International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Top-up Frequency Forecasting for Lao Telecommunication Company Using Time-Series Analysis

Soulixay Phomvongxay[1], Phouthone Vongpasith[2], Chitnavanh Phonekhamma[3], Bounmy Phanthavong[4]

[1,2,3,4]Department of Computer Science, Faculty of Natural Sciences, National University of Laos
Vientiane Capital, Lao P.D.R

Abstract
The study aimed to forecast the top-up frequency for Lao Telecommunication Company (LTC) customers using time-series analysis. Accurate forecasts help the company optimize resource allocation, improve customer service, and boost revenue by anticipating customer demand. The time-series analysis techniques are employed, such as autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average + exogenous variables (SARIMAX) and exponential smoothing (ETS) models, to analyze historical top-up data. The data is pre-processed, identified patterns, and trained the model to forecast future top-up frequencies. The model's predictions closely aligned with the actual data, indicating its effectiveness. The ARIMA model produced reliable forecasts for short-term top-up frequencies, showing the best performance overall with the lowest RMSE and MAE, and a positive R-squared indicating some ability to explain the data. SARIMAX and EST both perform poorly, with very high RMSE and MAE, and negative R-squared values. The results demonstrated that time-series analysis could enhance decision-making in the telecommunications industry by enabling better forecasting of customer behavior. This helps in optimizing marketing strategies, inventory management, and service availability.

Keywords: Time Series, ARMA, SARIMAX, EST

1. Introduction

The telecommunication industry in Laos has been growing rapidly, driven by increasing mobile phone usage [1], internet penetration, and government efforts to improve digital infrastructure. Major players, such as Lao Telecommunication Company (LTC), Unitel, and ETL, provide a range of services [2], including voice, data, and mobile financial services. With a significant portion of the population relying on prepaid mobile services, top-ups or recharges are essential for maintaining connectivity and access to services [3]. Top-up frequency is a critical metric for telecommunication companies in Laos [4]. It reflects customer engagement, service usage, and revenue generation. For prepaid customers, regular top-ups enable continuous access to services such as calls, SMS, and mobile internet. By understanding and predicting top-up behavior, companies can better manage resources, plan marketing campaigns, and enhance customer satisfaction. Additionally, accurate forecasts of top-up frequency help optimize network capacity, ensuring smooth service and improving operational efficiency [5]. Forecasting top-up behavior helps LTC design targeted marketing campaigns [6]. Predicting top-up frequency allows LTC to identify patterns of customer behavior, including when users are likely to reduce or stop topping up. By predicting future demand, LTC can optimize staffing and operational resources to meet customer needs effectively, reducing costs and improving service delivery. Accurate forecasting helps the company anticipate cash flow, manage resources, and optimize pricing strategies to maximize revenue. This enables the company to implement targeted retention campaigns, such as offering promotional deals or discounts, to prevent customer churn. In short, forecasting top-up frequency enhances LTC's ability to manage its business more effectively, improve customer experience, and sustain long-term growth in the competitive telecommunications market.

Time series forecasting is a technique for predicting future events by analyzing past trends, checking for patterns of time decomposition, such as trends, seasonal and cyclical patterns and regularity [7]. It involves building models through historical analysis and using them to make observations and drive future strategic decision-making, with the assumption that future trends will hold similar to historical trends. The goal of time series forecasting is to predict a future value or classification at a particular point in time [8].

In this paper we are going to study the frequency top up of the different mobile voucher cards, in Telecommunication mass company, between the 1st of April and the 30nd of June of 2024 and try to forecast the frequency top up between the 1st and the 11th of July of 2024, while comparing with the actual frequency. In this project we are going to study a total of 3 machine learning models prepared to work with time series, such as auto-regressive integrated moving average (ARIMA) [9] and seasonal autoregressive integrated moving average + exogenous variables (SARIMAX) [10] and exponential smoothing (ETS) models [11].

2. Literature Review

Time-series forecasting is based on analyzing historical data points collected at regular time intervals to predict future values [12]. The key concept underlying time-series analysis is that past behavior and patterns are likely to repeat, which enables predictions of future trends. A stationary time series has a constant mean, variance, and autocorrelation over time [13]. Many forecasting models assume stationarity, and techniques like differencing and seasonal decomposition can be used to achieve it. Several methodologies are commonly applied in time-series forecasting, each suited to different types of data and forecasting needs. ARIMA [9] is widely used for univariate time-series forecasting, focusing on short-term, non-seasonal data. ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. SARIMA [10] is applied when the data has clear seasonal patterns. SARIMA, which stands for Seasonal Autoregressive Integrated Moving Average, is a versatile and widely used time series forecasting model. It's an extension of the non-seasonal ARIMA model, designed to handle data with seasonal patterns. SARIMA captures both short-term and long-term dependencies within the data, making it a robust tool for forecasting. ETS [11] is a commonly-used local statistical algorithm for time-series forecasting. The ETS algorithm is especially useful for datasets with seasonality and other prior assumptions about the data. ETS computes a weighted average over all observations in the input time series dataset as its prediction. In summary, time-series forecasting is based on detecting patterns like trends and seasonality within historical data. A variety of models, ranging from simple ARIMA to SARIMA and ETS [14], are used to capture these patterns, depending on the complexity and nature of the data.

Research on top-up frequency forecasting and time-series analysis in telecommunications and similar industries primarily focuses on predicting customer behavior, optimizing resource allocation, and enhancing business operations [15]. Below is a review of some key studies and approaches that are related to this field. Several studies have explored the use of ARIMA models for forecasting customer behavior in prepaid mobile services, particularly top-up frequency [16]. These models are used to predict the likelihood of customers recharging their prepaid accounts based on historical data. However, its limitations become apparent when dealing with more complex patterns or seasonal fluctuations in customer activity. Accurate predictions can help telecommunications companies optimize their marketing efforts and ensure service availability for prepaid customers by predicting when and how often customers will top-up [17]. Existing research on top-up frequency forecasting highlights the increasing importance of time-series analysis in the telecommunications industry. The methods like ARIMA, SARIMA and ETS remain effective for simple patterns and gain popularity due to their ability to handle nonlinear trends in customer behavior.

3. Data Collection

The dataset, provided by the telecommunication mass company, it's divided in six different kinds vouchers, time series voucher_5000, voucher_10000, voucher_20000, voucher_25000, voucher_50000 and voucher_100000, each one consisting in frequency top up of a certain product in a specified period of time. For this problem we are only going to use the data in the time series total of voucher_5000, voucher_10000, voucher_20000, voucher_25000, voucher_50000 and voucher_100000. In Fig. 1, we can see the different vouchers cards per time stamp for the telecommunication mass company, we can see that there are 91 observations in the dataset.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 91 entries, 0 to 90
Data columns (total 8 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   date            91 non-null     object
 1   Voucher_5000    91 non-null     int64
 2   Voucher_10000   91 non-null     int64
 3   Voucher_20000   91 non-null     int64
 4   Voucher_25000   91 non-null     int64
 5   Voucher_50000   91 non-null     int64
 6   Voucher_100000  91 non-null     int64
 7   Total           91 non-null     int64
dtypes: int64(7), object(1)
memory usage: 5.8+ KB
```

Figure 1 The detail of dataset

The data on historical top-up frequency for the Lao Telecommunication Company (LTC) typically includes various attributes that track customer top-up behavior over time. The dataset described appears to focus on the top-up frequency and distribution of different voucher denominations used by prepaid customers. The table 1 below is a detailed description of the data structure, including date that represents the date when the top-ups were made, voucher columns that each column represents the number of top-ups made using a specific voucher denomination and total that is sum of top-ups across all voucher denominations for that specific date.

Table 1: Sample of historical top-up frequency **for LTC**

Date	Voucher_5000	Voucher_10000	Voucher_20000	Voucher_25000	Voucher_50000	Voucher_100000	Total
2024-04-01	208	1186	90	91	82	68	1725
2024-04-02	140	972	89	81	88	45	1415
2024-04-03	121	1170	134	121	119	48	1713
2024-04-04	76	1048	141	117	83	52	1517
2024-04-05	100	1164	116	123	124	39	1666

The column Date is the number of times a customer tops up their prepaid balance within a given time period in daily (top-ups per day). Other word this column is time stamps when top-ups occurred that recorded at daily intervals from 2024-04-01 to 2024-06-01. This is the core time-series component, allowing analysis of patterns, trends, and seasonality in top-up behavior. It helps in identifying peak periods or regular intervals of high top-up activity. The columns voucher_5000**,** voucher_10000**,** voucher_20000**,** voucher_25000**,** voucher_50000**,** voucher_100000 mentioned represent specific voucher values, it's the count of top-ups likely corresponding to different denominations of top-up vouchers that customers can purchase. These columns represent the number of top-up transactions or the monetary amount of top-ups made using vouchers valued at 5,000 LAK, 10000 LAK, 20000 LAK, 25000 LAK, 50000 LAK, 100000 LAK, respectively. This can help the telecom company understand the popularity of smaller denominations among customers and how frequently they are used. The column Total is the total number of top-ups or the total monetary value of all voucher denominations combined, for a specific daily period. It is representing the count of all transactions.

4. Parameters Setting

An ARIMA model is likely from a library like stats models. The order (p, d, q) of the ARIMA model is specified as (5, 1, 0). A SARIMAX model, also found in the stats models library in Python. It's a sophisticated statistical model designed 1to handle time series data with both trend and seasonality components. The argument order specifies the non-seasonal of ARIMA component of the model. It follows the same structure as in a regular ARIMA model as (p, d, q) that specified as (1, 1, 1). The seasonal order argument specifies the seasonal component of the SARIMAX model. It also follows the (p, d, q) structure but with an additional element with (P, D, Q, s) that specified as (1, 1, 1, 12). The important parameters of ETS model include an additive trend, additive seasonal and length of the seasonal cycle, in this case, `12` implies a yearly seasonality (e.g., monthly data with peaks and troughs repeated every 12 months). This model will consider both a linear trend and a yearly seasonal component. The "additive" nature of the components assumes that the trend and seasonal effects are consistent over time.

5. Simulation Result

Figures 1 show the history view of frequency with different vouchers for the dates April 1st to June 20th. The x-axis displays the date, while the y-axis represents the frequency of vouchers.

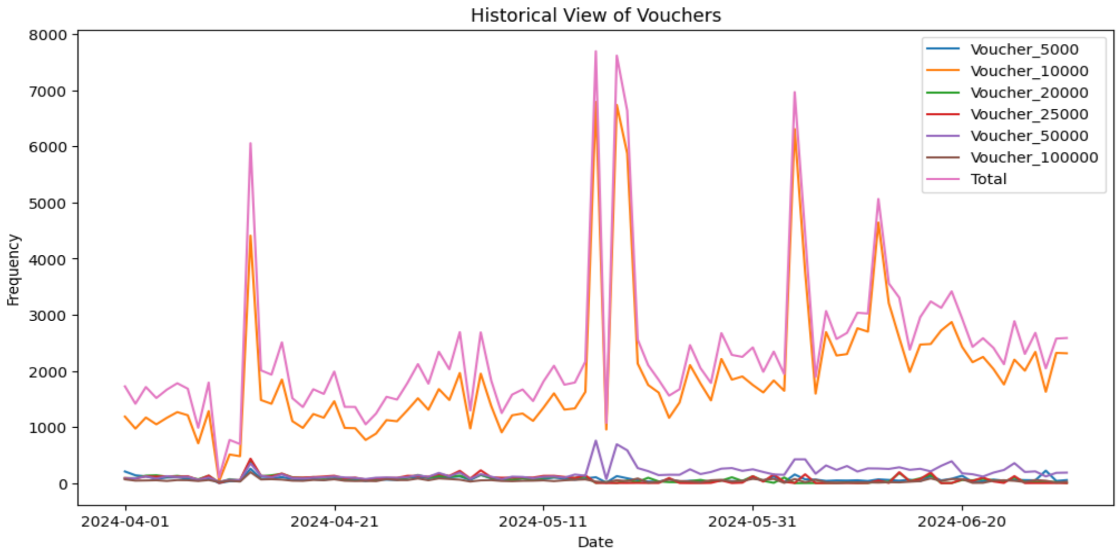

Fig. 1. History view of frequency with different vouchers

In Figure 1, there are six types of vouchers: Voucher_5000, Voucher_10000, Voucher_20000, Voucher_25000, Voucher_50000, Voucher_100000. The total frequency is also shown. The x-axis represents the date and the y-axis represents the frequency. The frequency of Voucher_5000, Voucher_10000, Voucher_20000, Voucher_25000, Voucher_50000, Voucher_100000 are all much lower than the total frequency. The total frequency has some spikes over the period and the highest frequency was around April 15th, when it reached around 7500. Overall, the graph provides a visual representation of the frequency of different types of vouchers over a specific period of time. The graph can be used to analyze trends and patterns in voucher usage.

Figures 2 to 4 show the he actual and forecast frequency of total vouchers using an ARIMA, SARIMAX and EST models, respectively.

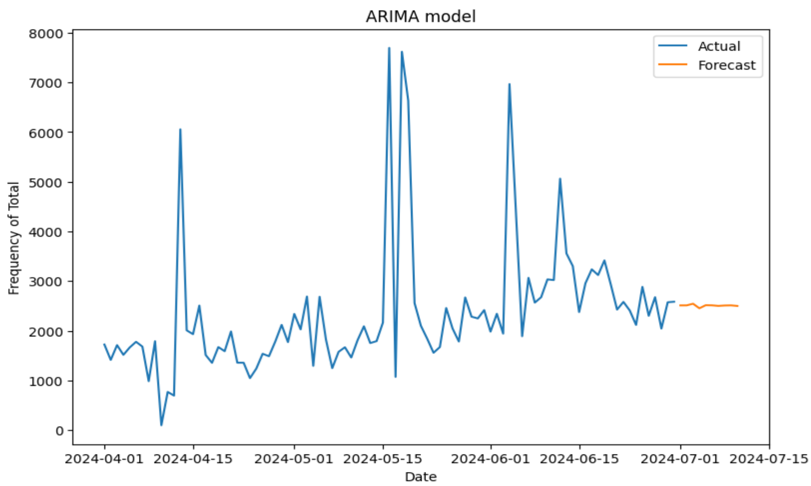

Fig. 2. The actual and forecast frequency using an ARIMA model

In Fig. 2, the blue line represents the actual frequency and the orange line represents the forecast. The model seems to be accurate for the most part but there are some deviations in the forecast. There are some noticeable spikes in the actual frequency that are not captured by the forecast, for example around April 15th and June 1st. The model suggests that the frequency would be stable around 2500 for the forecast period.

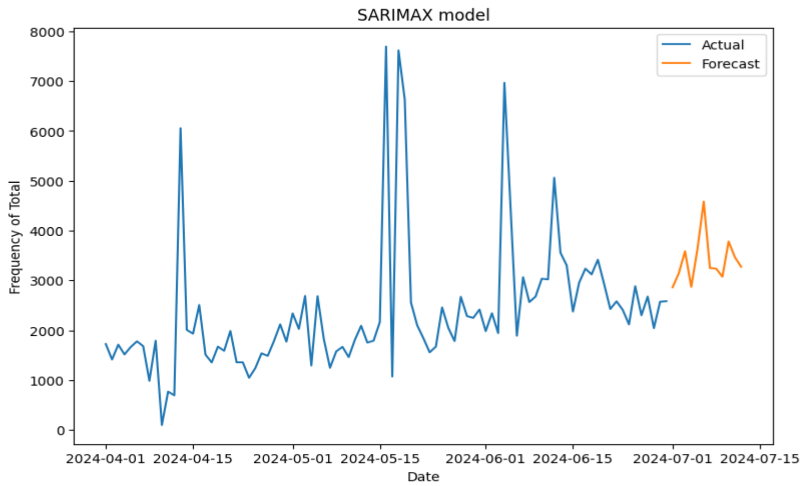

Fig. 3. The actual and forecast frequency using an SARIMAX model

Fig. 3 depicts the actual frequency of total vouchers, represented by the blue line, compared to the forecasted frequency using a SARIMAX model, represented by the orange line. The actual frequency exhibits a fluctuating pattern with pronounced spikes, especially around April 15th, May 15th, and June 1st. The SARIMAX model, while capturing the general trend, struggles to accurately predict these sharp peaks. This indicates that the model might be less effective at forecasting sudden, significant changes in frequency. The forecast for the subsequent period suggests a sharp increase in frequency around July 1st, followed by a decline. Overall, the SARIMAX model offers a general estimate of the frequency trend. However, its ability to accurately predict sharp fluctuations might be limited.

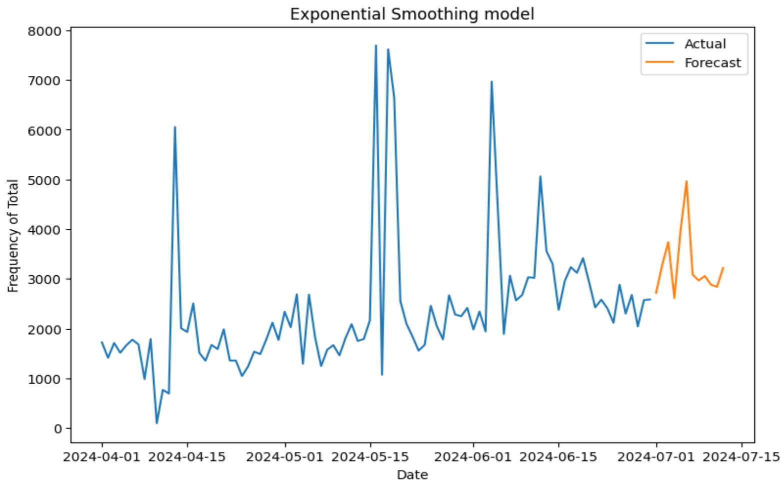

Fig. 4. The actual and forecast frequency of total vouchers using an ETS model

The Fig. 4 depicts the frequency of total vouchers, represented by the blue line, along with the forecasted frequency using an ETS model, represented by the orange line. The x-axis indicates the date, and the y-axis represents the frequency of vouchers. The actual frequency exhibits a fluctuating pattern with noticeable spikes, particularly around April 15th and June 1st. The ETS model, while capturing the general trend, underestimates the spikes. This suggests the model might be less effective at capturing sudden, significant changes in frequency. The forecast for the subsequent period indicates a gradual increase in frequency, reaching a peak around July 1st before stabilizing. Overall, the ETS model provides a general estimation of the frequency trend. However, its ability to accurately predict sharp fluctuations might be limited.

Figures 5 to 7 show the forecasting of frequency of different vouchers using an ARIMA, SARIMAX and EST models, respectively.

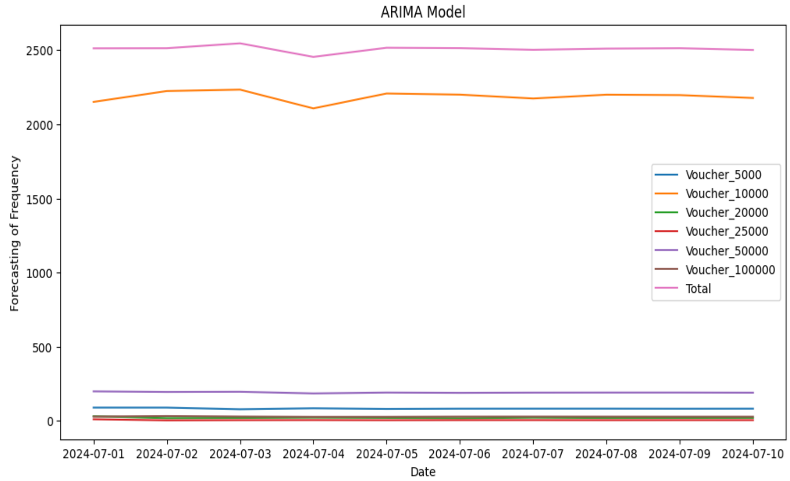

Fig. 5. The forecasting of frequency using an ARIMA model

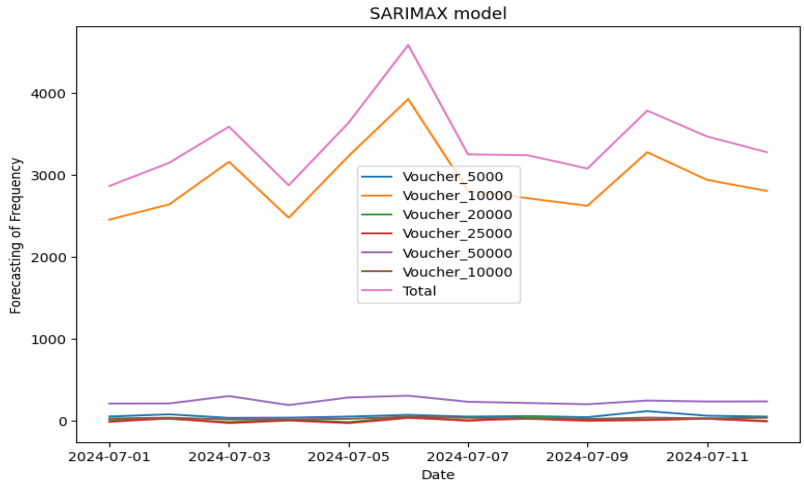

Fig. 5. The forecasting of frequency using a SARIMAX model

In figure 5, the frequency of vouchers of 5000, 10000, 20000 and 25000 are relatively low while the frequency of 50000 and 100000 are higher. The total frequency is highest.
Figure 6 depicts the forecasting of frequency of vouchers of different values using a SARIMAX model. The x-axis represents the date, and the y-axis represents the forecasting of frequency. The graph displays the frequency of vouchers of 5000, 10000, 20000, 25000, 50000, 100000 and the total. The frequency of vouchers of 5000, 10000, 20000, 25000 and 50000 are relatively low while the frequency of 100000 are higher. The total frequency is the highest. The frequency for all vouchers are expected to decrease in the future, with the total frequency expected to reach around 3200.

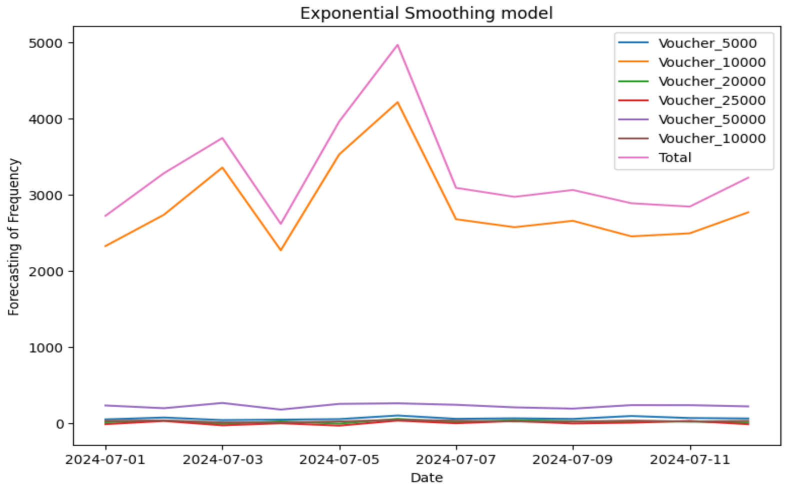

Fig. 7. The forecasting of frequency using an ETS model

Fig.7 shows the forecasting of frequency of vouchers of different values using an ETSmodel. The x-axis represents the date and the y-axis represents the forecasting of frequency. The graph displays the frequency of vouchers of 5000, 10000, 20000, 25000, 50000, 100000 and the total. The frequency of vouchers of 5000, 10000, 20000, 25000 and 50000 are relatively low while the frequency of 100000 are higher. The total frequency is the highest. The frequency for all vouchers are expected to increase in the future, with the total frequency expected to reach around 3300.

The performance of different time series forecasting models on a dataset, based on three common evaluation metrics are shown in Table 2.

Table 2 Three common evaluation metrics of different time series forecasting models

Model	RMSE	MAE	R-squared
ARIMA	241.48	199.58	0.01
SARIMAX	1111.55	910.47	-8.51
exponential smoothing	1104.01	814.23	-8.38

Table 2 presents the performance of different time series forecasting models on a dataset, based on three common evaluation metrics [18]: RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) and R-squared. ARIMA Shows the best performance overall with the lowest RMSE and MAE, and a positive R-squared indicating some ability to explain the data. SARIMAX and EST both perform poorly, with very high RMSE and MAE, and negative R-squared values, suggesting they are worse than simply using the average value as a prediction.

6. Conclusion

In this research, there was the forecasting of the top-up frequency for Lao telecommunication company using time-series analysis. It offers context-specific, advanced forecasting solutions tailored to the unique behaviors of prepaid customers in Laos, while also providing actionable insights for business operations, customer retention, and targeted marketing. The forecasting model was build using the ARIMA, SARIMAX and ETS models. Each model has its strengths and weaknesses. While all models capture the general trend, they struggle to accurately predict sharp peaks in the actual data. ARIMA is often suitable for time series data with seasonality, while ETS is simpler and can handle non-stationary data. SARIMAX is a more sophisticated model that can capture both seasonality and autoregressive components. This knowledge can inform model selection and feature engineering decisions. For instance, if voucher frequency is driven by seasonal promotions, incorporating seasonal factors into the model becomes crucial.

Reference

1. Lam P.L., Shiu A., “Economic Growth, Telecommunications Development and Productivity Growth of the Telecommunications sector: Evidence Around the World”, Telecommunications Policy, May 2010, 34(4), 185-99.
2. Vilaysouk K., “Porter’s Five Forces Analysis: An Exploratory Study of the Telecommunication Industry in Lao PDR”, Doctoral dissertation, UAJY, 2011.
3. ESCAP., UN. "Electric mobility in public transport: a guidebook for Asia-Pacific Countries", 2023.
4. Annear P.L., Akkhavong K., Thomé J.M., Haegeman F., Bonnet F., Phommavong C., Pholsena S., “Moving towards greater equity in health: recent initiatives in the Lao PDR and their implications”, Studies in Health Services Organisation and Policy, 2008, 23, 227-59.
5. Nishimura H., Kimura F., Ambashi M., Keola S., “Lao PDR at the crossroads: Industrial development strategies 2016–2030”, 2016.
6. Kingsada V., Duangvilaykeo T., Saybounheung S., “Factors Affecting for Using Services of Mobile Phone Network for Customer of the Lao Telecommunications Company, Oudomxay Province”, Souphanouvong University Journal Multidisciplinary Research and Development, Sep 2024, 10(4), 56-64.
7. Kolambe M., Arora S., “Forecasting the Future: A Comprehensive Review of Time Series Prediction Techniques”, Journal of Electrical Systems, 2024, 20(2s), 575-86.
8. Tong Y., Liu J., Yu L., Zhang L., Sun L., Li W., Ning X., Xu J., Qin H., Cai Q., “Technology investigation on time series classification and prediction”, PeerJ Computer Science, May 2022, 18(8), e982.
9. Panasa V., Kumari R.V., Ramakrishna G., Kaviraju S., “Maize price forecasting using auto regressive integrated moving average (ARIMA) model”, Int, J, Curr, Microbiol, App, Sci, 2017, 6(8), 2887-95.
10. Alharbi F.R., Csala D., “A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach”, Inventions, 2022, 7 (4), 94.
11. Tran Q.T., Hao L., Trinh Q.K., “A comprehensive research on exponential smoothing methods in modeling and forecasting cellular traffic”, Concurrency and Computation, Practice and Experience, Dec 2020, 32(23), e5602.
12. Kolambe M., Arora S., “Forecasting the Future: A Comprehensive Review of Time Series Prediction Techniques”, Journal of Electrical Systems, 2024, 20(2s), 575-86.
13. Elsamie M.A., Ali T., Zhou DY., “Using a dynamic time series model (ARIMA) for forecasting of Egyptian cotton crop variables”, 2021, 810-823.
14. Kuan M.M., “Applying SARIMA, ETS, and hybrid models for prediction of tuberculosis incidence rate in Taiwan”, PeerJ, Sep 2022. 21(10), 13117.
15. Alboukaey N., Joukhadar A., Ghneim N., “Dynamic behavior based churn prediction in mobile telecom”, Expert Systems with Applications, Dec 2020, 30(162), 113779.
16. Abdul M.F., Mohd M.A., Wong H.L., Yip S.C., Tan C.K., “A Prepaid Energy Billing System with Credit Shortage Forecasting”, In Proceedings of the 8th International Conference on Computational Science and Technology, Labuan, Malaysia, August 2022, 299-311.

17. Amajuoyi C.P., Nwobodo L.K, Adegbola A.E., “Utilizing predictive analytics to boost customer loyalty and drive business expansion”, GSC Advanced Research and Reviews, 2024, 19(3), 191-202.
18. Varol H., Ince M., Bıçaklı E.E., “Prediction of mechanical properties of the 2024 aluminum alloy by using machine learning methods”, Arabian Journal for Science and Engineering, Mar 2023, 48(3), 2841-50.