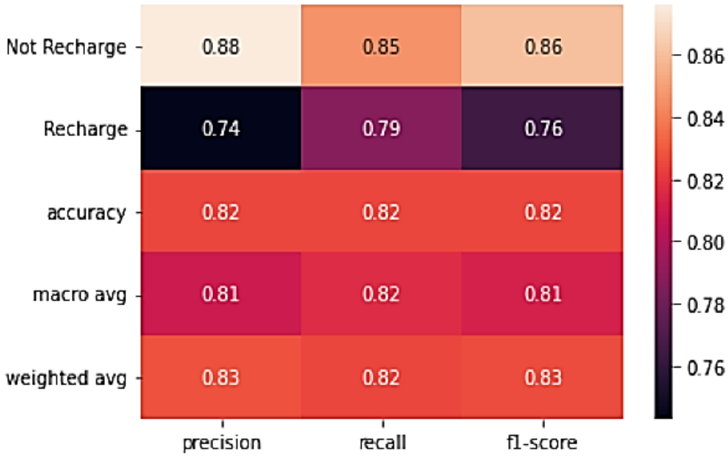


Figure 21. Confusion matrix of CNN + GRU results

Table 2. Comparison between time-series approaches

| Metric | GRU model | CNN+GRU model |
|---|---|---|
| Precision | 0.85 | 0.83 |
| Recall | 0.85 | 0.82 |
| F1-score | 0.85 | 0.83 |
| Accuracy | 0.85 | 0.82 |
| MAE | 7.32 | 7.4 |

### 6.2. Statistical approach

Six different algorithms were implemented and tested in this study as mentioned before. The RF and the SVM proved to be the best most of the time. The following subsections compare the results obtained by the sizx algortihms in detail.

### 6.2.1. Comparison between the used machine learning algorithms

The six different algorithms used in this study are: DT, RF, SVM, XGBoost, AdaBoost, and NN. The MAE of the predicted amount, the predicted hour, the predicted day, and the predicted month can be seen in Figures 22 to 25.

– Predicted amount: the best experiment with the lowest MAE was achieved by the SVM model which scored 14.48. Followed by NN model and RF with mean absolute error of 17 and 18.08 respectively.
– Predicted hour: the best experiment with the lowest MAE was achieved by the RF model which scored 4.28. Then the SVM model with mean absolute error of 4.33 and Adaboost with mean absolute error of 4.66.
– Predicted day: the best experiment with the lowest MAE was achieved by the RF model which scored 5.27. Furthermore, the second-best performing model was the SVM with a MAE of 5.43. Finally, the third best performing model was the Adaboost with a mean absolute error of 5.5.
– Predicted month: the best experiment with the lowest MAE was achieved by the SVM and NN models 0.14 followed by the Adaboost model with a very minor difference of 0.01 and followed by the RF with a difference of 0.02.

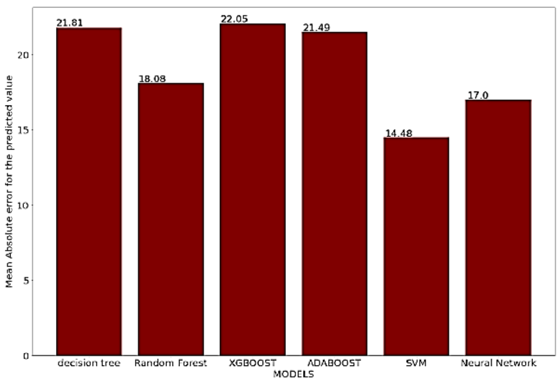


Figure 22. Mean absolute error for predicted amount

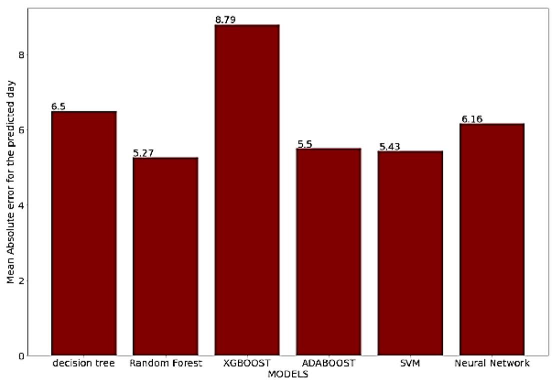


Figure 23. Mean absolute error for predicted hour

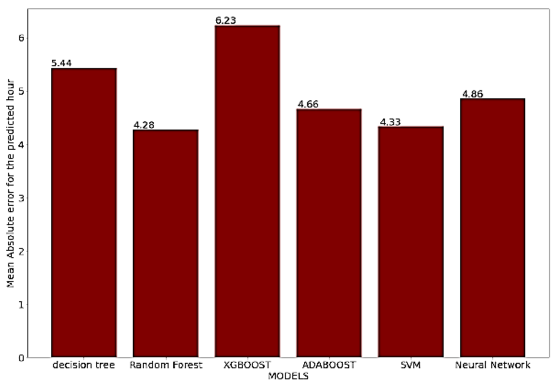


Figure 24. Mean absolute error for predicted day

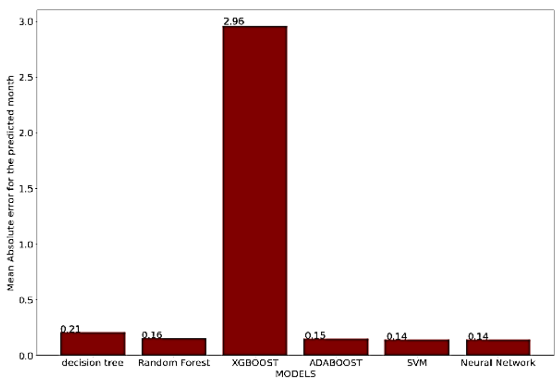


Figure 25. Mean absolute error for predicted month

### 6.2.2. True vs predicted values for best model

Figures 26 to 29 demonstrate the actual values versus the predicted values on a sample of customers. Figure 26 shows the predicted re-charge amount versus the actual amounts. As can be noticed from the figure, the system predictions are very close to the actual values. The figures after that demonstrate the predictions obtained by the proposed work in different time frames. The hourly predictions versus actual is shown in Figure 27, the daily predictions versus actual is shown in Figure 28, and the monthly predictions versus actual is shown in Figure 29. As can be shown in the figures, the longer the time frame, the better the predicted results, but overall, we can say that the proposed work succeeded in obtaining very good predictions that are very close to the actual values. Note: the actual values are represented using gray bars, and the predicted values are represented using magenta bars.

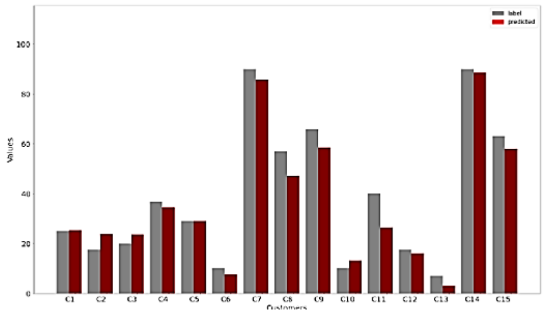

Figure 26. Predicted recharge amount vs actual amount for a sample of customers

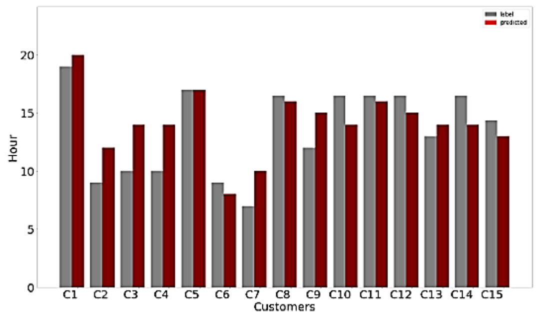


Figure 27. Predicted hour vs actual hour for a sample of customers

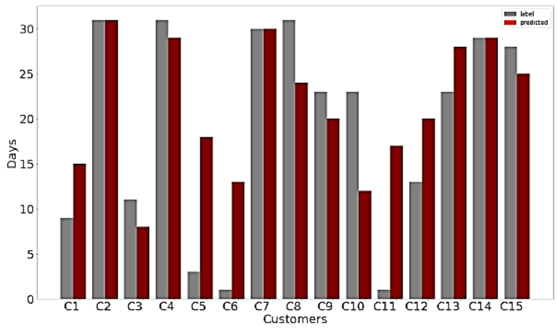


Figure 28. Predicted day vs actual day for a sample of customers

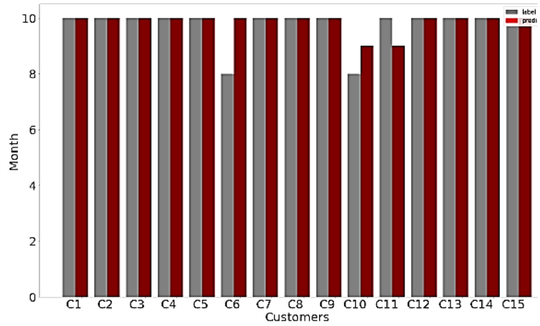


Figure 29. Predicted month vs actual month for a sample of customers

#### 6.2.3. Evaluation

Some analyses are conducted in this study to evaluate results. For each customer, a five-day period is cut out from the original dataset as an evaluation period. The start and end times of this period are not constant for all customers as it begins before the last recharge date for a given customer by two days and ends after five days. Only 500 customers are chosen for the evaluation task and only the time-series approach with pure GRUs is tested. The reason for choosing only 500 customers is that this task is computationally expensive and requires predicting every hour for the five days for each customer, and the reason for choosing the time-series approach over the statistical approach is because it was worse in requirements fulfillment. All requirements were fulfilled for both approaches, except for the fifth requirement (see Table 3). The statistical approach is hand crafted for this problem set and cannot be easily generalized to similar problems of transnational data. The time-series approach has achieved a mean accuracy of 84.5% and a mean absolute error for value prediction of 5.05.

Table 3. Requirements fulfillment for each approach

| Req. Number/approach | Time-series approach | Statistical approach |
|---|---|---|
| Recharge in the next timestep | Fulfilled | Fulfilled |
| Daily prediction of potential interaction | Fulfilled and also predicts hourly | Fulfilled and also predicts hourly |
| Real-time inference and the ability to show results almost instantly | Inference time is 0.8 seconds for a batch of 512 data points | Inference time is 26 seconds for a batch of 512 data points |
| Implementation on Spark for distributed system | Fulfilled | Fulfilled |
| Generic solution and can be applied to different problems of similar nature | Fulfilled | Solution is crafted only for this problem |

## 7. CONCLUSION AND FUTURE WORK

This work represents the first system to address the problem of predicting telecommunication customers' next interaction, as in literature all research done in this area was to predict the customer churn only. In this work a lot of data analysis was conducted to analyze telecom company customer's data. The paper also proposed two different approaches to predict the coming customer's interaction and the expected time of this interaction. One approach uses statistical features as inputs to a machine learning algorithm, and the other is a time-series based approach that uses deep learning networks such as CNNs and GRUs. Experiments show that both approaches can fulfill all the requirements needed by a telecom company to perform the task, but the statistical based approach can only be applied for that specific problem and cannot be used for different problems of the same nature. Therefore, the experiment proceeded with the time-series methodology which achieved an accuracy rate of 84.5% and MAE of 5.05 for value of prediction of customer’s next interation. Overall, the proposed methodologies succeeded in predicting the next recharge amount, as the results were very close to the actual recharge amount. The proposed methodologies also succeeded in predicting the next interaction on a monthly, daily, and even hourly basis. It is worth mentioning that the proposed approach also succeeded to fulfill all the telecom company requirements which are: predicting the recharge in the next time step, daily prediction of potential interaction, real-time inference, implmenetation on Spark for distributed systems, and applicability of the proposed approach for different problems.

Future work will focus on three main aspects. First, evaluating the proposed approaches using a larger number of customers. Second, predicting more types of events such as promo redemption, bundle renewal, and call center call. Third, deployment of the machine learning models proposed in this work on a distributed system to validate the performance of the proposed framework in a real production environment.

## ACKNOWLEDGMENT

The assistance, datasets, and technical support provided by Vodafone Egypt Company, espacially Dr. Rania El Badry is greatly appreciated by all of us.

## REFERENCES


[1] F. Tan *et al.*, “A blended deep learning approach for predicting user intended actions,” in *Proceedings - IEEE International Conference on Data Mining, ICDM*, Nov. 2018, vol. 2018-November, pp. 487–496, doi: 10.1109/ICDM.2018.00064.

[2] B. de Jong, “Next purchase prediction: improving customer interaction by using transactional data”, *Universiteit van Amsterdam*, 2019. [Online]. Available: https://scripties.uba.uva.nl/search?id=691699;setlang=en.

[3] A. Coullandreau, “Sparkify churn prediction, or the power of timeseries data,” 2019. https://towardsdatascience.com/sparkify-churn-prediction-or-the-power-of-time-series-data-7b2e8f49a30e.

[4] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, “A CNN-LSTM-based model to forecast stock prices,” *Complexity*, vol. 2020, pp. 1–10, Nov. 2020, doi: 10.1155/2020/6622927.

[5] M. Zhao, Q. Zeng, M. Chang, Q. Tong, and J. Su, “A prediction model of customer churn considering customer value: an empirical research of telecom industry in China,” *Discrete Dynamics in Nature and Society*, vol. 2021, pp. 1–12, Aug. 2021, doi: 10.1155/2021/7160527.

[6] A. M. A. Zamil, N. M. Nusairat, T. G. Vasista, M. M. Shammot, and A. Y. Areiqat, “Predicting telecom customer churn: an example of implementing in python,” *Journal of Legal, Ethical and Regulatory Issues*, vol. 24, no. Special Issue 1, pp. 1–11, 2021.

[7] R. Srinivasan, D. Rajeswari, and G. Elangovan, “Customer churn prediction using machine learning approaches,” in *Proceedings of the International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering, ICECONF 2023*, Jan. 2023, pp. 1–6, doi: 10.1109/ICECONF57129.2023.10083813.

[8] V. J. Nirmal, “Minimax probability-based churn prediction for profit maximization,” *Ictact Journal on Soft Computing*, p. 1, 2021, doi: 10.21917/ijsc.2021.0343.

[9] G. Suguna, “IRJET- churn analysis for maximizing profit using statistical model in telecom industry,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 8, no. 4, p. undefined-8, 2021.

[10] M. K. Awang, M. Makhtar, N. Udin, and N. F. Mansor, “Improving customer churn classification with ensemble stacking method,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, pp. 277–285, 2021, doi: 10.14569/IJACSA.2021.0121132.

[11] Y. Fareniuk, T. Zatonatska, O. Dluhopolskyi, and O. Kovalenko, “Customer churn prediction model: a case of the telecommunication market,” *ECONOMICS - Innovative and Economics Research Journal*, vol. 10, no. 2, pp. 109–130, Dec. 2022, doi: 10.2478/eoik-2022-0021.

[12] N. Y. Nguyen, V. L. Tran ,D. V. T. Son, “Churn prediction in telecommunication industry using kernel support vector machines,” *PLOS ONE*, vol. 17, no. 5, p. e0267935, May 2022, doi: 10.1371/journal.pone.0267935.

[13] S. M. Shrestha and A. Shakya, “A customer churn prediction model using XGBoost for the telecommunication industry in Nepal,” *Procedia Computer Science*, vol. 215, pp. 652–661, 2022, doi: 10.1016/j.procs.2022.12.067.

[14] D. Pebrianti, D. D. Istinabiyah, L. Bayuaji, and Rusdah, “Hybrid method for churn prediction model in the case of telecommunication companies,” in *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Oct. 2022, vol. 2022-October, pp. 161–166, doi: 10.23919/EECSI56542.2022.9946535.

[15] K. Peng and Y. Peng, “Research on telecom customer churn prediction based on GA-XGBoost and SHAP,” *Journal of Computer and Communications*, vol. 10, no. 11, pp. 107–120, 2022, doi: 10.4236/jcc.2022.1011008.

[16] L. Saha, H. K. Tripathy, T. Gaber, H. El-Gohary, and E. S. M. El-kenawy, “Deep churn prediction method for telecommunication industry,” *Sustainability (Switzerland)*, vol. 15, no. 5, p. 4543, Mar. 2023, doi: 10.3390/su15054543.

[17] Y. Bharambe, P. Deshmukh, P. Karanjawane, D. Chaudhari, and N. M. Ranjan, “‘Churn prediction in telecommunication industry,’” in *2023 International Conference for Advancement in Technology (ICONAT)*, Jan. 2023, pp. 1–5, doi: 10.1109/ICONAT57137.2023.10080425.

[18] O. Pandithurai, S. B, H. Narayan S, and H. Ahmed H, “Telecom churn prediction using voting classifier ensemble method and supervised machine learning techniques,” *ITM Web of Conferences*, vol. 56, p. 05012, Aug. 2023, doi: 10.1051/itmconf/20235605012.

[19] T. Zatonatska, Y. Fareniuk, and V. Shpyrko, “Churn rate modeling for telecommunication operators using data science methods,” *Marketing and Management of Innovations*, vol. 14, no. 2, pp. 163–173, 2023, doi: 10.21272/mmi.2023.2-15.

[20] R. Babatunde, S. O. Abdulsalam, O. A. Abdulsalam, and M. O. Arowolo, “Classification of customer churn prediction model for telecommunication industry using analysis of variance,” *IAES International Journal of Artificial Intelligence*, vol. 12, no. 3, pp. 1323–1329, Sep. 2023, doi: 10.11591/ijai.v12.i3.pp1323-1329.

[21] T. Dobson, M. Muzychuk, and P. Spiga, “Generalised dihedral CI-groups,” *Ars Mathematica Contemporanea*, vol. 22, no. 2, p. #P2.07, May 2022, doi: 10.26493/1855-3974.2443.02e.

[22] D. A. Anggoro and W. Supriyanti, “Improving accuracy By applying Z-score normalization in linear regression and polynomial regression model for real estate data,” *International Journal of Emerging Trends in Engineering Research*, vol. 7, no. 11, pp. 549–555, Nov. 2019, doi: 10.30534/ijeter/2019/247112019.

[23] P. Mooijman, C. Catal, B. Tekinerdogan, A. Lommen, and M. Blokland, “The effects of data balancing approaches: a case study,” *Applied Soft Computing*, vol. 132, p. 109853, Jan. 2023, doi: 10.1016/j.asoc.2022.109853.

[24] J. D. Karch, “Outliers May not be automatically removed,” *Journal of Experimental Psychology: General*, vol. 152, no. 6, pp. 1735–1753, Jun. 2023, doi: 10.1037/xge0001357.

[25] I. Ul Haq, I. Gondal, P. Vamplew, and S. Brown, “Categorical features transformation with compact one-hot encoder for fraud detection in distributed environment,” in *Communications in Computer and Information Science*, vol. 996, 2019, pp. 69–80.

[26] Z. Xiao, B. Xu, Y. Zhang, K. Wang, Q. Wan, and X. Tan, “Aspect ratio-based bidirectional label encoding for square-like rotation detection,” *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3247027.

[27] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[28] J.~S.~Bridle, “Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters,” *Advances in Neural Information Processing Systems*, vol. 2, pp. 211–217, 1990, doi: 10.5555/2969830.2969856.

[29] C. Chen, L. Xue, and W. Xing, “Research on improved GRU-based stock price prediction method,” *Applied Sciences (Switzerland)*, vol. 13, no. 15, p. 8813, Jul. 2023, doi: 10.3390/app13158813.

[30] K. Cho *et al.*, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724–1734, doi: 10.3115/v1/D14-1179.

[31] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, Mar. 1994, doi: 10.1109/72.279181.

[32] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[33] A. A. Rahman, S. S. Prasetiyowati, and Y. Sibaroni, “Performance analysis of the imbalanced data method on increasing the classification accuracy of the machine learning hybrid method,” *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 1, pp. 115–126, Feb. 2023, doi: 10.29100/jipi.v8i1.3286.

[34] A. Kumar, D. K. M, V. D. J. Macedo, B. R. Mohan, and A. P. N, “Machine learning approach for prediction of the online user intention for a product purchase,” *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 1s, pp. 43–51, 2023, doi: 10.17762/ijritcc.v11i1s.5992.

[35] A. Alqatawna, B. Abu-Salih, N. Obeid, and M. Almiani, “Incorporating time-series forecasting techniques to predict logistics companies’ staffing needs and order volume,” *Computation*, vol. 11, no. 7, p. 141, Jul. 2023, doi: 10.3390/computation11070141.


## BIOGRAPHIES OF AUTHORS

**Reyad Hussien** received the B.E. degree from Alexandria University, Egypt in 2020, and M.Eng. degree from University of Ottawa in 2022. He is currently a senior data scientist at Digital Egypt for Investment and a senior data scientist for a company called Vlense. He can be contacted at email: rmeli007@uottawa.ca.

**Mohamed Mahgoub** received the B.E. degree from Ain Shams University, Egypt in 2020, and M.Eng. degree from University of Ottawa in 2022. He is currently a teaching assistant at Digital Egypt Builders Initiative (DEBI) and a lead programmer at a stealth conversational AI startup. Previously he co-founded CGEagleTeam game studio which developed Inheritors2078 in 2021. He can be contacted at email: mmahg092@uottawa.ca.

**Shahenda Youssef** received the M.Eng. degree in Electrical and Computer Engineering from the University of Ottawa, Canada, in March 2022. She got her B.Sc. degree in Computer and Communications Engineering from the Faculty of Engineering at Alexandria University, Egypt, in May 2019. She is a Teaching Assistant at Queen's University, Canada. She is a senior Machine Learning Engineer in the Research and Development department at a conversational AI Company in Europe. Her research interests include machine learning, deep learning, computer vision, and natural language processing. She can be contacted at email: syous071@uottawa.ca.

**Ashraqat Torky** received the B.Sc. degree in communication and electronics engineering from Alexandria University, Alexandria, Egypt, in 2020 and the M.Eng. degree in electrical and computer engineering from University of Ottawa, Ontario, Canada, in 2022. From 2022, she was a teaching assistant at Queen's University, Kingston, Canada for a master's degree program. The courses she teaches include data mining, data analytics, and cloud computing. She can be contacted at email: atork095@uottawa.ca.

**Dr. Nermin K. Negied** is the Head of Data Science and Artificial Intelligence Track, at Digital Egypt Builders Initiative (DEBI), Ministry of Communication and Information Technology (MCIT). Nermin is an assistant professor at Cairo University and Zewail City of Science and Technology, Faculty of Engineering. She is a former assistant professor at Nile University, Arab Academy for Science and Technology and Maritime Transport (AASTMT), and October University for Modern Science and Arts (MSA). She worked as the Educational Quality Manager at October University for Modern Science and Arts (MSA), Faculty of Computer Science. Nermin was a lecturer at 6th of October University, Faculty of Engineering, Computer Engineering department from September 2012 till September 2015. She was a Teaching Assistant at the same place, September 2006 till September 2012. She got her Ph.D. degree in July 2016, from Cairo University, Faculty of Engineering, Computer Department. She got her M.Sc. degree in February 2012, from the same place. Nermin published many international journals and conference papers and shared in reviewing many scientific papers. Her research interests include image processing and computer vision, machine learning, artificial intelligence, expert systems, natural language processing, and genetic algorithms. She can be contacted at email: nnegied@nu.edu.eg.