

Análisis de demanda de ancho de banda en los servicios de internet de RedUNAM

Información del reporte:

Licencia Creative Commons



El contenido de los textos es responsabilidad de los autores y no refleja forzosamente el punto de vista de los dictaminadores, o de los miembros del Comité Editorial, o la postura del editor y la editorial de la publicación.

Para citar este reporte técnico:

Ramírez Fernández, E. R., Rivera Martínez, H. y Castañeda Ávila, L. I. (2025). Análisis de demanda de ancho de banda en los servicios de internet de RedUNAM. *Cuadernos Técnicos Universitarios de la DGTIC*, 3 (4) páginas (65 - 85). <https://doi.org/10.22201/dgtic.30618096e.2025.3.4.144>

Esteban Roberto Ramírez Fernández

Dirección General de Cómputo y de Tecnologías
de Información y Comunicación

Universidad Nacional Autónoma de México

esteban.ramirez@unam.mx

ORCID 0000-0002-2169-6233

Hugo Rivera Martínez

Dirección General de Cómputo y de Tecnologías
de Información y Comunicación

Universidad Nacional Autónoma de México

hugo.rivera@unam.mx

ORCID 0009-0007-1248-9412

Leonardo Isay Castañeda Ávila

Facultad de Estudios Superiores Aragón

Universidad Nacional Autónoma de México

leocastavi@gmail.com

ORCID 0009-0001-9257-5383

Resumen

El crecimiento en la demanda en los servicios de red ha elevado su importancia ya que éstos soportan procesos clave de las organizaciones, lo que crea la necesidad de contar con una perspectiva futura fiable para la contratación y renovación de servicios de internet. Este escenario y el surgimiento de diferentes modelos para el estudio y pronóstico de datos que permiten el análisis del comportamiento futuro usando información con estructura de series de tiempo, ha favorecido la implementación de estrategias que entregan resultados

en un menor tiempo. Tomando como modelo la mejora continua de las buenas prácticas de servicios de tecnologías de información, en el Departamento de Monitoreo de la Red, conocido como NOC RedUNAM, se propuso analizar el comportamiento de los datos recolectados referentes al consumo de ancho de banda de los enlaces de internet que integran la RedUNAM, los cuales cuentan con las características de series de tiempo. Con el objetivo de obtener información fiable del consumo futuro esperado de los servicios de internet, se probaron diferentes algoritmos de análisis que consideran las condiciones que afectan a la demanda de los servicios en la Universidad Nacional Autónoma de México (horarios, eventos extraordinarios, periodos escolares, entre otros). Considerando esta necesidad, se seleccionó el que mejor aceptó la influencia de los cambios de comportamiento cotidiano en los enlaces de datos. En consecuencia, se enfocó el análisis en algoritmos que dieran mayor importancia a las variables consideradas relevantes para el pronóstico de demanda de los enlaces de red, así como el comportamiento general de la operación de ésta. Para este ejercicio, se probaron distintos modelos, incluyendo regresión simple, promedios móviles integrados y otras técnicas. Finalmente, se seleccionó el algoritmo que ofreció el mejor margen de confiabilidad, permitiendo considerar aceptable el pronóstico para las condiciones de los datos de red recolectados

Palabras clave:

Análisis de datos, proyección de demanda, modelo de demanda, datos históricos.

Abstract

The growing demand for network services has raised their importance because they support organizations key processes, which creates the need for a reliable future perspective for the contracting and renewing internet services. This scenario, and the emergence of different models for data study and forecasting that allow the analysis of future behavior using information with time series data, has favored the implementation of strategies that deliver results in less time. Following the model of continuous improvement and good practices in information technologies services, the Departamento de Monitoreo de la Red, known as NOC RedUNAM, proposed the analysis of behavior of the collected data regarding the bandwidth consumption of internet links that integrate RedUNAM, which exhibit time series characteristics. With the aim of obtaining reliable information about future expected internet service consumption, different analysis algorithms that consider the conditions that affect the demand of services in the Universidad Nacional Autónoma de México (schedules, extraordinary events, school periods, among others) were tested. Based on this need, the algorithm that best accepted the influence of changes in data links daily behavior was selected. Consequently, the analysis was focused on algorithms that gave major importance to the variables considered relevant for the demand of network links forecast, as well as the network operation general behavior. For this exercise, simple regression models, integrated moving averages and other techniques were tested, concluding with the selection of an algorithm that offered a margin of reliability, allowing the forecast to be considered acceptable for the conditions of the collected network data.

Keywords:

Data analysis, demand projection, demand model, historical data.

1. INTRODUCCIÓN

Para llevar a cabo una gestión eficiente de servicios de Tecnologías de la Información y la Comunicación (TIC) en la Universidad Nacional Autónoma de México (UNAM), es necesaria la mejora continua en la implementación, operación y actualización de los mismos. En el caso de los servicios de red, la mejora en la gestión de los enlaces requiere implementar una estrategia para dimensionar la capacidad futura del ancho de banda que tiene cada enlace de internet. Esto es fundamental, debido a que dichos enlaces se contratan a proveedores externos de forma multianual y un correcto dimensionamiento permite aprovechar de mejor manera los recursos económicos que la UNAM destina a estos servicios.

Con el objetivo de dimensionar esta capacidad necesaria en los enlaces de internet y planear su ampliación antes de que la capacidad sea superada por la demanda, se ha buscado pronosticar de forma efectiva la demanda a futuro de las conexiones a internet de la UNAM, considerando que el costo asociado a los servicios, cuando éstos tienen grandes periodos de inactividad, crea una percepción de estar sobredimensionados.

Es importante resaltar que actualmente los enlaces de internet contratados a diferentes proveedores de servicios por la Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC) requieren un análisis individual, ya que cada uno de los contratos tiene diferente temporalidad y periodos de contratación.

Actualmente, dicha decisión suele sustentarse en diversas variables, tales como el crecimiento de la matrícula escolar en las escuelas, la demanda de mayor ancho de banda de las aplicaciones, el uso incorporado de servicios inalámbricos de acceso a la red, servicios en la nube o el acceso a recursos multimedia; para esto, se analiza mediante la observación visual del comportamiento en el tiempo.

El análisis de las variables se puede complicar debido al impacto de múltiples factores exógenos como periodos escolares, demandas excepcionales de los servicios, entre otras causas.

Al buscar resolver la influencia de todas estas variables sobre los datos y con la intención de crear una proyección de demanda que emplee datos históricos, se realizó una evaluación de los mecanismos más utilizados que permitieran apoyar en el pronóstico de los datos. El objetivo del presente reporte técnico es generar la proyección de la demanda de ancho de banda en enlaces de internet por medio de modelos y análisis de los datos de consumo que sean de utilidad en futuros procesos de toma de decisiones.

2. METODOLOGÍA

2.1 PROCESO DE RECOLECCIÓN DE DATOS EN EL SISTEMA CACTI

Para el desarrollo del análisis de los datos, se utiliza la información recolectada por el sistema de monitoreo Cacti, el cual emplea el protocolo SNMP (Protocolo simple de administración de red por sus siglas en inglés) de la siguiente forma:

1. El sistema de monitoreo realiza la recolección de los datos de utilización de ancho de banda consumido en cada ruteador y enlace de internet, el intervalo entre cada petición es de un minuto. La transmisión de datos desde el ruteador hacia el servidor que aloja el sistema Cacti se realiza

utilizando el SNMP.

2. El sistema de monitoreo trabaja con una base de datos de series de tiempo (guarda el valor y el momento que fue consultado) y se vincula a la gráfica a partir de una base de datos relacional (MariaDB), lo que permite su presentación vía web.
3. El tiempo de retención de los datos es diferente de acuerdo con la gráfica consultada. La base de datos que se usa es conocida como circular y tiene la bondad de que, una vez creada una gráfica, la base no seguirá creciendo, además de que promedia los datos en periodos de tiempo, los cuales son:
 - Monitoreo a 1 minuto con retención de 7 días
 - Monitoreo a 5 minutos con retención a 60 días
 - Monitoreo a 15 minutos con retención a 180 días
 - Monitoreo a 30 minutos con retención a 1 año

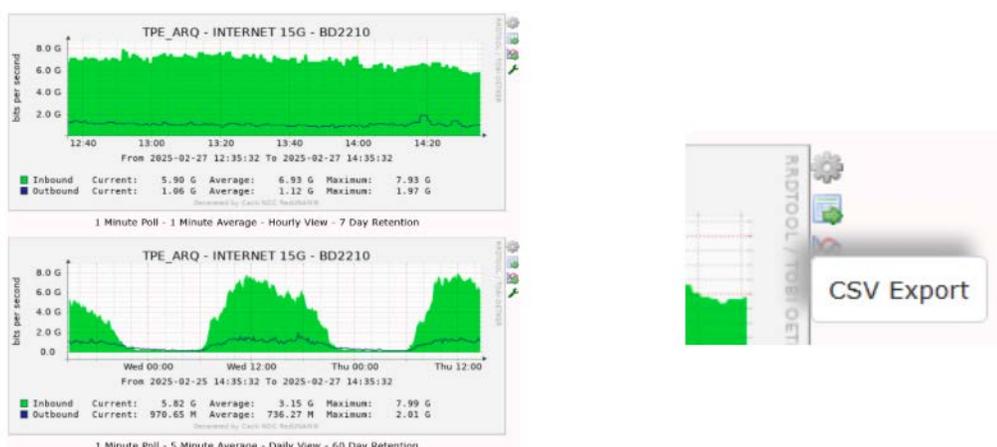
De esta forma, se obtiene una base de datos que no crece y puede administrarse dentro de un servidor con un disco duro de máximo 100 gigabytes, permitiendo conservar datos de hasta 1000 gráficas generadas.

Los datos de las gráficas se almacenan en archivos tipo RRD (*Roud Robin Database*). Posteriormente, esta información se utiliza para presentar las gráficas en el portal web del sistema de monitoreo Cacti.

Para la recuperación de estos datos, el sistema ofrece la posibilidad de exportar el conjunto de datos agrupados en formato CSV de forma gráfica por medio de un ícono en la parte superior derecha de cada gráfica, como se muestra en la Figura 1.

Figura 1

Gráfica de Cacti que muestra el ícono para exportar los valores



Adicionalmente, es posible obtener la información de cada una de las gráficas extrayendo el archivo individual de la base de datos desde el servidor vía el protocolo SSH (comunicación cifrada usando algoritmos ssl desde el servidor Cacti), que también permite exportar a formato separado por comas (CSV).

Para el análisis, se obtuvo la información en formato CSV y se procedió a aplicar las técnicas de filtrado y aplicación de algoritmos para evaluar resultados y seleccionar el algoritmo más útil.

2.2 FILTRADO

Como siguiente paso para validar que los datos utilizados en los pronósticos de la demanda de tráfico están completos, se validó la información recuperada en formato CSV, comparándola con la base de las gráficas en formato "round robin" (archivos de las gráficas en el servidor).

Una vez realizada la validación, se procedió a la limpieza y filtrado de los datos para trabajar con información que represente el comportamiento de la demanda sin la influencia de datos considerados producto de incidentes o anomalías en la recolección.

2.2.1 FILTRADO EMPLEANDO PYTHON

La siguiente fase del filtrado fue retirar datos innecesarios para el análisis. A continuación, se muestra un extracto de los datos de la base de una gráfica. Los datos en rojo se eliminaron y sólo se trabajó con los datos en verde recuperados del mismo archivo.

Las siguientes 11 líneas son una muestra de datos en formato CSV de una gráfica en el sistema de monitoreo Cacti:

Title	PRUEBAS RIU 2024 - HACIA RED UNAM - Gi0/0	
Vertical Label	bits per second	
Start Date	11/12/2024 9:34	
End Date	11/12/2024 11:34	
Step	60	
TotalRows	121	
Graph ID	11754	
Host ID	694	
Date	Inbound	Outbound
11/12/2024 9:34	523000000	120000
11/12/2024 9:34	N/A	N/A
...		...

2.2.2 FILTRADO DE VALORES

Para el análisis de una serie de tiempo, es muy importante la relación entre el tiempo y el dato consultado. Como existe la posibilidad de tener un valor fuera de rango por múltiples causas o incluso valores nulos, registrados en la base con un "N/A", se requirió filtrar cualquier valor no numérico y, en el caso de valores fuera de rango, sólo considerarlos si éstos demuestran continuidad en el tiempo.

2.3 SELECCIÓN DE MODELOS

La selección de un modelo útil y eficiente para los recursos que se tienen en los servidores de monitoreo consideró, en principio, mecanismos de análisis en hojas de cálculo para identificar si, con las funciones existentes, podría generarse una tendencia útil con referencia a datos previos.

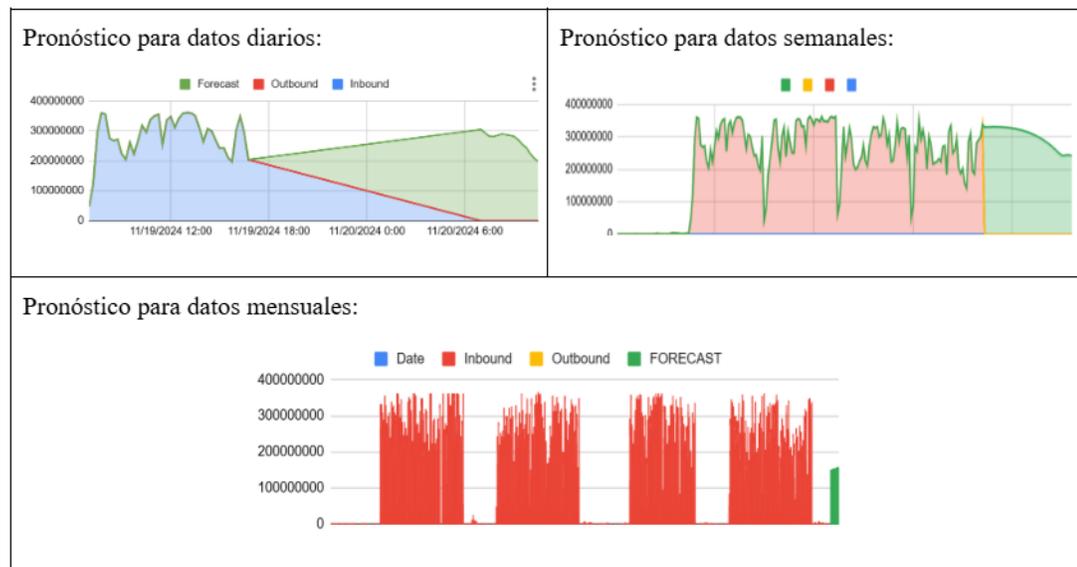
2.3.1 ANÁLISIS CON HOJAS DE CÁLCULO

Se realizó este análisis contemplando el filtrado antes mencionado y con el apoyo de las funciones existentes de hojas de cálculo, específicamente con la función “forecast”, que se apoya de la regresión lineal desplazada en el tiempo para crear los datos futuros.

Los resultados del análisis inicial reflejaron un pronóstico con tendencias consistentes a la baja, como se refleja en la Figura 2.

Figura 2

Gráficas de pronóstico empleando datos íntegros de monitoreo (sin aplicar filtros)



En la Figura 2, se muestran tres gráficas donde se observa un área verde que presenta tiempos prolongados de baja o nula utilización, éstos afectan el pronóstico de valores futuros con tendencias bajas.

A partir del primer análisis, realizado mediante la función “forecast” de las hojas de cálculo para los enlaces de internet de la UNAM, se deduce que se requería considerar la demanda centrada en los periodos específicos (picos de utilización), como son horarios laborales y periodos escolares, porque, al no hacerlo así y utilizar largos periodos de tiempo, provocaba un pronóstico siempre a la baja y poco apegado a la realidad debido a que hay periodos grandes de baja utilización como son horarios nocturnos y de fin de semana.

Esta primera prueba mostró la necesidad de definir criterios de filtrado y/o prioridad para aplicar durante el análisis de datos, tanto previo como durante el cálculo de los pronósticos. De esta manera, se busca conformar una metodología que se adecúe al comportamiento de uso de ancho de banda de los enlaces de internet que la UNAM utiliza para comunicar a sus usuarios.

2.3.1.1 PRONÓSTICO SUMANDO CRITERIOS DE FILTRADO

Se realizó un segundo pronóstico de valores aplicando filtros a partir de criterios de operación definidos por las condiciones conocidas de demanda (usando picos de utilización principalmente), generando pronósticos a través de una regresión lineal móvil.

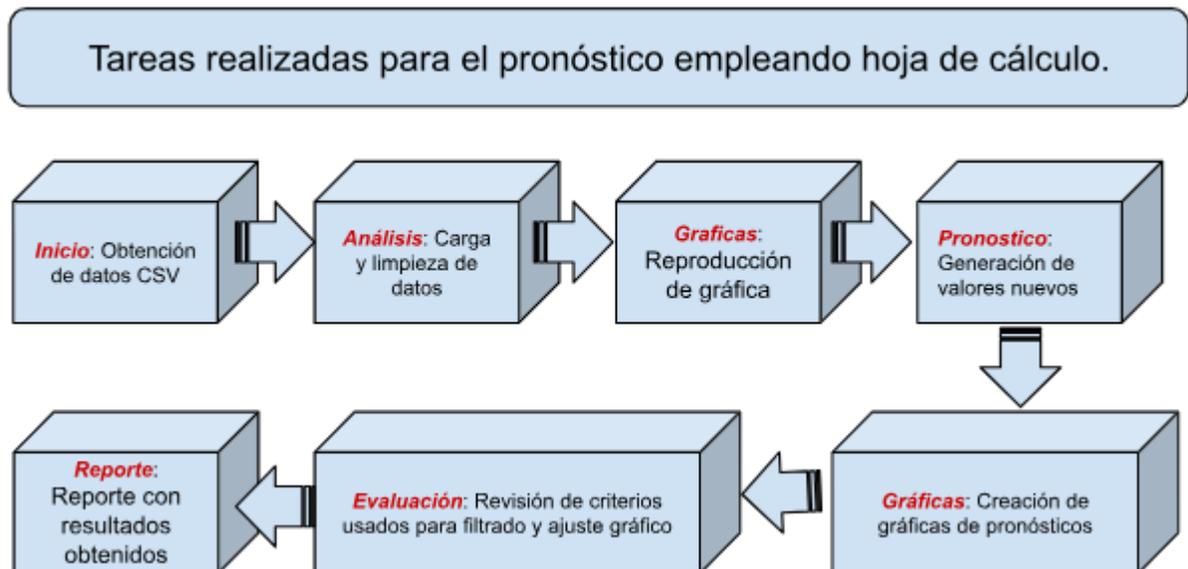
Para elaborar el primer pronóstico, se trasladaron las condiciones de demanda conocidas (variables definidas con base en el comportamiento de la demanda como las estacionalidades y horarios consumo bajo) por medio de filtros a una hoja de cálculo, donde se consideró:

- a) Valores por encima del 70% de su magnitud respecto a los máximos recuperados. Se definió con el propósito de centrar el cálculo de pronóstico en la demanda de tráfico que pudiera representar una saturación por consumo de ancho de banda.
- b) Eliminación de los datos registrados fuera de un horario laboral o actividad escolar. Esto permite eliminar posibles valores fuera de rango, propios de una consulta periódica o de ráfagas de consumo por funcionamiento anormal de las aplicaciones e incluso por transferencias de datos poco frecuentes.
- c) Eliminación de periodos de descanso (fines de semana y nocturnos) y vacaciones. Éstos se eliminaron porque no se cuenta con demanda del servicio significativa o que sature el enlace.

En la Figura 3, se muestran las tareas realizadas para este análisis.

Figura 3

Lista de tareas realizadas



Con estas condicionantes, se obtuvo:

1. Disminución de la cantidad de información para alimentarla al pronóstico.
2. Eliminación de datos fuera de rango que pudieran causar un pronóstico erróneo.
3. Utilización únicamente de información considerada útil para generar los nuevos datos.

2.3.1.2 ANÁLISIS DE RESULTADOS MEDIANTE HOJAS DE CÁLCULO

Para una inspección inicial, se descargaron los datos de día, semana, mes y año en formato CSV, el cual ofrece cada una de las gráficas mediante la plataforma de Cacti, empleando la función que tiene cada gráfica en la esquina superior derecha de las gráficas.

Las primeras evaluaciones se realizaron al usar la función FORECAST, incluida dentro de Google Sheets, que utiliza una regresión lineal simple; ésta obtuvo datos muy por debajo de los picos observados debido a que se usaron valores históricos sin filtrado para generar valores futuros que incluyen datos nulos y valores bajos derivados de largos periodos de baja utilización en días y horarios no hábiles para la universidad, tal y como se pudo observar previamente en la Figura 2.

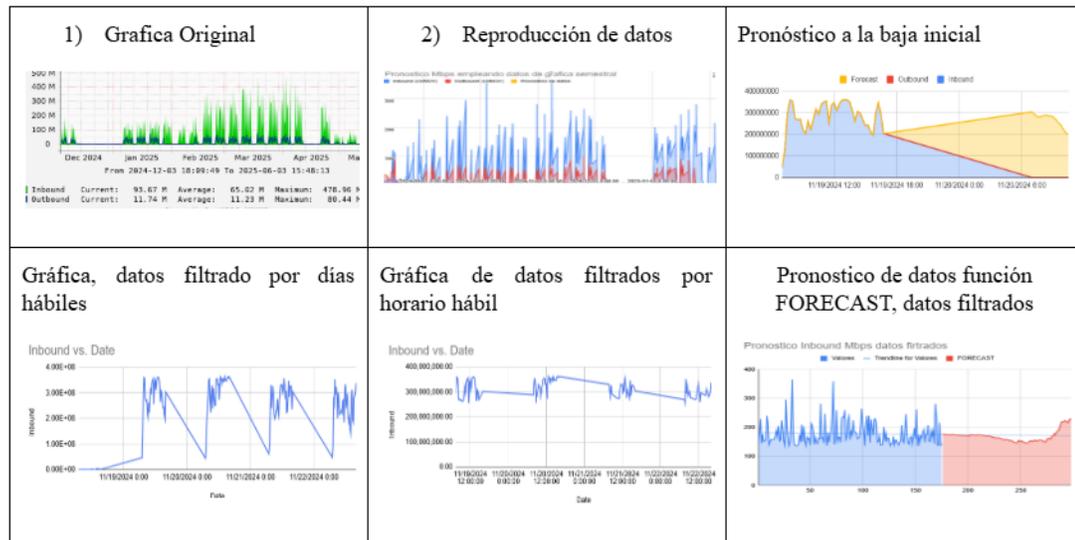
Dado que la función FORECAST arrojaba reiteradamente valores pronósticos a la baja, se efectuó un ejercicio de análisis de las variables empíricas utilizadas durante la proyección de los servicios de forma visual, empleando únicamente los datos de las gráficas y la experiencia al hacer un análisis de proyección de demanda, cuyos resultados fueron la identificación de:

- Valores identificados importantes para determinar tenencia
 - Todos aquellos valores no nulos positivos de las gráficas de consumo de ancho de banda de descarga de datos.
- Picos de datos
 - Se calculó el promedio de la demanda máxima registrada en un rango del 70% al 95% de los datos en un periodo de tiempo identificado.
- Periodos de demanda
 - Se definió la relevancia de los datos de demanda constantes por más de una hora
- Horarios de demanda
 - Se definió un horario común de incremento de demanda (10 a 16 horas) de acuerdo a la experiencia.
- Días de demanda alta de tráfico
 - Se determinó que los días inhábiles, periodos vacacionales y horarios no laborales son de baja relevancia para identificar la demanda futura.
- Ignorar eventos atípicos de demanda en caso de haber
 - Se consideran esporádicos y se excluyen del análisis los valores que no alcanzan un periodo mayor a una hora.

En la Figura 4, también se puede observar como el pronóstico, una vez aplicado con los filtros, reflejaba el comportamiento de la máxima demanda, objetivo del análisis y pronóstico.

Figura 4

Aplicación de criterios de filtrado



Las consideraciones antes mencionadas contemplaron que una alta demanda por periodos muy cortos no justifica una tendencia a incrementar la demanda de ancho de banda. Sin embargo, hay condiciones relevantes que escapan a este análisis y pueden ser consideradas y agregadas al pronóstico por medio de una variable adicional, si es considerada importante.

Debido al resultado de estos primeros análisis, se continuó con la búsqueda de algoritmos que permitieran un mejor manejo de los datos basados en series de tiempo. A raíz de esta búsqueda, se seleccionó Python como lenguaje donde se pudieran probar las bibliotecas y algoritmos de pronóstico.

2.3.2 PRONÓSTICO EMPLEANDO ALGORITMOS EN LENGUAJE PYTHON

Se seleccionó el lenguaje Python debido a su soporte, versatilidad de uso y la incorporación de bibliotecas que permitieron la ejecución de diversos modelos, siguiendo a Huang & Petukhina, (2022). Igualmente, fue clave la elección de modelos matemáticos que pudieran considerar la totalidad de elementos monitoreados, lo que derivó en una búsqueda, clasificación y evaluación de dichos modelos para su implementación por medio del lenguaje Python.

Para determinar el modelo más adecuado para las pruebas en Python, se revisaron aquellos con mayor popularidad, siguiendo lo sugerido por Fierro Torres, Castillo Pérez y Torres Saucedo (2022); esto permitió generar la lista contenida en las tablas 2 a 7.

En busca de un mejor mecanismo para el pronóstico de valores, se realizó la búsqueda de algoritmos de pronósticos disponibles en las bibliotecas de Python adecuados para un análisis con información de series de tiempo.

Se puede observar de la Tabla 1 a la 7 la evaluación de modelos para pronóstico:

Tabla 1

Código de colores de las tablas

Código de colores sobre modelos evaluados			
Modelo evaluado con posible aplicación	No es funcional para el problema	Modelo que no se recomienda aplicar	No es un modelo, sólo es apoyo para análisis

Tabla 2

Modelos tradicionales

Modelos de ARIMA	Utilización	Principales características
Simple Moving Average (SMA)	Promedios móviles de datos previos, para datos sin estacionalidad	Se utilizan pocas muestras del pasado para generar los valores futuros.
Exponential Smoothing (SES, Holt, Holt-Winters)	Promedio ponderado de datos (mayor peso a datos recientes).	Al revisar la documentación, se considera que no es aplicable porque no todos los datos tienen el mismo valor para el pronóstico.
Implementado en statsmodels, pmdarima	Se requiere establecer un parámetro de estacionalidad.	Este modelo no es utilizado porque se agrega complejidad al establecer múltiple estacionalidad.
Modelos ARIMA y Variantes (Autoregressive Integrated Moving Average)	Modelos autorregresivos que contemplan la relación de los datos pero no contemplan la estacionalidad.	No es elegido porque, de acuerdo a la documentación consultada, es más difícil considerar múltiple estacionalidad.
SARIMA (Seasonal ARIMA)	Modelos autorregresivos que contemplan la relación de los datos pero no contemplan la estacionalidad.	Este modelo no es utilizado por la complejidad de agregar variables adicionales a la estacionalidad.
SARIMAX (Seasonal ARIMA with Exogenous Variables)	Acepta variables exógenas (formadas en el exterior) y parámetros de estacionalidad.	El modelo es altamente preciso pero se vuelve complejo el análisis cuando se usan variables externas.

Tabla 3

Modelos Espacio estado

Modelos de Regresión Espacio-Estado	Utilización	Principales características
State Space Models (SSM)	Empleado con Sarimax.	No utilizado porque no contempla múltiple estacionalidad.
Kalman Filters	Estima a partir de muestras con ruido.	Los parámetros modificables incluyen un valor de ruido que no se tiene.

Tabla 4

Modelos basados en descomposición

Modelos Basados en Descomposición	Utilización	Principales características
Prophet (desarrollado por Facebook, robusto para datos con estacionalidad)	Diseñado por Facebook para series de tiempo.	Elegido para pruebas porque maneja automáticamente tendencias, estacionalidades y permite la personalización del modelo.
TBATS (modelo flexible para datos con múltiple estacionalidad)	Acepta múltiples estacionalidades. Funciona bien con datos irregulares y de alta frecuencia.	Mayor uso de recursos principalmente consumo alto de CPU.

Tabla 5

Modelos basados en Árboles de decisión

Modelos de Árboles de Decisión y Ensamblés	Utilización	Principales características
Gradient Boosting Machines (XGBoost, LightGBM, CatBoost)	Funcionan con paralelismo lo cual hace más eficiente.	Elegido para pruebas aunque más precisos en pronóstico, requieren mayor procesamiento y son menos personalizables.

Tabla 6

Modelos basados en redes neuronales

Modelos Basados en Redes Neuronales	Utilización	Principales características
MLP (Multi-Layer Perceptron)	Simulando neuronas humanas.	Alta eficiencia pero demanda alto procesamiento.
LSTM (Long Short-Term Memory)	Pronósticos empleando memoria de datos.	Elegido para pruebas, consume más CPU en la ejecución de pronósticos.
GRU (Gated Recurrent Unit)	Simulando neuronas humanas.	Elegido para pruebas, consumo alto de CPU.

Tabla 7

Modelos híbridos

Modelos Híbridos y de AutoML	Utilización	Hallazgo
AutoTS (AutoML para series de tiempo)	Sirve para seleccionar otros modelos como LSTM y XGBOOST.	Sólo sirve como librería para armar otros modelos.
Darts (interfaz unificada para múltiples modelos)	Agrupar el uso de diferentes modelos como LSTM.	Sólo sirve como librería para llamar a otros modelos.
tsfresh (extracción automática de características para ML)	Extracción de características en series de tiempo.	No diseñado para pronóstico.
tslearn (métodos de series de tiempo en aprendizaje profundo)	Clasificación.	Más útil para segmentación y clasificación.

A partir del análisis de las tablas mostradas, se seleccionaron los modelos aplicables para pruebas con el propósito de elegir el de mejor desempeño en cuanto a demanda de recursos de procesamiento y memoria. Esto tuvo la finalidad de insertarlo en un ambiente de producción como función para crear el pronóstico en tiempo real de una gráfica:

Redes neuronales:

- XGBoost
- LSTM
- MLP
- GRU

Autorregresivo aditivo:

- Prophet

2.4 EVALUACIÓN

Esta sección se centra en los modelos de pronóstico implementados en el lenguaje Python, debido a que los pronósticos basados en hojas de cálculo no requieren una validación, ya que los datos futuros sólo son el reflejo del promedio de una muestra de datos previos, generando una tendencia lineal.

2.4.1 DESCRIPCIÓN DE ANÁLISIS DE ALGORITMOS EMPLEANDO EL LENGUAJE PYTHON

En el análisis de modelos, se usaron todos los datos recolectados empleando ajustes de hiperparámetros para considerar las diferentes estacionalidades del consumo de ancho de banda (periodos en que se repite un comportamiento gráfico), lo que ayudó a tener resultados de pronóstico más fieles a los datos originales.

Los principales parámetros modificados para el análisis se describen en la siguiente lista. Se agregan los valores donde se obtuvo un pronóstico con un rango de confiabilidad aceptable (alrededor del 30%):

- Series de Fourier: 12
- Periodos de estacionalidad: día, semana, mes y cuatrimestre
- Énfasis en valores altos: se emplea transformación cuadrática
- Cantidad de períodos pronosticados: 69 días
- Ajustar variación máxima en el periodo de pronóstico: 30%

Las pruebas de código se realizaron con las aplicaciones Pycharm y Anaconda como entornos de desarrollo (IDE) del código probado.

Se emplearon para ello las bibliotecas:

- Pandas
- Matplotlib
- Prophet
- os
- sys
- Time
- Tensorflow, MinMaxscaler, Sequential, LSTM, Dense
- Numpy

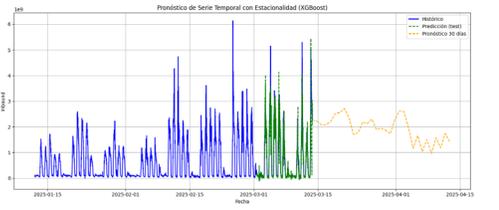
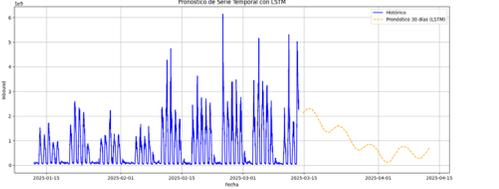
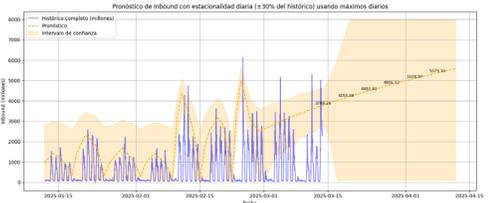
- Xgboost, XGBRegressor
- Sklearn

La comparación gráfica de las respuestas de los algoritmos ayudó a la elección, considerando la tendencia parecida al comportamiento de los datos reproducidos en las gráficas y la facilidad de modificación de los parámetros en la generación de datos pronosticados.

En la Tabla 8, se muestran ejemplos de las salidas obtenidas a partir de la aplicación de los modelos de pronóstico que ayudó a la elección del modelo final.

Tabla 8

Ejemplo de gráficas obtenidas con los modelos probado

Nombre de algoritmo	Gráfica	Observaciones
XGBOOST		<p>El modelo presentó problemas para ajustarse al comportamiento de los datos históricos ya que éstos tienen múltiple estacionalidad y comportamientos irregulares.</p>
LSTM		<p>El modelo consume muchos recursos durante el procesamiento y los datos pronosticados siempre fueron hacia la baja.</p>
PROPHET		<p>El modelo se ejecutó en un tiempo mucho menor y da la posibilidad de modificar con apoyo de variables internas la estacionalidad de los datos y da un margen de confiabilidad.</p>

2.4.2 RESULTADO DE LA EVALUACIÓN

Como resultado de las pruebas realizadas, se considera que el algoritmo Prophet ofrece la mejor respuesta a las necesidades del análisis debido a que:

- Es el algoritmo que mejor considera el comportamiento estacional de los datos debido a que ha sido empleado para analizar fenómenos económicos que tienen un comportamiento similar.
- Ofrece facilidad en el uso de variables aplicables a los datos para buscar que el pronóstico se adecúe al comportamiento histórico de la información.
- El consumo de recursos para cada pronóstico es menor que el de los demás algoritmos.

- Se cuenta con un respaldo amplio de la comunidad en el uso del algoritmo.
- Al facilitar el uso de hiperparámetros en Python, se pueden hacer ajustes de forma rápida.

2.5 VALIDACIÓN

Para la validación de los resultados, se considera la mayor cantidad de variables internas; sin embargo, éstas se aplican a fenómenos de comportamiento no lineal, lo que puede generar errores significativos debido a los ajustes en los hiperparámetros del algoritmo o a la influencia de variables exógenas que provoquen la generación de datos esperados incorrectos, como lo concluye Pooja Anand y Mayank Sharma (2024).

Sin embargo, con un correcto ajuste de los hiperparámetros y la vigilancia del comportamiento de los mecanismos de validación integrados, que permiten verificar si los pronósticos de los datos ofrecidos son aceptables, es valiosa la información entregada. La evaluación realizada para los algoritmos probados se describe a continuación.

Los resultados detallados del rendimiento de los algoritmos son mostrados en la Tabla 9, donde:

- MSE = Error Cuadrático Medio
- RMSE = Raíz del Error Cuadrático Medio
- MAE = Error Absoluto Medio
- R2= Coeficiente de determinación (qué tanto el modelo se ajusta a los datos originales)
- MAPE = Error Porcentual Absoluto Medio

Tabla 9

Comparación de rendimiento de modelos

Modelo	MSE	RMSE	MAE	R2	MAPE
XGBoost	1.149e+17	3.389e+08	1.628e+08	0.906	0.458
CNN-LSTM	3.974e+17	6.304e+08	3.431e+08	0.673	0.627
LSTM simple	4.576e+16	2.139e+08	1.064e+08	0.963	0.241
Prophet base	5.883e+17	7.670e+08	6.107e+08	0.517	4.483
Prophet ME	3.272e+18	1.809e+09	1.212e+09	0.025	0.525

La Tabla 9 refleja que el modelo mejor evaluado es LSTM simple debido a la respuesta ante errores y a su capacidad de conservar un valor a corto y mediano plazo, lo cual puede permitir que, en un conjunto de datos con múltiples estacionalidades, este comportamiento pueda tomarse en cuenta para el pronóstico a mediano plazo.

Es importante destacar que el ajuste en los parámetros operativos, durante la aplicación de un modelo derivado de la observación del desempeño, es de gran importancia, tal y como señalan Niamjoy, P. & Phumchusri, N. (2020).

Durante la evaluación del modelo XGBoost, se encontró que presenta problemas para adaptarse a cambios bajos de estacionalidad cuando ya ha considerado un comportamiento previo, lo que es congruente con los resultados de Xu, Zheng, Zhu, Wong, Wang y Lin (2024).

Un aspecto importante que apoya la elección de uno u otro modelo es la sensibilidad al comportamiento de los mismos, al sobreajuste que provoca generación de valores alejados de los esperados y para lo cual deben estar en constante vigilancia. Es el caso de LTSM, tal y como comentan Xue, Deng y Wang (2024).

En contraste con lo anterior, se identifican análisis como el de Zheng, Liu, Jiang, Tang y Xiang (2022), donde los valores de rendimiento para series de tiempo del modelo Prophet han presentado mejoría en comparación con LTSM.

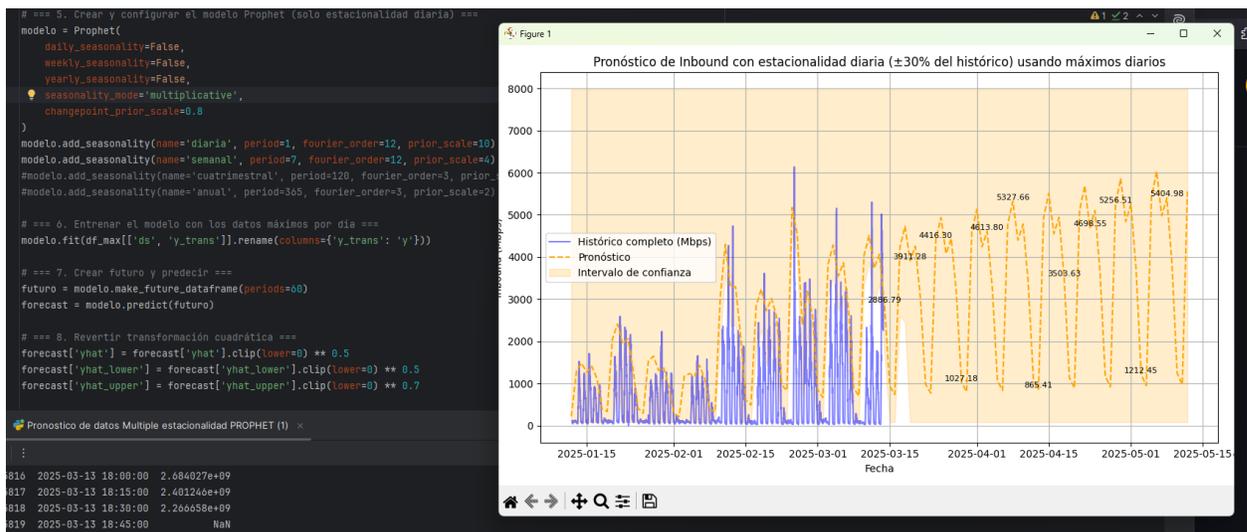
Con la información obtenida y debido a que los conjuntos de datos se adaptan de mejor manera al modelo Prophet, se determinó seleccionar este algoritmo para una implementación en la que, además, pueda realizarse una ejecución al interior del propio sistema de graficación con el apoyo del lenguaje Python.

Una vez hecha la evaluación de los modelos, se creó un programa desarrollado en Python con el fin de generar un código útil para la implementación en el sistema de gráficas Cacti.

Producto del desarrollo del *script*, se logró el pronóstico de la forma más acertada con el modelo de Prophet, mostrado en la Figura 5.

Figura 5

Resultado de ejecución del modelo de pronóstico en Python



Observaciones

- Se considera que los resultados pueden utilizarse para identificar la demanda futura, aún así, puesto que no se incluyen variables como mantenimientos y fallas en los servicios o periodos de saturación por fallas fuera de la demanda de los usuarios, no se recomienda usar únicamente el resultado del pronóstico para determinar un cambio en el ancho de banda contratado.

- El algoritmo Prophet funciona ajustando con un rango de confiabilidad alto (la posibilidad de que el pronóstico sea inexacto es alta), por lo que, para un pronóstico preciso, deben configurarse las variables que no están plasmadas en el histórico de datos y que reflejan un comportamiento de tendencia, como estacionalidad o bajas esperadas de demanda, entre otras. Cuando se aplica Prophet a rangos pequeños de datos, el pronóstico muestra una tendencia que no es precisa, esto derivado de que su funcionamiento se apoya de los cambios de tendencias (que son cíclicos). Para conjuntos de datos con estacionalidad (cambios periódicos) en un periodo corto de tiempo, como ejemplo en periodos vacacionales, si sólo se usa este periodo, generará un resultado poco confiable. Los datos históricos recuperados en un periodo anual se promedian a 30 minutos de acuerdo a la base de datos circular de RRA en Cacti, por lo que el mayor rango de datos con un intervalo de 15 minutos entre ellos es el de un semestre, así que éste es el rango que se considera para los pronósticos de ancho de banda.

3. IMPLEMENTACIÓN OPERATIVA

Con el objetivo de dar utilidad operativa al análisis (que se pueda emplear el pronóstico por todos los usuarios del sistema de gráficas Cacti), se adaptó el ícono para exportar datos del sistema Cacti, diseñado para exportarlos y para que éste también ejecute al interior del servidor el *script* con el algoritmo Prophet, generando los datos pronosticados en ese momento de esta manera, se obtiene un pronóstico de demanda futura para la gráfica que se encuentra en visualización.

Para lograr este objetivo, se realizaron las siguientes tareas:

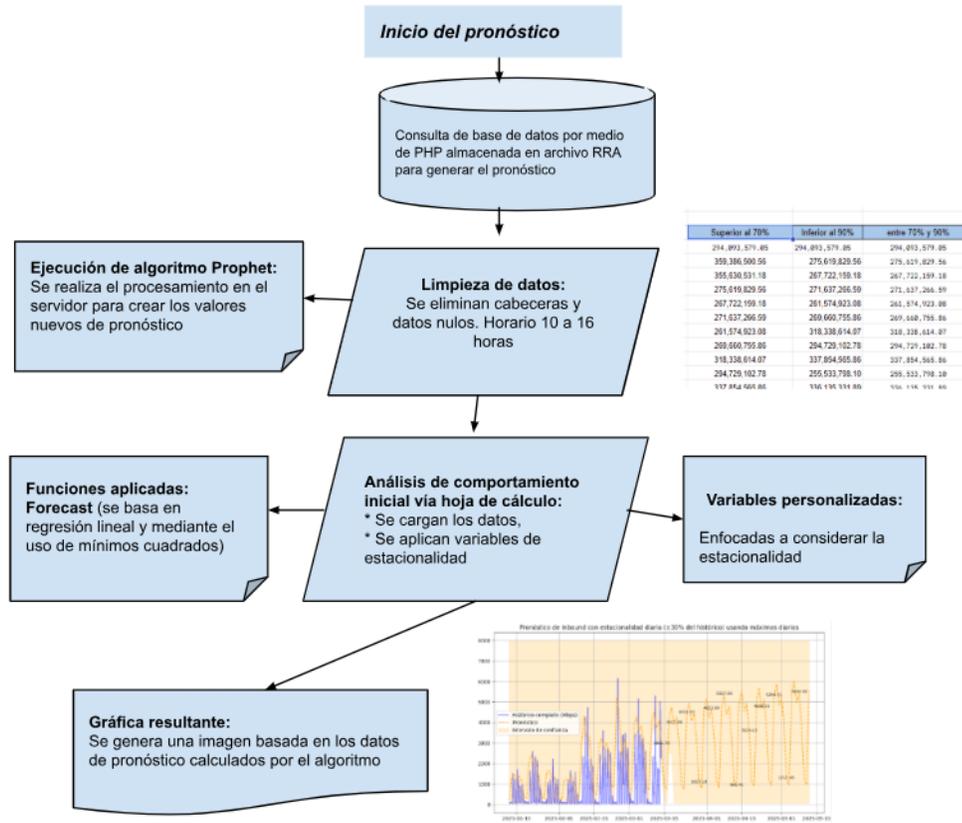
1. Implementación de bibliotecas en Python necesarias para la ejecución del pronóstico dentro del propio servidor Debian que aloja la aplicación de Cacti.
2. Implementación de *script* en PHP con el objetivo de ejecutar el código del análisis y generar la imagen interna para guardarla en el servidor.
3. Modificar la funcionalidad web de la función para exportar los datos que mostrará la gráfica de pronóstico en la sesión web del usuario.

La implementación de esta funcionalidad no está abierta a todos los usuarios debido a que, al momento de escribir el presente documento, se continúa con las pruebas sobre la aplicabilidad a todas las gráficas dentro del sistema de monitoreo.

La Figura 6 muestra el proceso aplicado dentro del sistema Cacti para la obtención del pronóstico en tiempo real de la información generada hasta el momento de ejecución de la función de pronóstico.

Figura 6

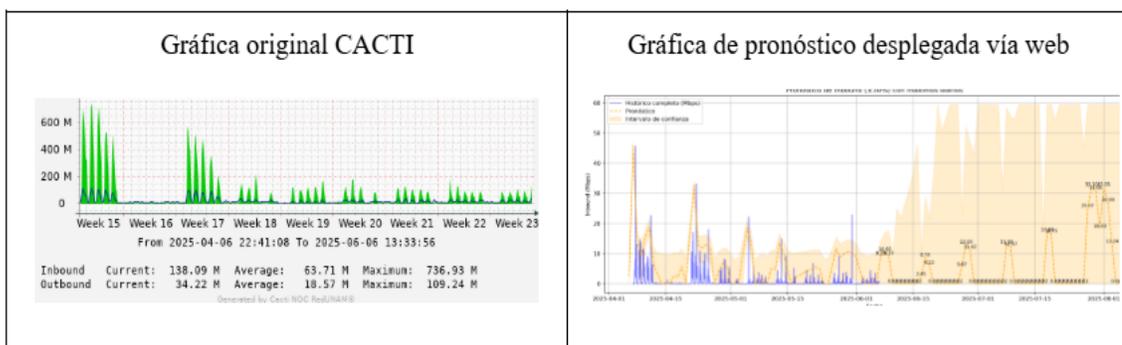
Proceso de generación de pronóstico en sistema Cacti



Para mostrar la diferencia en la Figura 6, se muestra la imagen generada por el sistema Cacti y la desplegada al ejecutar el pronóstico.

Figura 7

Pronóstico desplegado desde sitio web



Tareas pendientes:

1. El pronóstico se genera con valores predefinidos, por tanto, sería útil ofrecer al usuario un formulario para modificar los parámetros previo a su generación.
2. Mejorar el tiempo de espera asociado a la tarea de generación de pronósticos ejecutada por el servidor.

4. SIGUIENTES PASOS

Se sugiere evaluar la creación de un proyecto de generación para un *plugin* que pueda liberarse de forma oficial dentro del proyecto de Github del sistema de monitoreo Cacti, ya que, por el momento, lo que se ha hecho es modificar un *plugin* existente para insertar la función de pronóstico en el sistema.

- Evaluar la posibilidad de generar un reporte que sea realizado en segundo plano para no esperar la generación del pronóstico por parte del usuario de la herramienta Cacti.
- Ajustar los valores de los hiperparámetros a las necesidades de los usuarios del sistema Cacti.
- Evaluar la implementación de diferentes modelos para las gráficas de internet de la UNAM fuera del bachillerato para verificar su funcionalidad en todos los escenarios de demanda sobre los enlaces de la red de datos de la UNAM.
- Mejorar el tiempo de espera asociado a la tarea de generación de pronósticos ejecutada por el servidor.
- El pronóstico genera valores predefinidos, por tanto sería útil ofrecer al usuario un formulario para modificar los parámetros previo a la generación del pronóstico.

5. CONCLUSIONES

A lo largo del análisis de los datos obtenidos de las gráficas de consumo, se observó la importancia de reflejar las condiciones externas a los datos por medio del ajuste de hiperparámetros, de modo que el análisis se centre en los comportamientos de interés. En este caso, se analiza la demanda futura y configuración de parámetros para que el algoritmo dé importancia a los cambios cíclicos en día, semana y cuatrimestre. Esta personalización es clave para aplicar un modelo adecuadamente. Durante el análisis de comportamiento, también se identificó la necesidad de generar modelos diferentes para distintos tipos de demanda en los enlaces, como facultades y oficinas administrativas. La elección del algoritmo de Prophet ha permitido generar múltiples pruebas al adaptarse de mejor manera a estos cambios para realizar la mejor aproximación futura de los diferentes enlaces de internet, debido a que se adapta mejor en series temporales con patrones de datos irregulares, tal y como se muestra en otros estudios como el de Asirim, Aşirim y Salepçioğlu (2024), en donde también se concluye que la fortaleza de Prophet radica en la existencia de un calendario periódico de pruebas con patrones estacionales en los valores analizados. Además, es adecuada su elección para este estudio debido a la múltiple estacionalidad que presenta la demanda de los enlaces de internet.

Es importante reconocer las limitaciones del enfoque actual. El modelo depende de la calidad y granularidad de los datos históricos. El promedio de datos en periodos largos, una característica del sistema de almacenamiento de Cacti, puede enmascarar picos de tráfico de corta duración que son críticos. Asimismo, su capacidad predictiva se ve comprometida ante cambios estructurales abruptos en el comportamiento de la red (como la adopción masiva de una nueva tecnología de *streaming* para propósitos académicos) que no están reflejados en el histórico.

Se considera que la aplicación, por medio de series de tiempo, es indispensable en la generación de pronósticos para otros tipos de datos basados en series temporales, al poder determinar su ocurrencia y la relación con otros eventos registrados.

Este esfuerzo abre la puerta a diversas líneas de trabajo para ampliar y fortalecer significativamente sus capacidades. Se podría enriquecer el modelo incorporando variables exógenas como el número de dispositivos activos concurrentes en la red inalámbrica, métricas agregadas del tráfico de aplicaciones específicas o el número de docentes que incorporan plataformas de aprendizaje digital y recursos multimedia en sus cátedras, lo que permitiría correlacionar la demanda de ancho de banda con su origen. Adicionalmente, se podría explorar la implementación de modelos híbridos (ej., Prophet-LSTM) para mejorar la precisión a corto plazo y replicar esta metodología para pronosticar la demanda en otros dominios críticos como la ocupación de la red inalámbrica o la carga de los servidores de aplicaciones institucionales.

REFERENCIAS

- Anand, P., Sharma, M., & Saroliya, A. (2024). "A Comparative Analysis of Artificial Neural Networks in Time Series Forecasting Using Arima Vs Prophet", *International Conference on Communication, Computer Sciences and Engineering (IC3SE)*, Department of computer Science Engineering, Mody University of Science and Technology, Sikar, Rajasthan. <https://ieeexplore-ieee-org.pbidi.unam.mx:2443/document/10593482>
- Asirim, Ö. E., Aşirim, A., & Salepçioğlu, M. A. (2024). Performance of Prophet in stock-price forecasting: Comparison with ARIMA and MLP networks. *2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS)* (pp. 1–7). IEEE. <https://doi.org/10.1109/ICDS62089.2024.10756299>
- Fierro Torres, C. A., Castillo Pérez, V. H., & Torres Saucedo, C. I. (2022). Análisis comparativo de modelos tradicionales y modernos para pronóstico de la demanda: enfoques y características. *RIDE Revista Iberoamericana Para La Investigación Y El Desarrollo Educativo*, 12(24). <https://doi.org/10.23913/ride.v12i24.1203>
- Huang, C., & Petukhina, A. (2022). *Applied time series analysis and forecasting with Python*. Springer. <https://link-springer-com.pbidi.unam.mx:2443/book/10.1007/978-3-031-13584-2>
- Niamjoy, P., & Phumchusri, N. (2020). Forecasting inbound tour daily demand with multi seasonality pattern: A case study of a tour operator in Thailand. *Proceedings of the 2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)* (pp. 1044–1048). IEEE. <https://doi.org/10.1109/ICIEA49774.2020.9101918>
- Xu, Y., Zheng, S., Zhu, Q., Wong, K.-C., Wang, X., & Lin, Q. (2024). A complementary fused method using GRU and XGBoost models for long-term solar energy hourly forecasting. *Expert Systems with*

Applications, 254, 124286. <https://doi.org/10.1016/j.eswa.2024.124286>

Xue, X., Deng, S., & Wang, Y. (2024). Optimization design of predictive response based on time series forecast model LSTM in broadband high current feedback regulation. *2024 Asia-Pacific Conference on Software Engineering, Social Network Analysis and Intelligent Computing (SSAIC)* (pp. 237–241). IEEE. <https://doi.org/10.1109/SSAIC61213.2024.00051>

Zheng, Y., Liu, Y., Jiang, Z., Tang, Q., & Xiang, Y. (2022). Wind power forecasting based on Prophet model. En *2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)* (pp. 1544–1548). IEEE. <https://doi.org/10.1109/ICPSAsia55496.2022.9949918>