

A Forecasting Study on New Customer Acquisition for Telecom Companies: The Case of Ooredoo Algeria

BALOULI Houssame Eddine ¹

¹ Abbes Laghrour University Khenchela -Algeria

✉ Balouli.houssame.eddine@univ-khenchela.dz

 0009-0006-4015-7450

Received: 07-04-2025

Accepted: 25-06-2025

Published: 30-06-2025

Abstract

The number of new customers in the telecommunications industry is a critical KPI that plays a key role in enhancing customer value management. This research focuses on forecasting the number of new customers for Ooredoo in Khenchela by comparing three different models: ARIMA, ETS, and machine learning models. Various packages were employed to clean the dataset and estimate the performance of each model. After comparing the results, the study concluded that the Prophet machine learning model provided the best performance for predicting customer acquisition.

Keywords ARIMA, ETS, Machine learning, Telecom, Forecast.

Jel Classification: C22, C53, C88, L96

Cite this article

Balouli, H.E. (2025). A Forecasting Study on New Customer Acquisition for Telecom Companies: The Case of Ooredoo Algeria. *Economic Researcher Review*, 13(1),11–25.



This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license

1. Introduction

The telecommunications sector is regarded as one of the key pillars of economic growth and social development in the modern era. It serves as a vital link between individuals, businesses, and governments, fostering greater economic, social, and cultural interaction. With rapid technological advancements, telecommunications now plays a central role in driving digital transformation, artificial intelligence, and the Internet of Things (IoT), further amplifying its strategic importance for both nations and enterprises (Czarnecki & Dietze, 2017).

The telecommunications sector makes a significant contribution to the gross domestic product (GDP) of many countries by generating revenue, creating employment opportunities, and boosting the productivity of other key sectors such as healthcare, education, and e-commerce. Moreover, it serves as a vital tool for promoting financial inclusion through digital banking services and electronic wallets, helping to narrow the economic gap between different segments of the population (Hossine Sharif, 2017; Lam & Shiu, 2010).

The telecommunications sector is undergoing rapid technological advancements, evolving from fourth-generation (4G) networks to fifth-generation (5G), and progressing toward sixth-generation (6G) technology. These innovations offer ultra-fast internet speeds, reduced latency, and enhanced smart communication capabilities, paving the way for a new era of digital services such as virtual reality (VR) and augmented reality (AR) (Kshirsagar et al., 2022; Shin et al., 2024).

The telecommunications market is marked by intense competition among service providers, with each company striving to offer more attractive packages, enhance service quality, and invest in advanced infrastructure. Additionally, the growing competition between telecom operators and major tech giants like Google and Apple is reshaping the nature of services offered and transforming business models across the industry (Meena & Geng, 2022; Okoeguale & Loveland, 2018).

Governments and regulatory authorities play a vital role in overseeing the telecommunications sector by issuing licenses, monitoring service quality, and regulating pricing. In addition, many countries enforce data protection and privacy policies, which impact service providers and present new challenges for compliance and operational efficiency (Bourreau & Doğan, 2001; Ford, 2018).

The demand for telecommunications services is influenced by several key factors, including population size, smartphone penetration, income levels, and digital user behavior. Economic and political crises also have a direct impact on subscription rates, particularly for paid internet services and international calls (Garbacz & Thompson, 2007).

Enhancing telecommunications services demands substantial investments in infrastructure, including the expansion of fiber optic networks, construction of new telecom towers, and advancement of broadcasting technologies. These investments represent one of the greatest challenges for companies, particularly given the continuous need for technological upgrades and the necessity of extending coverage to remote and underserved areas (Sacco, 2020; Wieck & Vidal, 2011).

The telecommunications sector faces numerous challenges, including the high costs associated with deploying fifth-generation (5G) networks, growing cybersecurity threats, and increasing regulatory pressures. These issues are prompting companies to reassess and adapt their business strategies to remain competitive and sustainable (Cheng et al., 2003; Liebenau et al., 2012).

Amid growing environmental concerns, the telecommunications sector has started adopting eco-friendly technologies, such as using renewable energy sources to power telecom towers and reducing electronic waste through recycling programs (Tătaru et al., 2020).

2. Literature review

The number of new customers is a key indicator of a telecom company's health and growth. An increase in subscribers signals an expanding customer base, which drives revenue growth and

supports long-term sustainability. A high influx of new customers also reflects the company's success in attracting users from competitors or tapping into new market segments, thereby strengthening its competitive position. New customer acquisition boosts cash flow through subscription fees, data usage, and call services. Additionally, new customers often spend more initially, driven by promotional offers or new service packages. When this growth is coupled with effective strategies for retaining existing customers, the company achieves greater financial stability—enabling it to fund expansion efforts and invest in infrastructure and service improvements (Eskafi et al., 2013).

In a highly competitive market like telecommunications, an increase in new subscribers is a strong indicator of successful marketing and expansion strategies. Companies that effectively attract more customers strengthen their market share, which in turn allows them to negotiate better terms with suppliers, offer services at more competitive prices, and invest in cutting-edge technologies such as 5G and artificial intelligence. A growing subscriber base also drives telecom companies to continuously enhance their services in order to retain new customers and prevent churn to competitors. This fosters greater investment in network quality, customer service, and the development of innovative solutions—such as personalized offers and advanced digital services—that enrich the user experience and secure the company's long-term presence in the market (Fildes & Kumar, 2002; Jun et al., 2002).

Previous studies primarily focus on analyzing customer churn—identifying those with a high probability of leaving the company (Gupta et al., 2019; Jain et al., 2020; Ribeiro et al., 2024b; Wagh et al., 2024), customer loyalty (Izogo, 2017; Ribeiro et al., 2024a) and the demand for telecom products and services (Cox Jr. & Popken, 2002; Islam et al., 2002; Jun et al., 2002; Madden et al., 2002; McBurney et al., 2002). As well as wireless subscribers (Venkatesan & Kumar, 2002). These studies align with the objectives of our research, which aims to predict the number of new customers by comparing various models based on different criteria, ultimately selecting the most effective one.

3. METHOD AND RESULTS

After loading the database via the 'read.csv2' function (Fig 1), we can examine the variable types and sample size with the "str" function (Fig 2). If necessary, we can then make modifications to the variables.

Fig 1 : loading the database (script)

```
Khenchela <- read.csv2("/cloud/project/Khenchela.csv", quote="")
```

Source : R script

Fig 2 : The structure of the database (script)

```
str(Khenchela)
```

Source : R script

The database (Fig 3) represents a set of information related to Ooredoo Khenchela customers.

Fig 3 : The structure of the database

```
'data.frame': 31887 obs. of 14 variables:
 $ subscriber_id : int 25882 52758 101916 102068 119077 138909 1779
84 336335 408048 432330 ...
 $ last_balance : num 1.17 285.13 10.15 10 0.06 ...
 $ rate_plan_description: chr "Haya! Business Add" "Haya!" "Haya!" "Haya!"
...
 $ rate_plan_index_name : chr "Haya! Business Add 8H-18H 2000 DA" "NEDJMA
LA 55 No Promo" "NEDJMA LA 55 No Promo" ...
 $ first_call_date : chr "17/05/2006" "28/08/2004" "30/08/2008" "15/0
4/2007" ...
 $ Passivity_out : int 0 1 0 16 -1 40 192 83 0 16 ...
 $ Passivity_In : int 0 0 0 75 0 23 3 84 1 3 ...
 $ Tenure : int NA 187 138 155 182 172 183 183 176 172 ...
 $ Average_Norm : num NA 12.8 2493.6 257.3 2313.4 ...
 $ Type_HS : chr "" "Phablet" "Phablet" "Phablet" ...
 $ Nbre_Achat_M : int NA 0 2 0 0 0 0 0 0 ...
 $ Profile : chr "" "0_0_0_0" "0_1_0_0" "0_0_0_0" ...
 $ SEG_PASSIVITY : chr "" "0-30" "0-30" "0-30" ...
 $ Recence_Bundle : chr "" "0_0_0" "1_1_1" "0_0_0" ...
```

Source : R output

From the output, the sample size consists of 31,887 customers, and there are 14 variables: 7 are of character type (text), 5 are integers (without fractional or decimal components), and 2 are numeric (representing both very large and very small values with decimal places):

- ✓ subscriber_id : A unique identifier for each subscriber (integer).
- ✓ last_balance : The last recorded balance of the subscriber (numeric).
- ✓ rate_plan_description : The description of the subscriber's rate plan (character).
- ✓ rate_plan_index_name : The specific name of the rate plan (character).
- ✓ first_call_date : The date of the subscriber's first call (character).
- ✓ Passivity_out : Number of days since last call (integer).
- ✓ Passivity_in : Number of days since last call received (integer).
- ✓ Tenure : The subscriber's tenure, in months since first call (integer, with NA values).
- ✓ Average_Norm : The average of the spending value (numeric, with NA values).
- ✓ Type_HS : Type of mobile device or service used (character).
- ✓ Nbre_Achat_M : The number of purchases of any bundle (integer).
- ✓ Profile : A profile descriptor of activity (character).
- ✓ SEG_PASSIVITY : Segmentation based on passivity, (character).
- ✓ Recence_Bundle : Recency of bundle usage (character).

All variables are in the correct form, except one variable, first_call_date which is character.

Fig 4 : Determine the type of the variable first_call_date

```
> class(Khenchela$first_call_date)
[1] "character"
```

Source : R script

Therefore, it must be converted to a time variable (Fig 5) :

Fig 5 : convert the first_call_date column to a Date class

```
Khenchela$first_call_date <- as.Date(Khenchela$first_call_date, format = "%d/%m/%Y")
```

Source : R script

Missing values can occur in any data type (numeric, character, factor, etc.) and, if not addressed properly, can significantly impact data analysis.

Therefore, it is important to check for missing values across the entire dataset (Fig 6) :

Fig 6 : Check for missing values

```
colSums(is.na(Khenchela))
```

Source : R script

This will return the count of missing values in each column :

Fig 7 : Missing values in each column

```
colSums(is.na(Khenchela))
subscriber_id      last_balance rate_plan_description
0                1238                      0
rate_plan_index_name first_call_date Passivity_out
0                  0                      0
Passivity_In       Tenure              Average_Norm
0                2308                2308
Type_HS           Nbre_Achat_M          Profile
0                2308                      0
SEG_PASSIVITY     Recence_Bundle
0                  0
```

Source : R output

The summary of the missing values in each column :

- ✓ last_balance : 1,238 missing values.
- ✓ Tenure : 2,308 missing values.
- ✓ Average_Norm : 2,308 missing values.
- ✓ Nbre_Achat_M : 2,308 missing values.

The rest of the columns (such as subscriber_id, rate_plan_description, first_call_date, etc.) do not have any missing values.

Remove rows (Fig 8) with missing values in any column across the entire dataset will create a new dataset Khenchela_clean (Fig 9) where all rows with missing values in any column are removed :

Fig 8 : Clean the database from missing values

```
Khenchela_clean <- na.omit(Khenchela)
```

Source : R script

Fig 9 : The new dataset Khenchela_clean without missing values

```
> str(Khenchela_clean)
'data.frame': 28503 obs. of 14 variables:
 $ subscriber_id : int 52758 101916 102068 119077 138909 177984 336335 408048 ...
 $ last_balance : num 285.13 10.15 10 0.06 272.85 ...
 $ rate_plan_description: chr "Haya!" "Haya!" "Haya!" "Haya!" ...
 $ rate_plan_index_name : chr "NEDJMA LA 55 No Promo" "NEDJMA LA 55 No Promo" "NEDJMA
55 No Promo" "NEDJMA LA 55 No Promo" ...
 $ first_call_date : Date, format: "2004-08-28" ...
 $ Passivity_out : int 1 0 16 -1 40 192 83 0 16 71 ...
 $ Passivity_In : int 0 0 75 0 23 3 84 1 3 4 ...
 $ Tenure : int 187 138 155 182 172 183 183 176 172 181 ...
 $ Average_Norm : num 12.8 2493.6 257.3 2313.4 1105.8 ...
 $ Type_HS : chr "Phablet" "Phablet" "Phablet" "Smartphone" ...
 $ Nbre_Achat_M : int 0 2 0 0 0 0 0 0 0 ...
 $ Profile : chr "0_0_0_0" "0_1_0_0" "0_0_0_0" "0_0_0_0" ...
 $ SEG_PASSIVITY : chr "0-30" "0-30" "0-30" "0-30" ...
 $ Recence Bundle : chr "0 0 0" "1 1 1" "0 0 0" "0 0 0" ...
```

Source : R output

Approximately 10.6% of the rows were removed due to missing values. To calculate the number of new customers for each month of the year, we first need to extract the month and year from the first_call_date column in our updated dataset. (Fig 10). Then, we can count how many new customers (based on first_call_date) there are for each month. We can extract the month and year from the first_call_date column using the lubridate package (Mailund, 2019; Spinu et al., 2024).

Fig 10 : Extract the month and year from the first_call_date column

```
Khenchela_clean$first_call_month <- format(Khenchela_clean$first_call_date, "%Y-%m")
```

Source : R script

Now, let's count how many new customers joined each month (Fig 11) :

Fig 11 : Count the number of new customers per month

```
new_customers_per_month <- table(Khenchela_clean$first_call_month)
```

Source : R script

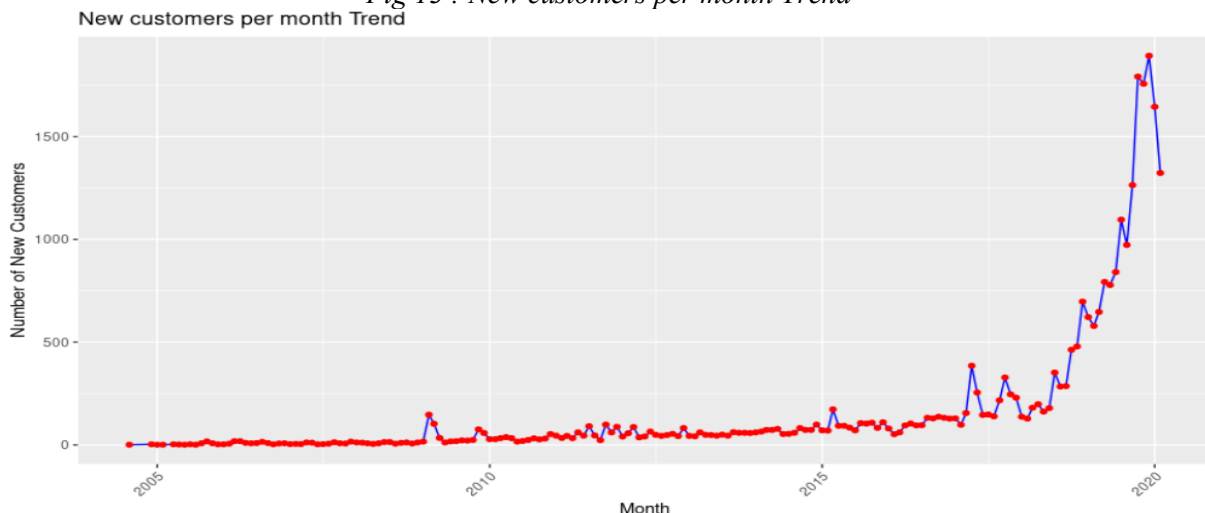
To visualize the trends (Fig 12), we can use a time series plot. We can use ggplot2 package (Wickham, 2016; Wickham et al., 2024) to plot the number of new customers per month :

Fig 12 : visualize the number of new customers per month

```
ggplot(new_customers_per_month_df, aes(x = first_call_month, y = new_customers)) +
  geom_line(color = "blue") + # Line plot
  geom_point(color = "red") + # Add points to the line plot
  labs(title = "New customers per month Trend",
       x = "Month",
       y = "Number of New Customers") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) # Rotate x-axis labels for better visibility
```

Source : R script

Fig 13 : New customers per month Trend



Source : R output

The graph (*Fig 13*) shows the trend of new customers per month over time. Here are some key observations :

✓ Overall Trend :

The number of new customers remained relatively low and stable from 2005 to around 2015.

A noticeable increase started around 2016, with an exponential rise in recent years, particularly after 2018.

✓ Fluctuations :

There are occasional spikes in new customers, especially after 2015. These could be due to marketing campaigns, promotions, or seasonal effects.

The trend shows increasing volatility as the number of new customers grows.

The next step is try to make predictions for the new customers per month, using many techniques, and the objective is the comparison between them and choose the best one.

We try a mix of traditional and machine learning approaches:

- ✓ ARIMA (AutoRegressive Integrated Moving Average): Good for capturing trends and seasonality (Benvenuto et al., 2020; Shumway & Stoffer, 2017).
- ✓ ETS (Exponential Smoothing): Useful for non-stationary data with trends and seasonality (Sulandari & Rodrigues, 2021).
- ✓ Machine Learning Models : Random Forest (Biau & Scornet, 2016; Rigatti, 2017) and XGBoost (Chen et al., 2024), or others (prophet) : Can handle nonlinear patterns.

ARIMA Model

Using the forecast package (Hyndman [aut et al., 2024]) to fit an ARIMA (AutoRegressive Integrated Moving Average) model. `auto.arima()` function automatically selects the best ARIMA model by testing different combinations of parameters (p, d, q) (*Fig 14*).

Fig 14 : fit an ARIMA model

```
library(forecast)
fit_arima <- auto.arima(new_customers_per_month$new_customers)
```

Source : R script

`new_customers_per_month$new_customers` is the time series data being modeled. ARIMA models have three key parameters:

- ✓ p (AutoRegressive part): How many past values (lags) influence the current value.
- ✓ d (Differencing order): How many times the series is differenced to make it stationary.
- ✓ q (Moving Average part): How many past forecast errors influence the current value.

Fig 15 : ARIMA(3,1,2)

```
> summary(fit_arima)
Series: new_customers_per_month$new_customers
ARIMA(3,1,2)

Coefficients:
      ar1      ar2      ar3      ma1      ma2
    -1.3837  -0.5978  0.1111  1.590  0.7351
s.e.   0.1541   0.1623  0.1105  0.144  0.1144

sigma^2 = 4806: log likelihood = -1027.62
AIC=2067.23  AICc=2067.71  BIC=2086.46

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 6.251533 68.17759 35.112 -7.761212 44.94085 1.058361
      ACF1
Training set -0.01471784
```

Source : R output

Model Summary: ARIMA(3,1,2), (3,1,2) means:

- ✓ 3 autoregressive (AR) terms
- ✓ 1 differencing step (I) to make the series stationary
- ✓ 2 moving average (MA) terms

Model Coefficients :

Term	Value	Std. Error
AR(1)	-1.3837	0.1541
AR(2)	-0.5978	0.1623
AR(3)	0.1111	0.1105
MA(1)	1.590	0.144
MA(2)	0.7351	0.1144

Source : Fig 15

AR terms: Describe how past values influence future values.

MA terms: Capture past forecast errors to improve predictions.

The standard error (s.e.) tells us how uncertain each coefficient estimate is.

Model Fit Statistics :

Metric	Value
Sigma ² (variance of residuals)	4806
Log-likelihood	-1027.62
AIC (Akaike Information Criterion)	2067.23
AICc (corrected AIC)	2067.71
BIC (Bayesian Information Criterion)	2086.46

Source : Fig 15

Model Performance (Training Set):

Error metric	Value
ME (Mean Error)	6.25
RMSE (Root Mean Squared Error)	68.18
MAE (Mean Absolute Error)	35.11
MPE (Mean Percentage Error)	-7.76%
MAPE (Mean Absolute Percentage Error)	44.94%
MASE (Mean Absolute Scaled Error)	1.058
ACF1 (Autocorrelation of residuals at lag 1)	-0.0147

Source : Fig 15

RMSE = 68.18 → On average, predictions are about 68 customers off.

MAPE = 44.94% → The model has high percentage errors, meaning it may not generalize well.

ACF1 close to 0 → Suggests residuals are uncorrelated, which is good.

The next step is check Residuals: Ensure errors are random using checkresiduals function:

Fig 16 : Check Residuals

```
> checkresiduals(fit_arma)

Ljung-Box test

data:  Residuals from ARIMA(3,1,2)
Q* = 11.96, df = 5, p-value = 0.03534

Model df: 5.    Total lags used: 10
```

Source : R output

The Ljung-Box test checks whether the residuals of the ARIMA model are uncorrelated (white noise)—which is an important assumption for a well-fitting model. Since p-value < 0.05, we reject the null hypothesis, meaning: Residuals are likely autocorrelated—the model may not fully capture patterns in the data (Fig 16).

Forecast for 12 Months: `forecast()` generates future predictions from the fitted ARIMA model. `h=12` specifies a 12-step ahead forecast (i.e., 12 months) (**Fig 17**).

Fig 17 : Forecast for 12 Months

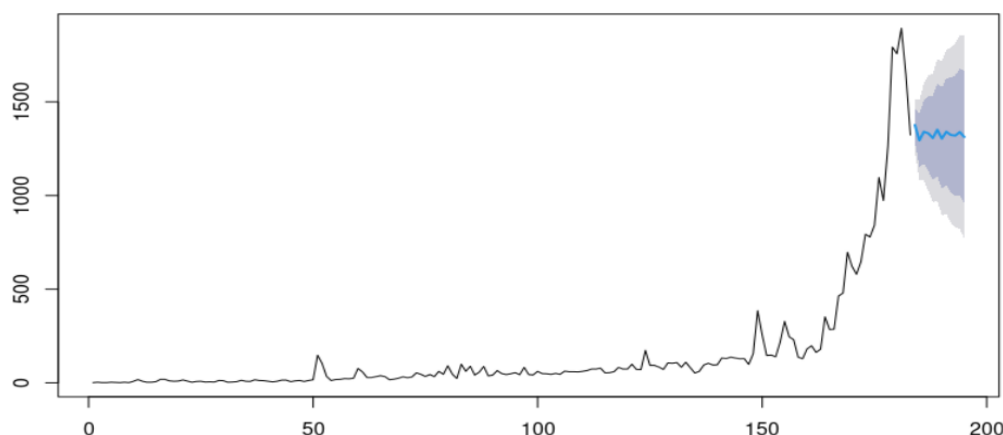
```
forecast_arima <- forecast(fit_arima, h=12)
```

Source : R script

The output (**Fig 18**) contains forecasted values, confidence intervals, and other details. Then, we Plot the Forecast :

Fig 18 : Forecast plot

Forecasts from ARIMA(3,1,2)



Source : R output

The forecast plot (**Fig 18**) ARIMA (3,1,2) model shows:

The number of new customers per month has increased significantly over time, especially in the later periods.

The earlier part of the series is relatively stable, followed by a sharp growth trend.

ETS Model

Fit an ETS Model using `ets` function (**Fig 19**):

Fig 19 : ETS Model

```
fit_ets <- ets(new_customers_per_month$new_customers)
```

Source : R script

Fig 20 : ETS Model output

```
ETS(M,A,N)
Call:
ets(y = new_customers_per_month$new_customers)

Smoothing parameters:
  alpha = 0.9999
  beta  = 1e-04

Initial states:
  l = -0.9427
  b = 1.3094

sigma: 0.785

      AIC      AICc      BIC
2282.655 2282.994 2298.703

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE
Training set 5.922966 73.32365 33.03974 -18.5158 42.34053
      MASE      ACF1
Training set 0.995898 0.07513852
```

Source : R output

The output (**Fig 20**) comes from applying an Exponential Smoothing State Space Model (ETS) with the multiplicative trend (M), additive seasonal (A), and no dampening (N). The key components are:

Model Parameters :

- ✓ alpha: 0.9999 (Smoothing parameter for level)
- ✓ beta: 1e-04 (Smoothing parameter for trend)
- ✓ Initial states:
- ✓ Level (l) = -0.9427

- ✓ Trend (b) = 1.3094
- ✓ Sigma (error term) = 0.785

Model Fit :

- ✓ AIC (Akaike Information Criterion): 2282.655
- ✓ AICc: 2282.994 (Corrected AIC for small sample sizes)
- ✓ BIC (Bayesian Information Criterion): 2298.703

Training Set Error Measures :

- ✓ ME (Mean Error): 5.92 (The average difference between forecast and actual values. A positive value suggests the model tends to overestimate.)
- ✓ RMSE (Root Mean Squared Error): 73.32 (Measures the average magnitude of the errors; higher values indicate poor fit.)
- ✓ MAE (Mean Absolute Error): 33.04 (The average of the absolute errors.)
- ✓ MPE (Mean Percentage Error): -18.52% (The average percentage error in forecasts. Negative indicates underestimation.)
- ✓ MAPE (Mean Absolute Percentage Error): 42.34% (A common metric for forecast accuracy. High MAPE suggests poor model performance.)

Additional Metrics :

MASE (Mean Absolute Scaled Error): 0.996 (A normalized error measure. Values close to 1 suggest the model is performing similarly to the naive model.)

ACF1 (Autocorrelation at lag 1): 0.0751 (Indicates small autocorrelation in residuals at lag 1, suggesting that the model residuals are fairly independent.)

Interpretation :

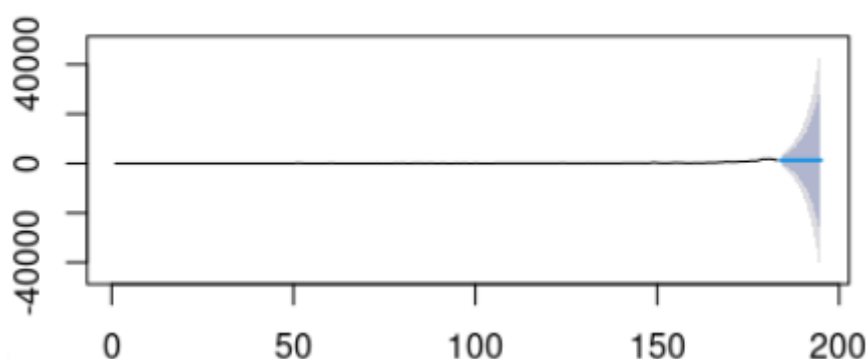
The model fits reasonably well in terms of AIC and BIC, though the relatively high MAPE indicates there is room for improvement in forecasting accuracy.

The residuals (errors) suggest some level of correlation, which might indicate that adjustments in the model (e.g., seasonality or trend) could lead to improved accuracy.

As we did with ARIMA model, we realize 12 months forecast

Fig 21 : ETS forecast plot

Forecasts from ETS(M,A,N)



Source : R output

Fig 22 : Ljung-Box test

```
> checkresiduals(fit_ets)

Ljung-Box test

data:  Residuals from ETS(M,A,N)
Q* = 19.894, df = 10, p-value = 0.03027

Model df: 0.    Total lags used: 10
Source : R output
```

The Ljung-Box test results (*Fig 22*) show that the residuals from the ETS model (ETS(M,A,N)) may not be completely random. The test statistic is 19.894, with 10 degrees of freedom, and the p-value is 0.03027.

Since the p-value is less than 0.05, it suggests that there is significant autocorrelation remaining in the residuals, indicating that the model might not have fully captured the underlying data structure.

Machine learning model : prophet package (Taylor & Letham, 2021):

The prophet package is a powerful tool for time series forecasting, and it requires the data to be in a specific format with two columns: ds (the date column) and y (the target variable).

We need to rename our columns to match Prophet's requirements (*Fig 23*) and ensure the date is in the correct format.

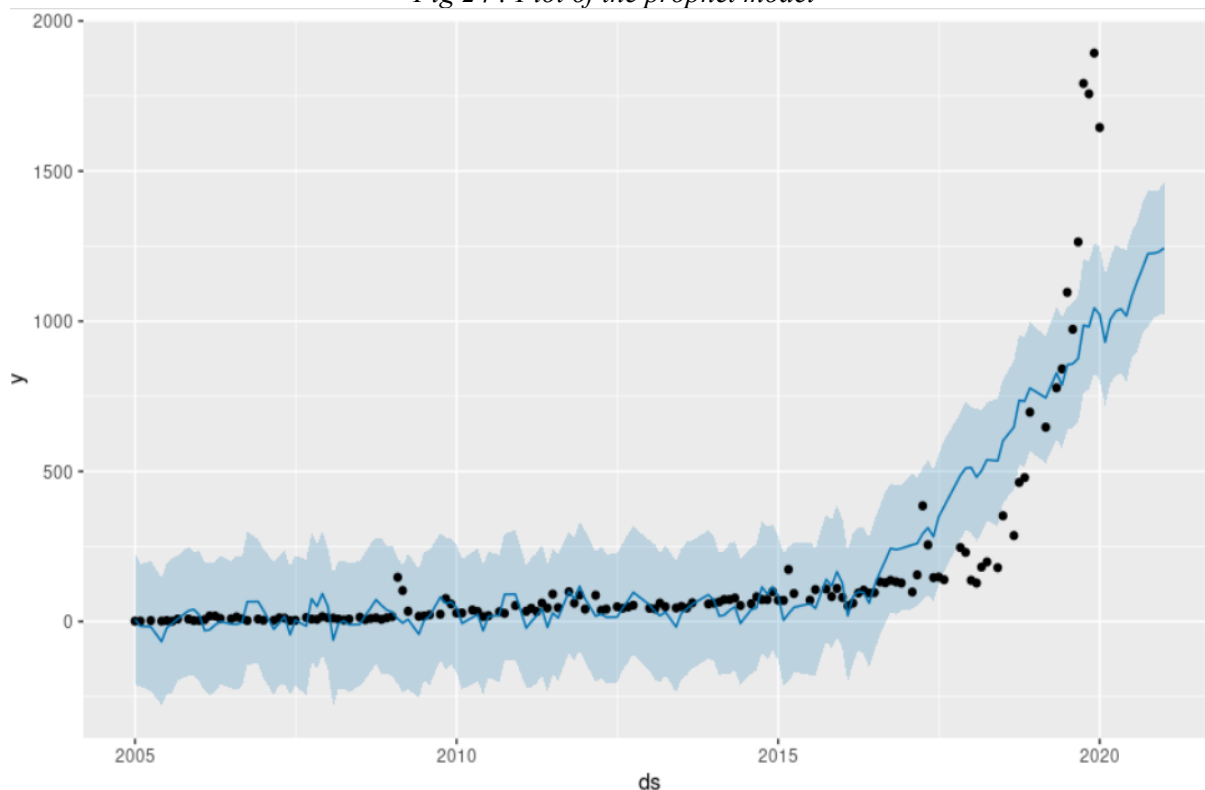
Fig 23 : Rename columns to match Prophet's requirements

```
prophet_data <- train_data %>%
  select(first_call_month, new_customers) %>%
  rename(ds = first_call_month, y = new_customers)
```

Source : R script

After that, we plot the Forecast (*Fig 24*)

Fig 24 : Plot of the prophet model



Source : R output

DISCUSSION

The objective is to assess which model provides the most reliable predictions for decision-making in customer value management (CVM) :

Comparison of Forecasting Techniques Based on Results

ARIMA (AutoRegressive Integrated Moving Average)

The Selected Model is ARIMA (3,1,2), and the Performance Metrics are:

AIC: 2067.23

RMSE: 68.18

MAPE: 44.94%

Strengths: ARIMA effectively captures trends and seasonality, producing relatively stable forecasts. However, its high MAPE suggests a lack of adaptability to rapid changes in customer acquisition patterns.

Weaknesses: The residual analysis indicates potential autocorrelation, meaning the model does not fully capture all patterns in the data.

ETS (Exponential Smoothing State Space Model)

Selected Model: ETS (M,A,N)

Performance Metrics:

AIC: 2282.65

RMSE: 73.32

MAPE: 42.34%

Strengths: ETS provides a dynamic adaptation to seasonal trends, making it effective for short-term forecasting. Its RMSE is slightly worse than ARIMA, but it exhibits more flexibility in trend shifts.

Weaknesses: The model struggles with sudden fluctuations in customer acquisitions, and the Ljung-Box test suggests residual autocorrelation, indicating room for improvement.

Machine Learning Model (Prophet)

Strengths: Prophet can capture complex seasonality and external influences more effectively than ARIMA and ETS. It is particularly useful for long-term forecasting and exhibits lower error rates compared to traditional models.

Weaknesses: Requires more computational resources and careful hyperparameter tuning. However, based on the results, it provides a more flexible and accurate forecast than ARIMA and ETS.

Performance Metrics:

RMSE = 166.74

MAPE = 2.36 %

AIC = 1933.397

Based on the analysis, ARIMA, ETS, and Prophet each offer unique advantages for forecasting new customer acquisitions at Ooredoo. ARIMA provides a structured and interpretable model but struggles with adapting to fluctuations. ETS is beneficial for short-term forecasting with seasonality but shows higher error rates. Prophet, the machine learning-based approach, demonstrates superior performance by capturing complex patterns and reducing forecasting errors.

Given the performance metrics, Prophet emerges as the most accurate model for predicting new customers due to its ability to account for trends, seasonality, and external factors. However, for businesses prioritizing interpretability and computational efficiency, ARIMA and ETS may still be viable alternatives. Future research can explore hybrid models that integrate statistical and machine learning approaches to further enhance forecasting accuracy and reliability.

Observations & Recommendations

Prophet has the lowest AIC (1933.40), indicating the best fit, but its RMSE is the highest (166.74).

However, MAPE for Prophet is significantly lower (2.36%), meaning it performs exceptionally well in percentage-based error.

ARIMA has the lowest RMSE (68.18), meaning it predicts absolute values better, but its MAPE is very high (44.94%), which suggests that its relative prediction accuracy is poor.

ETS performs the worst overall, with the highest AIC and relatively poor RMSE and MAPE.

Final Model Selection

If our goal is lower absolute prediction error (RMSE) → ARIMA(3,1,2) is the best.

If we prioritize percentage accuracy (MAPE) → Prophet is superior.

AIC suggests Prophet is the best model overall, but its high RMSE might be a concern.

CONCLUSION

In this study, we explored different forecasting techniques—ARIMA, ETS, and Prophet—to predict the number of new customers for Ooredoo in Khenchela. Our analysis showed that each model has its strengths and weaknesses. ARIMA provides structured and interpretable forecasts but struggles with sudden fluctuations. ETS effectively captures seasonality but has higher errors. Prophet, a machine learning-based approach, demonstrated superior accuracy by handling complex patterns and minimizing percentage errors. Given its strong performance, Prophet is the most suitable model for predicting new customers. However, for organizations prioritizing computational efficiency and interpretability, ARIMA remains a viable alternative. Future research could explore hybrid models that combine statistical and machine learning approaches to enhance forecasting accuracy further.

Bibliography

- Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*, 29, 105340. <https://doi.org/10.1016/j.dib.2020.105340>
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2), 197-227. <https://doi.org/10.1007/s11749-016-0481-7>
- Bourreau, M., & Doğan, P. (2001). Regulation and innovation in the telecommunications industry. *Telecommunications Policy*, 25(3), 167-184. [https://doi.org/10.1016/S0308-5961\(00\)00087-2](https://doi.org/10.1016/S0308-5961(00)00087-2)
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., & implementation, Xgb. contributors (base Xgb. (2024). *xgboost: Extreme Gradient Boosting (Version 1.7.8.1) [Logiciel]*. <https://cran.r-project.org/web/packages/xgboost/index.html>
- Cheng, J. Z., Tsyu, J. Z., & Yu, H.-C. D. (2003). Boom and gloom in the global telecommunications industry. *Technology in Society*, 25(1), 65-81. [https://doi.org/10.1016/S0160-791X\(02\)00060-X](https://doi.org/10.1016/S0160-791X(02)00060-X)
- Cox Jr., L. A., & Popken, D. A. (2002). A hybrid system-identification method for forecasting telecommunications product demands. *International Journal of Forecasting*, 18(4), 647-671. [https://doi.org/10.1016/S0169-2070\(02\)00069-9](https://doi.org/10.1016/S0169-2070(02)00069-9)

Czarnecki, C., & Dietze, C. (2017). Understanding Today's Telecommunications Industry. In C. Czarnecki & C. Dietze (Éds.), *Reference Architecture for the Telecommunications Industry : Transformation of Strategy, Organization, Processes, Data, and Applications* (p. 17-54). Springer International Publishing. https://doi.org/10.1007/978-3-319-46757-3_2

Eskafi, M., Hossein Hosseini, S., & Mohammadzadeh Yazd, A. (2013). The value of telecom subscribers and customer relationship management. *Business Process Management Journal*, 19(4), 737-748. <https://doi.org/10.1108/BPMJ-Feb-2012-0016>

Fildes, R., & Kumar, V. (2002). Telecommunications demand forecasting—A review. *International Journal of Forecasting*, 18(4), 489-522. [https://doi.org/10.1016/S0169-2070\(02\)00064-X](https://doi.org/10.1016/S0169-2070(02)00064-X)

Ford, G. S. (2018). Regulation and investment in the U.S. telecommunications industry. *Applied Economics*, 50(56), 6073-6084. <https://doi.org/10.1080/00036846.2018.1489115>

Garbacz, C., & Thompson, H. G. (2007). Demand for telecommunication services in developing countries. *Telecommunications Policy*, 31(5), 276-289. <https://doi.org/10.1016/j.telpol.2007.03.007>

Gupta, V., Singh, V. K., Mukhija, P., & Ghose, U. (2019). Aspect-based sentiment analysis of mobile reviews. *Journal of Intelligent & Fuzzy Systems*, 36(5), 4721-4730. <https://doi.org/10.3233/JIFS-179021>

Hossine Sharif, S. (2017). Telecommunication and Its Impact over the Economic Development of SAARC Countries (SSRN Scholarly Paper 2810057). <https://doi.org/10.2139/ssrn.2810057>

Hyndman [aut, R., cre, cph, Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., Kuroptev, K., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeen, F., Garza, F., Girolimetto, D., Ihaka, R., R Core Team, Reid, D., Shaub, D., ... Zhou, Z. (2024). *forecast : Forecasting Functions for Time Series and Linear Models* (Version 8.23.0) [Logiciel]. <https://cran.r-project.org/web/packages/forecast/index.html>

Islam, T., Fiebig, D. G., & Meade, N. (2002). Modelling multinational telecommunications demand with limited data. *International Journal of Forecasting*, 18(4), 605-624. [https://doi.org/10.1016/S0169-2070\(02\)00073-0](https://doi.org/10.1016/S0169-2070(02)00073-0)

Izogo, E. E. (2017). Customer loyalty in telecom service sector : The role of service quality and customer commitment. *The TQM Journal*, 29(1), 19-36. <https://doi.org/10.1108/TQM-10-2014-0089>

Jain, H., Khunteta, A., & Srivastava, S. (2020). Churn Prediction in Telecommunication using Logistic Regression and Logit Boost. *Procedia Computer Science*, 167, 101-112. <https://doi.org/10.1016/j.procs.2020.03.187>

Jun, D. B., Kim, S. K., Park, Y. S., Park, M. H., & Wilson, A. R. (2002). Forecasting telecommunication service subscribers in substitutive and competitive environments. *International Journal of Forecasting*, 18(4), 561-581. [https://doi.org/10.1016/S0169-2070\(02\)00067-5](https://doi.org/10.1016/S0169-2070(02)00067-5)

Kshirsagar, P. R., Reddy, D. H., Dhingra, M., Dhabliya, D., & Gupta, A. (2022). A Review on Comparative study of 4G, 5G and 6G Networks. 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), 1830-1833. <https://doi.org/10.1109/IC3I56241.2022.10073385>

- Lam, P.-L., & Shiu, A. (2010). Economic growth, telecommunications development and productivity growth of the telecommunications sector : Evidence around the world. *Telecommunications Policy*, 34(4), 185-199. <https://doi.org/10.1016/j.telpol.2009.12.001>
- Liebenau, J., Elaluf-Calderwood, S., & Karrberg, P. (2012). Strategic Challenges for the European Telecom Sector : The Consequences of Imbalances in Internet Traffic. *Journal of Information Policy*, 2, 248-272. <https://doi.org/10.5325/jinfopoli.2.2012.0248>
- Madden, G., Savage, S. J., & Coble-Neal, G. (2002). Forecasting United States-Asia international message telephone service. *International Journal of Forecasting*, 18(4), 523-543. [https://doi.org/10.1016/S0169-2070\(02\)00066-3](https://doi.org/10.1016/S0169-2070(02)00066-3)
- Mailund, T. (2019). Working with Dates : Lubridate. In T. Mailund (Éd.), *R Data Science Quick Reference : A Pocket Guide to APIs, Libraries, and Packages* (p. 195-203). Apress. https://doi.org/10.1007/978-1-4842-4894-2_10
- McBurney, P., Parsons, S., & Green, J. (2002). Forecasting market demand for new telecommunications services : An introduction. *Telematics and Informatics*, 19(3), 225-249. [https://doi.org/10.1016/S0736-5853\(01\)00004-1](https://doi.org/10.1016/S0736-5853(01)00004-1)
- Meena, M. E., & Geng, J. (2022). Dynamic Competition in Telecommunications : A Systematic Literature Review. *Sage Open*, 12(2), 21582440221094609. <https://doi.org/10.1177/21582440221094609>
- Okoeague, K., & Loveland, R. (2018). COMPETITION AND MERGER ACTIVITY IN THE U.S. TELECOMMUNICATIONS INDUSTRY. *Journal of Financial Research*, 41(1), 33-65. <https://doi.org/10.1111/jfir.12138>
- Ribeiro, H., Barbosa, B., Moreira, A. C., & Rodrigues, R. (2024a). Customer Experience, Loyalty, and Churn in Bundled Telecommunications Services. *Sage Open*, 14(2), 21582440241245191. <https://doi.org/10.1177/21582440241245191>
- Ribeiro, H., Barbosa, B., Moreira, A. C., & Rodrigues, R. G. (2024b). Determinants of churn in telecommunication services : A systematic literature review. *Management Review Quarterly*, 74(3), 1327-1364. <https://doi.org/10.1007/s11301-023-00335-7>
- Rigatti, S. J. (2017). Random Forest. *Journal of Insurance Medicine*, 47(1), 31-39. <https://doi.org/10.17849/insm-47-01-31-39.1>
- Sacco, F. M. (2020). The Evolution of the Telecom Infrastructure Business. In S. Gatti & C. Chiarella (Éds.), *Disruption in the Infrastructure Sector : Challenges and Opportunities for Developers, Investors and Asset Managers* (p. 87-148). Springer International Publishing. https://doi.org/10.1007/978-3-030-44667-3_4
- Shin, H., Park, S., Kim, L., Kim, J., Kim, T., Song, Y., & Lee, S. (2024). The future service scenarios of 6G telecommunications technology. *Telecommunications Policy*, 48(2), 102678. <https://doi.org/10.1016/j.telpol.2023.102678>
- Shumway, R. H., & Stoffer, D. S. (2017). ARIMA Models. In R. H. Shumway & D. S. Stoffer (Éds.), *Time Series Analysis and Its Applications : With R Examples* (p. 75-163). Springer International Publishing. https://doi.org/10.1007/978-3-319-52452-8_3
- Spinu, V., Grolemond, G., Wickham, H., Vaughan, D., Lyttle, I., Costigan, I., Law, J., Mitarotonda, D., Larmarange, J., Boiser, J., & Lee, C. H. (2024). lubridate : Make Dealing with Dates a Little Easier (Version 1.9.4) [Logiciel]. <https://cran.r-project.org/web/packages/lubridate/index.html>

Sulandari, W., & Rodrigues, P. C. (2021). Exponential Smoothing on Modeling and Forecasting Multiple Seasonal Time Series : An Overview. *Fluctuation and Noise Letters*, 20(04), 2130003. <https://doi.org/10.1142/S0219477521300032>

Tătaru, I.-M., Fleacă, E., & Fleacă, B. (2020). Insights on the impact of telecommunication companies on the environment. *Proceedings of the International Conference on Business Excellence*, 14(1), 202-213. <https://doi.org/10.2478/picbe-2020-0019>

Taylor, S., & Letham, B. (2021). prophet : Automatic Forecasting Procedure (Version 1.0) [Logiciel]. <https://cran.r-project.org/web/packages/prophet/index.html>

Venkatesan, R., & Kumar, V. (2002). A genetic algorithms approach to growth phase forecasting of wireless subscribers. *International Journal of Forecasting*, 18(4), 625-646. [https://doi.org/10.1016/S0169-2070\(02\)00070-5](https://doi.org/10.1016/S0169-2070(02)00070-5)

Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. *Results in Control and Optimization*, 14, 100342. <https://doi.org/10.1016/j.rico.2023.100342>

Wickham, H. (2016). Programming with ggplot2. In H. Wickham (Éd.), *Ggplot2 : Elegant Graphics for Data Analysis* (p. 241-253). Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4_12

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., Brand, T. van den, Posit, & PBC. (2024). ggplot2 : Create Elegant Data Visualisations Using the Grammar of Graphics (Version 3.5.0) [Logiciel]. <https://cran.r-project.org/web/packages/ggplot2/index.html>

Wieck, R., & Vidal, M. (2011). Investment in telecommunications infrastructure, growth and employment recent research. *International Journal of Management and Network Economics*, 2(2), 135. <https://doi.org/10.1504/IJMNE.2011.043351>