A close-up of a logo

Description automatically generatedA close-up of a logo

Description automatically generated

**Universidad Tecnológica de Panamá**

Maestría en Analítica de Datos

**Curso**:

Modelos Predictivos

**Avances del Proyecto Final:**

Análisis y Predicción del Consumo de Recursos en Entornos de Computación en la Nube.

**Facilitador**:

Juan Marcos Castillo, PhD

**De Gracia, Evaristo 8-788-1186**

**Año**

2025

**Introducción**

Los modelos predictivos son una herramienta poderosa que permite a usuarios y empresas realizar evaluaciones sobre sus sistemas para identificar información valiosa que les apoye en la toma de decisiones de una manera oportuna.

Durante el desarrollo de este proyecto estaré aplicando los conocimientos adquiridos en la materia de modelos predictivos de tal forma que me ayude a analizar y determinar de una manera más exacta que modelo se adecua más al estudio que quiero realizar.

El tema que he escogido para mi proyecto es la evaluación de los recursos al utilizar computación en la nube. La computación en la nube tiene la ventaja que dependiendo de la necesidad te permite escalar en recursos, pero si no se utilizan de una manera óptima y eficiente pueden llevar a generar costos innecesarios a las empresas.

**Justificación**

Actualmente me encuentro trabajando en el área de ingeniería de datos y dentro de mi último proyecto laboral nos encontramos en la migración de un sistema local a computación en la nube. Este proyecto surge de la necesidad de predecir y gestionar eficientemente el consumo de recursos en un entorno dinámico, donde la sobre provisión o subutilización de capacidades impacta directamente en la facturación operativa y la calidad del servicio. Con este análisis mi meta es poder entender y crear modelos predictivos que ayuden a la toma de decisiones y presentar un informe que sirva como base para la configuración y utilización de recursos en la nube enfocado en los costos eficientes.

**Antecedentes**

La computación en la nube ha experimentado un auge en la última década, cambiando la manera como las organizaciones manejan sus recursos tecnológicos, esto ha permitido reducir los costos operativos a las empresas ya que pueden delegar el aprovisionamiento de sus recursos computacionales a proveedores dedicados.

En este contexto, la aplicación de técnicas de modelado predictivo se ha convertido en una herramienta fundamental para las organizaciones que buscan maximizar la inversión de sus recursos en la nube. Al analizar el comportamiento histórico de sus datos como lo son consumo de CPU, memoria y/o GPU, estos modelos pueden ser utilizados de manera proactiva en vez de reactivo.

El objetivo de estos modelos y de nuestro proyecto es lograr un equilibrio entre costo-eficiencia para garantizar que los recursos asignados a un proyecto sean los correctos eliminando gastos innecesarios.

**Definición de Problema**

Con nuestro dataset actual queremos predecir los tiempos de ejecución en función del uso de recursos computacionales como CPU, memoria, GPU y como se relacionan las variables con la eficiencia de los recursos.

**Avances de análisis predictivo**

1. Selección del dataset e importarlo de csv a un dataframe para realizar el análisis descriptivo en Python utilizando VsCode.
2. El dataset contiene 8 columnas y 7064 filas sin valores nulos. Los recursos computacionales nos lo dan las columnas:
   * **cpu\_mili:** Uso de cpu en milisegundos
   * **memory\_mib:** Memoria en milisegundos
   * **gpu\_milli:** Uso de gpu en milisegundos
3. Análisis descriptivo de los datos:

**A screenshot of a computer

Description automatically generated**

1. Luego de esto, identifique mis valores numéricos y los grafique para ver como se comportaban cada uno:

**A screenshot of a graph

Description automatically generated**

1. Continuando con el análisis generé un boxplot de los mismos valores numéricos para identificar outliers en las columnas: Aquí podemos observar que hay valores dispersos en todas nuestras columnas pero son pocos puntos, por lo cual utilizaremos el dataset completo

**A screenshot of a graph

Description automatically generated**