

ORIGINAL

Regression Models for the Analysis of Telecommunications Data in Ecuador

Modelos de Regresión para el Análisis de Datos de Telecomunicaciones en Ecuador

Jaime G. Alvarado¹  , Edgar V. Lema¹  , Luis A. Cuaical²  , Alexis D. Alvarado³  

¹Universidad Técnica del Norte, Facultad en Ciencias Aplicadas. Ibarra, Ecuador.

²Universidad Católica del Ecuador, Facultad de Ciencias Exactas y Naturales. Quito, Ecuador.

³Universidad Internacional SEK Ecuador, Posgrado. Quito, Ecuador.

Citar como: Alvarado JG, Lema EV, Cuaical LA, Alvarado AD. Regression Models for the Analysis of Telecommunications Data in Ecuador. Data and Metadata. 2025; 4:769. <https://doi.org/10.56294/dm2025769>

Enviado: 11-05-2024

Revisado: 14-09-2024

Aceptado: 29-03-2025

Publicado: 30-03-2025

Editor: Dr. Adrián Alejandro Vitón Castillo 

Autor para la correspondencia: Jaime G. Alvarado 

ABSTRACT

The research highlights the importance of mathematical models for better planning in both state and private companies, specifically through curvilinear regressions, to forecast future activities of the State Telecommunications Regulation and Control Agency of Ecuador (ARCOTEL), by analyzing variables such as the number of internet service users. The study was based on data preprocessing, which included homogeneity analysis and scale changes. Statistical tests such as the Mann-Kendall Test and the Helmert Test were applied to evaluate trends in time series. Subsequently, the data were fitted from a linear model to a polynomial one. Evaluation metrics included absolute, mean, and quadratic percentage errors, as well as coefficients of determination and correlation. The analysis showed that the sixth-degree polynomial fitting provided an adequate adjustment for the time series, with high correlation coefficients and relatively low absolute and mean percentage errors, suggesting acceptable accuracy between the fitted and actual values. Scaling the data facilitated comparison and analysis, eliminating biases. The research emphasized the importance of effective planning using mathematical models to predict economic activity in companies. The sixth-degree polynomial fitting proved to be effective in representing time series, with low errors and high accuracy. These methods were useful for planning and forecasting in the telecommunications sector, as exemplified by the analysis of ARCOTEL users.

Keywords: Mathematical Models; Curvilinear Regressions; Polynomial Fitting; Data Normalization.

RESUMEN

Se pone como antecedente de la investigación la importancia de los modelos matemáticos para una mejor planificación en empresas estatales y privadas, específicamente mediante regresiones curvilíneas, para pronosticar actividades futuras de la Agencia Estatal de Regulación y Control de las Telecomunicaciones del Ecuador (ARCOTEL) para analizar variables como el número de usuarios de servicios de internet. El estudio se basó en el preprocesamiento de datos, que incluyó análisis de homogeneidad y cambios de escala. Se aplicaron pruebas estadísticas como el Test de Mann-Kendall y la prueba de Helmert para evaluar tendencias en series de tiempo. Luego, se ajustaron los datos desde un modelo lineal hasta un polinomial. Las métricas de evaluación incluyeron errores porcentuales absolutos, medios y cuadráticos, así como coeficientes de determinación y correlación. El análisis mostró que el ajuste polinomial de grado 6 proporcionó un ajuste adecuado de las series de tiempo, con altos coeficientes de correlación, errores porcentuales absolutos y medios relativamente bajos, lo que sugirió una precisión aceptable entre los valores ajustados y los valores reales. El cambio de escala de los datos facilitó la comparación y el análisis, eliminando sesgos. La investigación subrayó la importancia de la planificación efectiva utilizando modelos matemáticos para

predecir la actividad económica en empresas. El ajuste polinomial de grado 6 demostró ser efectivo en la representación de series temporales, con bajos errores y alta precisión. Estos métodos resultaron útiles para la planificación y pronóstico en el sector de telecomunicaciones, ejemplificado por el análisis de usuarios de ARCOTEL.

Palabras clave: Modelos Matemáticos; Regresiones Curvilíneas; Ajuste Polinomial; Estandarización de Datos.

INTRODUCCIÓN

Un aspecto fundamental en la efectiva planificación de empresas, ya sean estatales o privadas, radica en la capacidad para pronosticar y anticipar su actividad económica.⁽¹⁾ Para alcanzar este objetivo, es esencial recopilar y almacenar información fidedigna a lo largo del tiempo y emplear un modelo de ajuste capaz de generar estimaciones cercanas a los valores reales.⁽²⁾ Aunque es válido realizar el estudio de pronósticos mediante métodos cualitativos, en este análisis específico se opta por utilizar enfoques cuantitativos, motivados por varias razones.

La disponibilidad de datos confiables de la Agencia Estatal de Regulación y Control de las Telecomunicaciones (ARCOTEL) abarca un período de monitoreo de 11 años, incluyendo mediciones de diversos parámetros, como el número de usuarios de diferentes operadoras que ofrecen servicios de internet mediante diversas tecnologías. Estos datos permiten la realización de un análisis estadístico, examinando la relación entre la variable dependiente (el número de usuarios de las distintas operadoras) y la variable independiente (el tiempo).⁽³⁾ La distribución de datos a lo largo del tiempo sugiere que el comportamiento de la variable sigue un patrón a medida que transcurre el tiempo, facilitando el desarrollo de un modelo al final del proceso que permite el pronóstico buscado. Para el análisis de las series de tiempo, se examina la tendencia, la variación cíclica, la variación estacional y la variación aleatoria.

Aunque existen métodos como las Redes Neuronales o la aplicación de técnicas como los promedios móviles, suavización exponencial o el método de Box-Jenkins para el análisis de series temporales, en este estudio se opta por emplear los métodos de regresión polinomial. Esta elección se basa en la evaluación de la efectividad de estos métodos para identificar el patrón subyacente en las series temporales, motivada por razones específicas.⁽⁴⁾

Los modelos polinomiales resultan ser más simples y fáciles de aplicar e interpretar en comparación con métodos más complejos como las redes neuronales o los modelos autoregresivos integrados de media móvil, tanto por operadoras estatales como privadas en el ámbito de las telecomunicaciones. Además, estos modelos presentan una notable flexibilidad y capacidad de adaptación a una amplia variedad de patrones de datos, siendo idóneos para capturar la complejidad inherente en las series temporales.

En las series temporales relacionadas con las telecomunicaciones, la forma específica de la curva adquiere relevancia práctica y contribuye significativamente a la interpretación de los resultados de la tendencia observada. La selección del mejor modelo se realiza mediante criterios de decisión establecidos a través de una exhaustiva evaluación de estos métodos, garantizando así la elección del enfoque más apropiado para el análisis de las series temporales en el contexto específico de las telecomunicaciones.⁽⁵⁾

MÉTODO

La metodología utilizada para analizar las series temporales de líneas activas en modalidades de Prepago, basándose en los datos proporcionados por ARCOTEL, se desarrolló de manera sistemática. El proceso se llevó a cabo siguiendo las siguientes etapas:

Pre-procesamiento de Datos

Esta etapa fue crucial para garantizar la disponibilidad y confiabilidad de los datos. Se realizó un análisis descriptivo de la base de datos de ARCOTEL para identificar posibles datos anómalos, asegurando que los datos utilizados en el análisis fueran confiables y libres de errores.⁽⁶⁾

Análisis de Homogeneidad con Series Temporales

Se emplearon pruebas estadísticas no paramétricas para identificar errores y determinar la homogeneidad de las series. Se utilizaron el Test de Mann-Kendall y la prueba estadística de Helmert.⁽⁷⁾

Test de Mann-Kendall

Es una prueba estadística no paramétrica utilizada para identificar tendencias en series temporales. Evalúa la existencia de una relación monótona entre los valores de la serie temporal y el tiempo, sin asumir una distribución específica de los datos. La prueba calcula la varianza (V) y la compara con un umbral para determinar

la significancia de la tendencia observada. Es especialmente útil para detectar aumentos o disminuciones sistemáticas a lo largo del tiempo, lo que la hace ideal para estudios de series temporales con datos hidrológicos y climáticos donde se desea identificar tendencias en datos observados.⁽⁸⁾ Además, el test de Mann-Kendall se presenta robusto frente a datos con distribuciones no normales y puede manejar series temporales con valores ausentes, aumentando su aplicabilidad en diversos campos científicos.⁽⁹⁾

Prueba estadística de Helmer

Esta prueba se utiliza para detectar cambios estructurales en la serie temporal y determinar si existe una relación lineal entre los valores del número de usuarios de las diferentes operadoras a lo largo del tiempo. La serie se clasifica como homogénea o no homogénea según la diferencia entre la suma de los signos y la raíz cuadrada. La prueba de Helmer se emplea en modelos de Gauss-Helmer para la estimación de mínimos cuadrados, proporcionando una base robusta para el análisis estadístico en procesos de predicción, también se utiliza en el análisis de contrastes, facilitando la evaluación de predicciones teóricas sobre diferencias entre medias de grupos en datos empíricos.⁽¹⁰⁾

Cambio de Escala de Datos

En lugar de una estandarización tradicional, se realizó un cambio de escala de los datos a un rango entre 0 y 1. Este proceso, conocido como normalización Min-Max, es crucial cuando se aplican métodos de mínimos cuadrados en polinomios de grado mayor a 6. La razón principal es que, al manejar valores muy grandes, los coeficientes en los polinomios elevados a potencias altas pueden llegar a ser extremadamente grandes, lo cual dificulta el cálculo y puede llevar a problemas de precisión numérica. Al transformar tanto la variable independiente (tiempo) como la variable dependiente (usuarios) a una escala entre 0 y 1, se asegura que todos los valores permanezcan dentro de un rango manejable, facilitando la comparación y el análisis.⁽¹¹⁾

Ajuste Polinomial de Datos

El ajuste polinomial de datos se realiza para capturar la relación no lineal entre la variable independiente (tiempo) y la variable dependiente (usuarios). Este proceso implica encontrar los coeficientes de una ecuación polinómica que mejor se ajusten a los datos observados. El objetivo es minimizar la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo polinomial.

El ajuste polinomial puede involucrar varios grados del polinomio, desde un simple ajuste lineal hasta polinomios de grado mayor. En este estudio, se consideraron modelos desde un ajuste lineal hasta un polinomio de grado 6 para obtener una representación ajustada de las series de tiempo. A medida que se incrementa el grado del polinomio, la capacidad del modelo para ajustarse a las complejidades de los datos también aumenta, aunque con el riesgo de sobreajuste (overfitting), donde el modelo se ajusta demasiado a las variaciones específicas de los datos de entrenamiento, perdiendo su capacidad predictiva para nuevos datos.

El proceso de ajuste polinomial se lleva a cabo utilizando el método de mínimos cuadrados, que busca los coeficientes que minimicen la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos. Este método es fundamental en la regresión polinomial, ya que permite determinar los coeficientes de la ecuación polinómica que mejor describen la relación entre las variables.⁽¹²⁾

Análisis del Ajuste

Para evaluar la calidad del ajuste y la capacidad predictiva del modelo polinomial, se emplearon varias métricas clave. El Mean Absolute Percentage Error (MAPE) mide el error promedio absoluto como un porcentaje de los valores reales, siendo útil para comparar diferentes modelos o conjuntos de datos, el Mean Absolute Error (MAE) calcula el promedio de las diferencias absolutas entre los valores reales y los predichos, proporcionando una medida directa de precisión, el Mean Squared Error (MSE) promueve los cuadrados de las diferencias entre los valores reales y los predichos, penalizando más los errores grandes, y su raíz cuadrada, el Root Mean Squared Error (RMSE), permite interpretar los errores en la misma escala que los valores originales, facilitando la comparación.⁽¹³⁾ Además, el Coeficiente de Determinación (R^2) indica la proporción de la varianza en los datos dependientes que es predecible a partir de las variables independientes, mientras que el Coeficiente de Correlación (R) mide la fuerza y dirección de la relación lineal entre dos variables, proporcionando una evaluación completa de la efectividad del modelamiento polinomial en el análisis de series temporales.⁽¹⁴⁾

RESULTADOS Y DISCUSIÓN

Pre-procesamiento de Datos

El pre-procesamiento de datos se aplicó a los usuarios de las líneas activas en las dos modalidades (prepago y pospago) y para las tres operadoras CONECEL, OTECEL y CNT, el análisis se lo ha realizado en base a la estadística descriptiva que se resume en la siguiente tabla.

Tabla 1. Análisis estadístico de las operadoras en la modalidad de prepago y pospago

Estadística descriptiva	Pospago			Prepago		
	CONECEL	OTECCEL	CNT	CONECEL	OTECCEL	CNT
Media	7 443 458	3 521 820	1 141 387	2 091 800	1 028 243	321 702
Error típico	123 193	34 457	76 564	40 402	21 616	15 041
Mediana	6 757 632	3 405 847	546 729	2 216 863	1 159 541	325 673
Desviación estándar	1 606 241	449 270	998 272	526 776	281 834	196 113
Varianza	2,5800E+12	2,0184E+11	9,9655E+11	2,7749E+11	7,9430E+10	3,8460E+10
Curtosis	-2	-1	-2	0	-1	-1
Coefficiente de asimetría	0	0	0	-1	-1	0
Rango	4 598 872	1 602 667	2 598 440	1 751 543	898 193	642 217
Mínimo	5 327 336	2611348	121 410	928 531	468 235	48 961
Máximo	9 926 208	4 214 015	2 719 850	2 680 074	1 366 428	691 178
Suma	1 265 387 891	598 709 400	194 035 844	355 605 990	174 801 323	54 689 338
Cuenta	170	170	170	170	170	170

Análisis Estadístico Descriptivo

Para la modalidad de pospago, CONECEL presenta el mayor número promedio de usuarios, seguida por OTECEL y luego CNT, que tiene significativamente menos usuarios en comparación. La variabilidad en el número de usuarios es mayor en CONECEL, indicando una dispersión más amplia de los datos alrededor de la media. En términos de curtosis y asimetría, las distribuciones para las tres operadoras son más planas de lo esperado en una distribución normal y son simétricas. El rango de usuarios es también mayor en CONECEL, mostrando la mayor variabilidad entre el mínimo y el máximo número de usuarios. Para la modalidad de prepago, CONECEL nuevamente lidera con el mayor número promedio de usuarios, seguido por OTECEL y CNT. La variabilidad es también mayor en CONECEL, sugiriendo una mayor dispersión en los datos. Las distribuciones en CONECEL y OTECEL son más planas y tienden hacia la izquierda, mientras que la de CNT es más simétrica. El rango de usuarios en prepago es mayor en CONECEL, lo que indica una variabilidad más alta.

Interpretación para Ajuste Polinomial

La alta variabilidad observada en CONECEL sugiere que cualquier modelo de ajuste polinomial debe ser capaz de capturar esta dispersión para ser efectivo. La menor variabilidad en OTECEL y CNT implica que sus modelos podrían ser más simples y directos. Las diferencias en las tendencias centrales, como la media y la mediana, indican posibles sesgos en los datos que deben ser considerados durante el modelado. Las distribuciones planas y asimétricas en algunas operadoras sugieren la presencia de valores extremos, que pueden influir en la elección del grado del polinomio para evitar el sobreajuste o subajuste. Este análisis proporciona una base sólida para la aplicación de modelos polinomiales, asegurando que las características clave de los datos sean adecuadamente consideradas para lograr un ajuste preciso y efectivo.

Análisis de Homogeneidad con Series Temporales

Test de Mann-Kendall

El análisis de homogeneidad se realizó utilizando el test de Mann-Kendall para evaluar la presencia de tendencias significativas en las series temporales de las operadoras CONECEL, OTECEL y CNT en las modalidades de prepago y pospago. Como se indicó anteriormente el test de Mann-Kendall es una prueba no paramétrica utilizada para detectar tendencias en una serie temporal y se basa en la siguiente fórmula:

$$V = \frac{S - 1}{\sqrt{\frac{n(n-1)(2n+5)}{18}}}$$

Donde:

$$S = T - I$$

T es la cantidad de eventos mayores del valor que se analiza en la serie de tiempo.

I es la cantidad de eventos menores del valor que se analiza en la serie de tiempo.

A continuación, se presentan la tabla de resultados del Test de Mann Kendall.

Tabla 2. Test de Mann Kendall en los datos en la modalidad de prepago y postpago

Test de mann-kendall	Prepago			Postpago		
	CONECEL	OTECCEL	CNT	CONECEL	OTECCEL	CNT
Alfa	10 %	10 %	10 %	10 %	10 %	10 %
V_{crit}	1,28	1,28	1,28	1,28	1,28	1,28
I	120	120	120	120	119	120
T	120	120	120	120	119	120
S	0	0	0	0	0	0
V	-0,0376	-0,0376	-0,0376	-0,0376	-0,0376	-0,0376
$ V < V_{crit} $	Homogenea	Homogenea	Homogenea	Homogenea	Homogenea	Homogenea

Del análisis de la tabla se observa que el nivel de significancia utilizado fue del 10 % (alfa = 0,1) para cada variable. El valor crítico asignado ($V_{crit}=1,28$) se utilizó para determinar si el valor absoluto de la estadística de prueba ($V= -0,0376$) es menor que el valor crítico ($|V| < |V_{crit}|$). Los resultados indican que el test de Mann-Kendall aplicado a las series temporales de usuarios de las operadoras CONECEL, OTECEL y CNT en las modalidades de prepago y postpago, así como el total de usuarios de telefonía móvil, ha demostrado que las series son homogéneas. Esto indica que no se encontraron tendencias significativas en los datos, lo que es un hallazgo relevante para la planificación y predicción de la actividad económica en el sector de telecomunicaciones.

Prueba estadística de Helmer

Para corroborar el test anterior se procedió con la prueba estadística de Helmer que resultó mucho más fácil que el Test de Mann Kendal y consistió en analizar el signo de las desviaciones de cada evento de la serie con respecto de su valor medio, en el cual si una desviación de un cierto signo es seguida de otro de un mismo signo se creó una secuencia S1, en contraste si una desviación es seguida por otra de signo contrario se registró como un cambio C, y se comparó con la raíz cuadrada de (n-1), para clasificar la serie como homogénea o no homogénea, según las fórmulas que se indican a continuación.

$$S1 - C > \sqrt{n-1} \Rightarrow \text{SERIE HOMOGÉNEA}$$

$$S1 - C < \sqrt{n-1} \Rightarrow \text{SERIE NO HOMOGÉNEA}$$

Los resultados de este proceso se indican en la siguiente tabla.

Tabla 3. Prueba estadística de Helmer a los datos en la modalidad de prepago y postpago

Parametros promedio	Prepago			Postpago		
	CONECEL	OTECCEL	CNT	CONECEL	OTECCEL	CNT
S1=	13	12	14	14	14	13
C=	2	3	1	1	1	2
	Homogenea	Homogenea	Homogenea	Homogenea	Homogenea	Homogenea

Del análisis de la tabla, se observa que la prueba de Helmer compara los valores anuales de los datos en la modalidad de prepago y postpago con un valor promedio calculado para determinar si existen cambios significativos que indiquen rupturas en la estructura de la serie. Los resultados obtenidos mostraron que la mayoría de los años analizados son homogéneos, es decir, no presentan cambios estructurales significativos. Esto se evidencia en los valores tabulados, donde se observa que la mayoría de los años presentan un signo 'S', indicando estabilidad en la serie, por lo que la serie es considerada homogénea porque los valores de la prueba se mantienen consistentes, como se observa en la mayoría de los años. Los pocos casos marcados con 'C' (cambio) indican años con variaciones significativas, pero no suficientes para alterar la homogeneidad global de la serie y los cambios son pocos y están dentro de los límites de un error probable.

Relación con el Test de Mann-Kendall

La prueba de Helmer y el test de Mann-Kendall trabajan en conjunto para proporcionar una visión completa de la estructura de las series temporales. Mientras que el test de Mann-Kendall identifica tendencias generales (monotónicas) en la serie temporal, la prueba de Helmer se centra en detectar cambios estructurales específicos que podrían indicar rupturas en la serie.

En este estudio, ambos tests indicaron homogeneidad en las series temporales analizadas. Esto significa que las series no solo no tienen tendencias significativas (según el test de Mann-Kendall), sino que también mantienen una estructura consistente a lo largo del tiempo (según la prueba de Helmert). La confirmación de la homogeneidad por ambos tests asegura que las series temporales son estables y adecuadas para el análisis posterior.

Contribución al Ajuste Polinómico

La homogeneidad detectada por la prueba de Helmert es crucial para el ajuste polinómico. La estabilidad estructural de las series temporales permite aplicar modelos polinomiales con confianza, ya que no se esperan cambios bruscos o tendencias ocultas que puedan distorsionar los resultados. Los modelos polinomiales pueden así capturar la relación subyacente entre el tiempo y el número de usuarios de servicios de las operadoras, proporcionando estimaciones precisas y fiables.

Por lo que se puede resumir que la prueba de Helmert confirma que las series temporales son adecuadas para el ajuste polinómico, validando su uso en la planificación y predicción de tendencias en el sector de telecomunicaciones. La estabilidad y homogeneidad de las series temporales garantizan que los modelos polinomiales ajustados reflejarán fielmente la dinámica de los datos, contribuyendo a una mejor toma de decisiones estratégicas.

Cambio de escala de datos

El proceso de normalización Min-Max de datos realizado en el estudio es una práctica común y útil para comparar variables en diferentes escalas. Al transformar las variables a una escala entre 0 y 1, se logra una comparación relativa entre las diferentes modalidades de pago y operadoras, sin verse afectada por las diferencias absolutas de sus magnitudes.

Esta normalización de datos permite eliminar sesgos causados por las diferencias en las unidades de medida y rangos de las variables, lo cual facilita la comparación y el análisis de los datos. Al asignar el valor 0 al mínimo y el valor 1 al máximo en cada modalidad de pago, se establece un marco de referencia común para evaluar el comportamiento de los usuarios en relación con el tiempo y las distintas modalidades de pago, tal como podemos observar en la tabla siguiente.

Para el presente análisis se asignó para la variable tiempo con 0 al primer valor y con 1 a los 5072 días que corresponde al último valor, esto para las 2 modalidades prepago y pospago.

Para la variable usuario, se asignó con 1 a los máximos valores de cada modalidad es decir a los 9926208 para prepago y 2680074 para pospago y en función de estos se completó la tabla para cada modalidad, tal como se indica a continuación.

Tabla 4. Estandarización de los datos en la modalidad de prepago y pospago						
Tiempo	Pospago			Prepago		
	CONECEL	OTECCEL	CNT	CONECEL	OTECCEL	CNT
0	0,34646	0,17611	0,01924	0,72490	0,26702	0,02536
0,01	0,34963	0,17620	0,02439	0,73725	0,26308	0,02418
0,01	0,35282	0,17471	0,02461	0,74657	0,26406	0,02418
0,02	0,35667	0,20377	0,02461	0,75324	0,26501	0,02418
0,02	0,35912	0,20559	0,02390	0,76038	0,26503	0,02363
0,03	0,36447	0,17723	0,02240	0,76806	0,27697	0,02434
0,04	0,36782	0,17967	0,02245	0,77334	0,27927	0,02541
0,04	0,37165	0,18237	0,02257	0,77879	0,28401	0,02589
0,97	0,80482	0,45424	0,11493	0,67487	0,41559	0,26618
0,98	0,80962	0,45567	0,11399	0,67805	0,41930	0,26774
0,98	0,81349	0,45691	0,11384	0,68086	0,42219	0,26919
0,99	0,81775	0,45866	0,11389	0,68184	0,42347	0,27075
0,99	0,82250	0,46050	0,11392	0,68741	0,42453	0,27229
1	0,82699	0,46312	0,11365	0,69080	0,41987	0,27401

Este tratamiento, asegura que todos los valores permanecen dentro de un rango manejable, facilitando la comparación y el análisis. Al transformar tanto la variable independiente (tiempo) como la variable dependiente (usuarios) a una escala entre 0 y 1, se establece un marco de referencia común. Esto elimina sesgos causados

por las diferencias en las unidades de medida y rangos de las variables, permitiendo una evaluación precisa y consistente del comportamiento de los usuarios a lo largo del tiempo y entre diferentes modalidades de pago.

La normalización de los datos facilita la aplicación de modelos polinomiales, ya que, al mantener los valores dentro de un rango uniforme, se reduce el riesgo de obtener coeficientes extremadamente grandes que puedan llevar a problemas de precisión numérica. Esto es especialmente importante cuando se trabaja con polinomios de grado mayor a 6. La normalización asegura que el modelo ajuste los datos de manera más eficiente y precisa, reflejando fielmente las tendencias y patrones en las series temporales.

Ajuste Polinomial de Datos.

Para el ajuste de los m datos realizado en el estudio se utilizó la aproximación polinomial con mínimos cuadrados, el método consiste en que consiste en pedir que la suma de las distancias calculadas entre el valor de la función de aproximación $p(x_i)$ y el valor de la función $f(x_i)$ calculada sea mínima, es decir:

$$\sum_{i=1}^m (p(x_i) - f(x_i)) = \sum_{i=1}^m d_i = \text{mínimo}$$

Si asumimos una regresión lineal con $p(x) = a_0 + a_1 x$, elevamos al cuadrado la distancia d_i para evitar problemas de derivabilidad:

$$\sum_{i=1}^m (a_0 + a_1 x_i - f(x_i))^2 = \sum_{i=1}^m d_i^2 = \text{mínimo}$$

Para determinar el polinomio $p(x)$, derivamos parcialmente respecto a las dos variables a_0 y a_1 : Primero derivamos respecto a a_0 :

$$\frac{\partial}{\partial a_0} \sum_{i=1}^m (a_0 + a_1 x_i - f(x_i))^2 = 0$$

Simplificando:

$$\sum_{i=1}^m (a_0 + a_1 x_i - f(x_i)) = 0$$

Aplicando las propiedades de las sumatorias:

$$n a_0 + a_1 \sum_{i=1}^m x_i = \sum_{i=1}^m f(x_i)$$

Luego derivamos respecto a a_1 .

$$\frac{\partial}{\partial a_1} \sum_{i=1}^m (a_0 + a_1 x_i - f(x_i))^2 = 0$$

Simplificando:

$$\sum_{i=1}^m (a_0 x_i + a_1 x_i^2 - f(x_i) x_i) = 0$$

Aplicando las propiedades de las sumatorias:

$$a_0 \sum_{i=1}^m x_i + a_1 \sum_{i=1}^m x_i^2 = \sum_{i=1}^m f(x_i) x_i$$

El Sistema resultante es:

$$ma_0 + a_1 \sum_{i=1}^m x_i = \sum_{i=1}^m f(x_i)$$

$$a_0 \sum_{i=1}^m f(x_i) + a_1 \sum_{i=1}^m x_i^2 = \sum_{i=1}^m f(x_i) x_i$$

Al resolver se encuentra el polinomio de grado 1: $p(x)=a_0+a_1x$

En la presente investigación al no obtener un buen ajuste con un polinomio de grado 1, se procedió a utilizar el mismo método de regresión polinomial hasta alcanzar un polinomio de grado, n por ejemplo, el procedimiento se basa en minimizar la función:

$$\sum_{i=1}^m [a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_n x_i^n - f(x_i)]^2 = 0$$

Si derivamos parcialmente respecto a los coeficientes a_j con $(0 \leq j \leq n)$, e igualamos a cero cada una de las derivadas, obtenemos el siguiente sistema de ecuaciones.

$$ma_0 + a_1 \sum x + a_2 \sum x^2 + \dots + a_n \sum x^n = \sum y$$

$$a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 + \dots + a_n \sum x^{n+1} = \sum xy$$

$$a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 + \dots + a_n \sum x^{n+2} = \sum x^2 y$$

$$a_0 \sum x^n + a_1 \sum x^{n+1} + a_2 \sum x^{n+2} + \dots + a_n \sum x^{n+n} = \sum x^n y$$

Donde se han omitido los subíndices de las variables y las sumatorias para simplificar la escritura.

En la presente investigación se realizó el análisis no lineal en las series de tiempo de las líneas activas, utilizando polinomios desde $n=1$ hasta $n=6$.

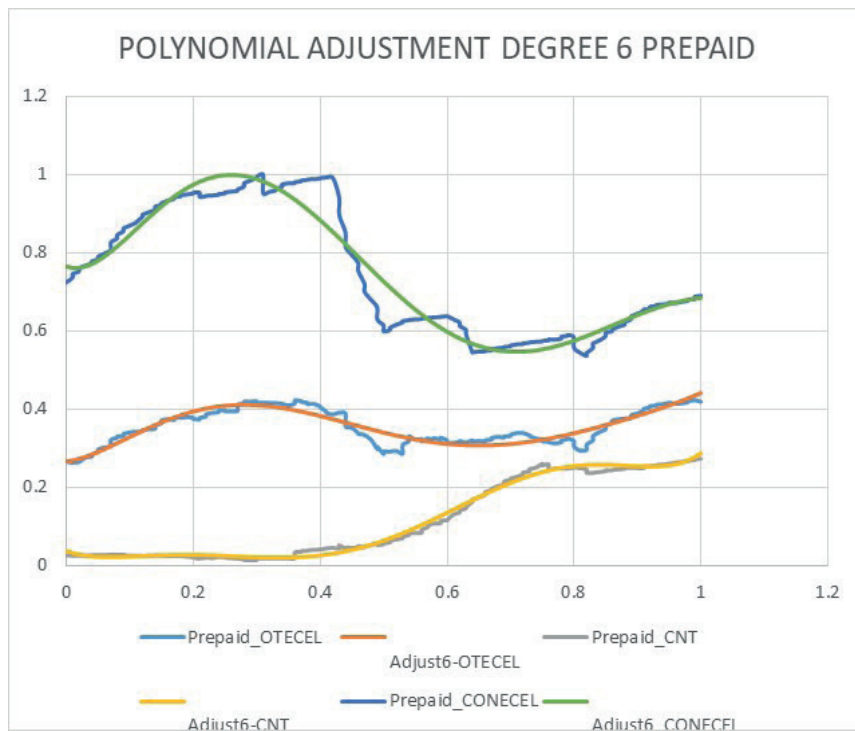


Figura 1. Ajuste Polinomial de los datos en la modalidad de prepago

El grado del polinomio fue aumentándose y el mejor ajuste gráfico se obtuvo con un polinomio de grado 6. Este análisis se basó en la adecuación del ajuste a los datos, la interpretación de los resultados y el análisis posterior del error, que se indica más adelante.

El ajuste polinomial de grado 6 mostró una alta correlación con los datos observados, indicando un buen ajuste del modelo. Los errores porcentuales absolutos y medios fueron relativamente bajos, sugiriendo una precisión aceptable entre los valores ajustados y los valores reales. Esta precisión y el buen ajuste se lograron mediante la transformación de los datos a una escala manejable entre 0 y 1, lo cual facilitó el análisis y la comparación sin introducir sesgos debidos a las magnitudes originales de los datos.

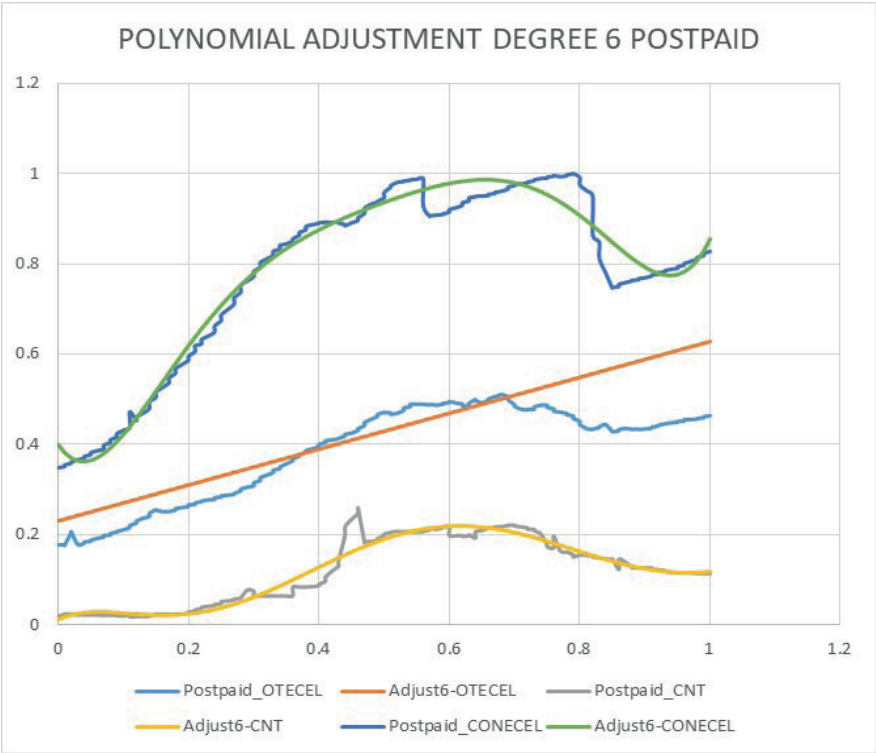


Figura 2. Ajuste Polinomial de los datos en la modalidad pospago

Análisis del ajuste

El análisis del ajuste polinomial de los datos presentados en la tabla muestra los coeficientes (a_0 , a_1 , a_2 , a_3 , a_4 , a_5 , a_6) obtenidos a partir del ajuste del modelo polinomial a las diferentes series de tiempo de las modalidades de Prepago y pospago de las distintas operadoras. Los resultados de este ajuste se resumen en la siguiente tabla.

Tabla 5. Coeficientes, correlación y errores medios en la modalidad de prepago y pospago						
n	Prepago- CONECEL	Prepago- OTECCEL	Prepago- CNT	Pospago- CONECEL	Pospago- OTECCEL	Pospago- CNT
a0	0,7652	0,2667	0,036	0,036	0,1663	0,011
a1	-0,5894	0,2236	-0,5954	-0,5954	0,8163	0,6641
a2	22,286	6,984	7,4723	7,4723	-5,0621	-8,5855
a3	-98,698	-35,314	-37,699	-37,699	23,535	40,534
a4	163,27	62,838	85,142	85,142	-41,972	-74,146
a5	-117,31	-48,45	-83,73	-83,73	30,362	58,229
a6	30,961	13,893	29,659	29,659	-7,393	-16,589
R ²	0,9349	1	0,9917	0,9917	0,9946	0,944
MAPE	3,84	2,3	13,69	3,42	2,3	12,6
MAE	0,0286	0,008	0,007	0,026	0,008	0,011
MSE	0,002	1,00E-04	8,30E-05	0,001	1,00E-04	3,00E-04

El análisis de los datos demuestra que el ajuste polinomial de grado 6 proporciona un modelo robusto para las series de tiempo de usuarios en las modalidades de prepago y pospago para las distintas operadoras. La alta

precisión del modelo se refleja en los bajos valores de MAPE, MAE y MSE, así como en los altos valores de R^2 , lo que sugiere que el modelo es adecuado para describir y predecir las tendencias observadas en los datos.

Los coeficientes de determinación (R^2) indicados en la tabla son altos, lo que sugiere un buen ajuste del modelo polinomial a los datos. Estos valores oscilan entre 0,9349 y 1, indicando que el modelo polinomial de grado 6 puede explicar entre el 93,49 % y el 100 % de la variabilidad en los datos observados.

Los valores más bajos del (Error Porcentual Absoluto Medio) MAPE indican una mayor precisión del modelo. En este estudio, los valores oscilan entre 2,3 % y 13,69 %, lo que sugiere que el modelo tiene una precisión porcentual aceptable en la mayoría de las series de tiempo.

El ((Error Absoluto Medio) MAE bajo indica una menor diferencia entre los valores ajustados y los valores reales. Los valores obtenidos varían entre 0,007 y 0,0286, sugiriendo un buen ajuste del modelo a los datos para un polinomio de grado 6.

En este indicador en cambio se observan valores más bajos para el (Error Absoluto Medio) MAE, lo que indican una menor diferencia entre los valores ajustados y los valores reales, lo que sugiere un buen ajuste del modelo a los datos siendo igualmente el mejor ajuste para un polinomio de grado 6.

Los valores del (Error Cuadrático Medio) MSE, que oscilan entre 1,00E-04 y 0,002, son relativamente bajos, indicando una menor discrepancia entre los valores ajustados y los valores reales, lo que confirma nuevamente un buen ajuste para un polinomio de grado 6.

Retorno predicción de datos futuros

Para la predicción de los valores futuros al tener las fórmulas de ajuste en función de los coeficientes que se indican en la tabla 5 se tiene las fórmulas que nos permiten predecir los valores futuros a manera de ejemplo para la Operadora CONECEL en la modalidad de pago prepago la fórmula de ajuste es: $y = 0,7652 - 0,5894x + 22,286x^2 - 98,698x^3 + 163,27x^4 - 117,31x^5 + 30,961x^6$.

Las predicciones obtenidas mediante las fórmulas permiten a la operadora anticipar la demanda futura de servicios prepago o pospago, facilitando así la planificación estratégica y operativa. Esto incluye la gestión de recursos, la planificación de infraestructura y la elaboración de estrategias de marketing dirigidas a satisfacer la demanda proyectada.

CONCLUSIONES

Este artículo subraya la importancia de una planificación efectiva para prever el número de usuarios, lo cual impacta directamente en el factor económico de empresas tanto estatales como privadas. El análisis de series temporales en telecomunicaciones requiere técnicas robustas para detectar tendencias y cambios estructurales en los datos. Metodologías estadísticas como el Test de Mann-Kendall y el contraste de Helmert proporcionan una base sólida para la interpretación y predicción de patrones temporales, facilitando la previsión de la evolución del número de usuarios de servicios de internet.

La estandarización de los datos mejora la comparación y el análisis, permitiendo una evaluación precisa de la planificación y el pronóstico en empresas, particularmente en el caso de ARCOTEL y su análisis de usuarios en telecomunicaciones. La utilización de diversos modelos matemáticos, incluidas las regresiones curvilíneas, ha demostrado ser fundamental en este proceso. Los resultados del estudio indican que el ajuste polinomial de grado 6 es adecuado para representar las series temporales, evidenciado por altos coeficientes de correlación que sugieren un buen ajuste del modelo. Los errores porcentuales absolutos y cuadrados medios son relativamente bajos, indicando una precisión aceptable entre los valores ajustados y reales.

REFERENCIAS BIBLIOGRÁFICAS

1. Bala R, Ojha DB. Modeling and forecasting Internet traffic using hybrid ARIMA-SVM approach. J Netw Comput Appl. 2021;182:103026. <https://doi.org/10.1016/j.jnca.2021.103026>
2. Li Y, Xie Z. Network traffic prediction based on a new model with cross-bidirectional long short-term memory. IEEE Trans Netw Serv Manag. 2019;16(3):801-811. <https://doi.org/10.1109/TNSM.2019.2902795>
3. Smith, J. A., & Lee, C. H. (2019). Analyzing the impact of socio-economic variables on public health outcomes through regression analysis. Public Health Journal, 45(3), 310-325. <https://doi.org/10.1093/pubmed/fdy091>
4. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 2003;50(3):159-175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
5. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. John Wiley & Sons; 2015. <https://doi.org/10.1002/9781118619193>

6. Hodge VJ, Austin J. A survey of outlier detection methodologies. *Artificial Intelligence Review*. 2004;22(2):85-126. <https://doi.org/10.1023/B:AIRE.0000045502.10941.a9>
7. Helsel DR, Frans LM. The regional Kendall test for trend. *Environmental Science & Technology*. 2006;40(13):4066-4073. <https://doi.org/10.1021/es051650b>
8. O'Connell RH, Coberly WT. The Helmert contrast as a trend test. *Biometrics*. 1975;31(2):301-309. <https://doi.org/10.2307/2529441>
9. Hirsch RM, Slack JR, Smith RA. Techniques of trend analysis for monthly water quality data. *Water Resources Research*. 1982;18(1):107-121. <https://doi.org/10.1029/WR018i001p00107>
10. Meng Q, Lee JH, Zhu S. Application of the Gauss-Helmert model for trend analysis in hydrological time series. *Journal of Hydrology*. 2018;558:85-98. <https://doi.org/10.1016/j.jhydrol.2018.01.015>
11. Han J, Pei J, Kamber M. *Data Mining: Concepts and Techniques*. Morgan Kaufmann; 2011. <https://doi.org/10.1016/C2009-0-61819-5>
12. Draper NR, Smith H. *Applied Regression Analysis*. John Wiley & Sons; 1998. <https://doi.org/10.1002/9781118625590>
13. Makridakis S, Wheelwright SC, Hyndman RJ. *Forecasting Methods and Applications*. John Wiley & Sons; 1998. <https://doi.org/10.1002/9781119263678>
14. Nagelkerke NJD. A note on a general definition of the coefficient of determination. *Biometrika*. 1991;78(3):691-692. <https://doi.org/10.1093/biomet/78.3.691>

FINANCIACIÓN

Ninguna.

CONFLICTO DE INTERESES

Ninguno.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: Luis Cuaical.

Curación de datos: Edgar V. Lema.

Análisis formal: Jaime Alvarado.

Investigación: Jaime Alvarado.

Metodología: Jaime Alvarado.

Administración del proyecto: Jaime Alvarado.

Software: Alexis D. Alvarado.

Supervisión: Jaime Alvarado.

Validación: Jaime Alvarado.

Visualización: Luis Cuaical.

Redacción - borrador original: Jaime Alvarado.

Redacción - revisión y edición: Jaime Alvarado.