



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**DISEÑO DE INVESTIGACIÓN PARA LA CONSTRUCCIÓN DE UN MODELO MATEMÁTICO
PARA ESTIMAR EL RENDIMIENTO DE UN SERVIDOR MOODLE DURANTE EL PERIODO
DE EXÁMENES EN LÍNEA DE LA FACULTAD DE ARQUITECTURA
DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**

José Victor Tobias Romero

Asesorado por el M.A. Ing. Rogelio Augusto Arroyo Castillo

Guatemala, febrero de 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**DISEÑO DE INVESTIGACIÓN PARA LA CONSTRUCCIÓN DE UN MODELO MATEMÁTICO
PARA ESTIMAR EL RENDIMIENTO DE UN SERVIDOR MOODLE DURANTE EL PERIODO
DE EXÁMENES EN LÍNEA DE LA FACULTAD DE ARQUITECTURA
DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

JOSÉ VICTOR TOBIAS ROMERO

ASESORADO POR EL M.A. ING. ROGELIO AUGUSTO ARROYO CASTILLO

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, FEBRERO DE 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Bladimir Cruz Lorente
VOCAL V	Br. Fernando José Paz Gonzales
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANA	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Ing. Marlon Francisco Orellana López
EXAMINADOR	Ing. Neftali de Jesús Calderón Méndez
EXAMINADOR	Ing. José Manuel Ruíz Juárez
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**DISEÑO DE INVESTIGACIÓN PARA LA CONSTRUCCIÓN DE UN MODELO MATEMÁTICO
PARA ESTIMAR EL RENDIMIENTO DE UN SERVIDOR MOODLE DURANTE EL PERIODO
DE EXÁMENES EN LÍNEA DE LA FACULTAD DE ARQUITECTURA
DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Posgrado, con fecha 10 de noviembre de 2022.



José Victor Tobias Romero



EEPFI-PP-1795-2022

Guatemala, 10 de noviembre de 2022

Directo:
Carlos Gustavo Alonzo
Escuela De Ingeniería En Sistemas
Presente.

Estimado Ing. Alonzo

Reciba un cordial saludo de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería.

El propósito de la presente es para informarle que se ha revisado y aprobado el Diseño de Investigación titulado: **CONSTRUCCIÓN DE UN MODELO MATEMÁTICO PARA ESTIMAR EL RENDIMIENTO DE UN SERVIDOR MOODLE DURANTE EL PERIODO DE EXÁMENES EN LÍNEA DE LA FACULTAD DE ARQUITECTURA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**, el cual se enmarca en la línea de investigación: **Todas las áreas - Pronósticos**, presentado por el estudiante **José Victor Tobias Romero** carné número **200714364**, quien optó por la modalidad del "PROCESO DE GRADUACIÓN DE LOS ESTUDIANTES DE LA FACULTAD DE INGENIERÍA OPCIÓN ESTUDIOS DE POSTGRADO". Previo a culminar sus estudios en la Maestría en ARTES en Estadística Aplicada.

Y habiendo cumplido y aprobado con los requisitos establecidos en el normativo de este Proceso de Graduación en el Punto 6.2, aprobado por la Junta Directiva de la Facultad de Ingeniería en el Punto Décimo, Inciso 10.2 del Acta 28-2011 de fecha 19 de septiembre de 2011, firmo y sello la presente para el trámite correspondiente de graduación de Pregrado.

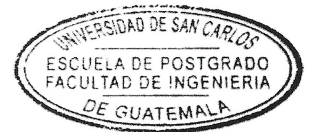
Atentamente,

"Id y Enseñad a Todos"

Mtro. Rogelio Augusto Arroyo Castillo
Asesor(a)

Rogelio Augusto Arroyo Castillo
Ingeniero en Ciencias y Sistemas
Colegiado No. 19,935

Mtro. Edwin Adalberto Bracamonte Orozco
Coordinador(a) de Maestría



Mtro. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería





EEP-EICS-1449-2022

El Director de la Escuela De Ingenieria En Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador y Director de la Escuela de Estudios de Postgrado, del Diseño de Investigación en la modalidad Estudios de Pregrado y Postgrado titulado: **CONSTRUCCIÓN DE UN MODELO MATEMÁTICO PARA ESTIMAR EL RENDIMIENTO DE UN SERVIDOR MOODLE DURANTE EL PERIODO DE EXÁMENES EN LÍNEA DE LA FACULTAD DE ARQUITECTURA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**, presentado por el estudiante universitario **José Victor Tobias Romero**, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingeniería en esta modalidad.

ID Y ENSEÑAD A TODOS

Ing. Carlos Gustavo Alonzo
Director
Escuela De Ingenieria En Sistemas

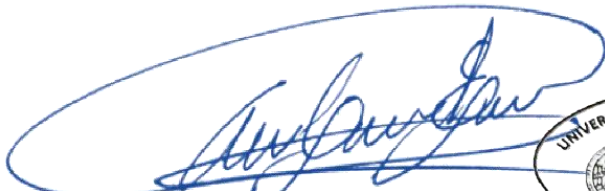
Guatemala, noviembre de 2022


Decanato
Facultad de Ingeniería
24189101- 24189102
secretariadecanato@ingenieria.usac.edu.gt

LNG.DECANATO.OI.143.2023

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al Trabajo de Graduación titulado: **DISEÑO DE INVESTIGACIÓN PARA LA CONSTRUCCIÓN DE UN MODELO MATEMÁTICO PARA ESTIMAR EL RENDIMIENTO DE UN SERVIDOR MOODLE DURANTE EL PERIODO DE EXÁMENES EN LÍNEA DE LA FACULTAD DE ARQUITECTURA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**, presentado por: **José Victor Tobias Romero**, después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:


Inga. Aurelia Anabela Cordova Estrada
Decana



Guatemala, febrero de 2023

AACE/gaoc

ACTO QUE DEDICO A:

Mi padre

Por ser la fuente de inspiración para seguir superándome.

Mis hermanos

Por todo el apoyo que me han brindado a lo largo de mi vida.

Mis amigos

Por siempre apoyarme de forma incondicional y presionarme para continuar mis estudios.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por ser mi casa de estudios y brindarme un lugar de trabajo.
Escuela de Estudios de Postgrado	Por permitirme continuar con mis estudios y seguirme desarrollando como profesional.
Facultad de Arquitectura	Por su apoyo para el desarrollo de esta investigación.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES	V
LISTA DE SÍMBOLOS	VII
GLOSARIO	IX
RESUMEN	XI
1. INTRODUCCIÓN	1
2. ANTECEDENTES	5
3. PLANTEAMIENTO DEL PROBLEMA	11
3.1. Contexto general	11
3.2. Descripción del problema	11
3.3. Formulación del problema	12
3.4. Delimitación del problema	13
4. JUSTIFICACIÓN	15
5. OBJETIVOS	17
5.1. General.....	17
5.2. Específicos	17
6. NECESIDADES A CUBRIR Y ESQUEMA DE LA SOLUCIÓN	19
7. MARCO TEÓRICO.....	21
7.1. Estadística.....	21

7.1.1.	Estadística descriptiva.....	21
	7.1.1.1. Utilidad	21
	7.1.1.2. Medidas.....	22
7.1.2.	Estadística inferencial	24
	7.1.2.1. Medidas.....	25
	7.1.2.2. Intervalos de confianza	26
7.1.3.	Regresión.....	27
	7.1.3.1. Tipos de variables	27
	7.1.3.2. Regresión lineal.....	28
	7.1.3.3. Regresión no lineal.....	29
	7.1.3.4. Selección del modelo	29
7.1.4.	Diseño de experimentos.....	30
	7.1.4.1. Requisitos de un diseño experimental.....	30
7.2.	Plataforma Moodle	31
7.2.1.	Servidor Web.....	31
	7.2.1.1. Rendimiento	32
	7.2.1.2. Modelo de rendimiento	33
	7.2.1.3. Recursos de un servidor web	33
7.2.2.	Plataforma Moodle	35
	7.2.2.1. Tipo de licencia	36
	7.2.2.2. Escalabilidad	36
	7.2.2.3. Rendimiento	36
	7.2.2.4. Mecanismos para determinar el rendimiento.....	37
7.2.3.	Pruebas de software	38
	7.2.3.1. Tipos de pruebas.....	38
	7.2.3.2. Pruebas web automatizadas	40
	7.2.3.3. Métricas de rendimiento	41

8.	PROPUESTA DE ÍNDICE DE CONTENIDOS	43
9.	METODOLOGÍA.....	45
9.1.	Características del estudio	45
9.2.	Unidades de análisis	46
9.3.	Variables.....	46
9.4.	Fases del estudio	47
10.	TÉCNICAS DE ANÁLISIS DE INFORMACIÓN.....	53
10.1.	Estadística descriptiva:.....	53
10.2.	Estimación de medias.....	53
10.3.	Correlación	54
10.4.	Pruebas de bondad de ajuste.....	54
10.5.	Regresión lineal y multilínea	54
10.6.	Regresión no lineal.....	55
11.	CRONOGRAMA.....	57
12.	FACTIBILIDAD DEL ESTUDIO	59
12.1.	Recurso humano	59
12.2.	Recursos tecnológicos.....	59
12.3.	Acceso a información y permisos	59
12.4.	Equipo e infraestructura.....	60
12.5.	Recursos financieros	60
13.	REFERENCIAS.....	61
14.	APÉNDICE	67

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Scatter Plot – Emisiones de CO ₂ vs tamaño del motor	23
2.	Representación básica de la conexión cliente/servidor	32
3.	Esquema de solución	51
4.	Cronograma de actividades.....	57

TABLAS

I.	Operativización de variables	47
II.	Detalle de los recursos financieros	60

LISTA DE SÍMBOLOS

Símbolo	Significado
%	Porcentaje
Q	Quetzal

GLOSARIO

CPU	Abreviatura de <i>Central Processing Unit</i> , que es el componente de la computadora en la que se encuentran los elementos que sirven para el procesamiento de datos.
HTTP	Es la abreviatura de <i>Hypertext Transfer Protocol</i> , y es el conjunto procedimientos y estándares que permiten la transferencia de datos en el Internet, generalmente documentos HTML.
Memoria RAM	Componente de la computadora donde se almacenan los datos o instrucciones que posteriormente son utilizados por el CPU. Es la memoria principal de una computadora y es de alta velocidad, a diferencia de los dispositivos de almacenamiento convencionales.
Servidor Web	Un servidor web es un software y un hardware que utiliza el protocolo HTTP y otros protocolos para responder a las peticiones de los clientes.
Usuario concurrente	Son aquellos usuarios para los que el servidor está realizando una tarea activa.

RESUMEN

Moodle es una plataforma diseñada para brindar un ambiente robusto, seguro e integrado a estudiantes, maestros y administradores. Moodle es una plataforma gratuita.

Moodle es utilizado como plataforma principal para el desarrollo de las asignaturas en la Facultad de Arquitectura.

El crecimiento de usuarios que utilizan la plataforma Moodle impacta de forma directa al consumo de recursos del servidor en donde se aloja la plataforma. Para muchos administradores es difícil determinar un estimado de recursos necesarios para poder brindar un servicio adecuado a los usuarios porque no existe un modelo matemático que permita realizar una proyección de uso de recursos.

El objetivo de la presente investigación es generar un modelo matemático para poder determinar el estimado de recursos necesarios para que una plataforma Moodle pueda brindar un servicio adecuado. Los datos necesarios para la construcción del modelo matemático serán generados a través de una serie de experimentos en donde la carga de usuarios y recursos del sistema serán controlados.

Para determinar el mejor modelo matemático se ajustarán un conjunto de modelos de regresión y luego serán evaluados para determinar el modelo con mejor ajuste y mejores criterios de aceptación.

1. INTRODUCCIÓN

El presente estudio de investigación es una sistematización de las cargas que experimenta un servidor Moodle, durante los periodos de exámenes en línea y consiste en identificar, ordenar y clasificar los elementos de un servidor Moodle y el impacto de los usuarios concurrentes, sobre los recursos del Servidor que lo alberga.

Al momento de realizar la investigación, no existe un modelo matemático que permita predecir la carga sobre las plataformas o el rendimiento total de usuarios, concurrentes que podría soportar un servidor Moodle. Exceder la capacidad de los recursos, con los que cuenta la plataforma produce problemas, de rendimiento o la interrupción del servicio.

Para las instituciones que utilizan Moodle, es importante estimar el uso de los recursos del servidor Moodle, porque deben de poder garantizar la continuidad de las evaluaciones de los estudiantes, sin que el servicio se vea interrumpido por la saturación de los servidores. Cuando se implementa un modelo para estimar el rendimiento de las plataformas, se puede brindar un mejor servicio y también se puede optimizar recursos o la inversión necesaria, para garantizar un servicio continuo y adecuado.

El enfoque de la investigación es cuantitativo y con un diseño experimental. El estudio también tiene un alcance descriptivo-correlacional, en donde se estudia la importancia de la relación entre las variables independientes y dependientes.

Como resultado de la investigación, se espera obtener un modelo matemático que permita proyectar la cantidad de usuarios concurrentes, que puede soportar un servidor Moodle. El modelo será útil para el personal administrativo y técnico de la institución, ya que con él pueden proyectar la cantidad de usuarios que pueden tener en la plataforma de forma concurrente, sin tener inconvenientes de rendimiento o falla en los sistemas.

Para la presente investigación se desarrollan seis fases. En la primera fase se hará una revisión documental, con la que se fundamentan los temas en la investigación. En la segunda fase, extraerá la información de la base de datos histórica de Moodle y se hará un análisis descriptivo. La tercera fase consiste en recolectar los datos, bajo los ambientes controlados de la plataforma Moodle y el análisis descriptivo de los datos. En la cuarta fase se construirán los modelos matemáticos, para determinar el rendimiento de los servidores Moodle, con base en la información recolectada en la fase anterior. En la quinta fase se evaluará el desempeño de cada uno de los modelos, para determinar el modelo que mejor se ajuste al problema. Por último, en la sexta fase se redactará el informe final.

Se cuenta con todos los recursos necesarios para poder llevar a cabo el trabajo de investigación. El informe final está constituido por los siguientes capítulos:

En el primer capítulo, se encuentra el marco referencial de la investigación.

En el segundo capítulo, se encuentra el marco teórico, el cual está constituido de dos secciones. La primera sección, detalla todos los fundamentos teóricos necesarios para llevar a cabo la investigación. La segunda sección detalla todos los fundamentos que sustentan el estudio, es decir, la plataforma Moodle y los conceptos relacionados con la plataforma.

En el tercer capítulo se hará la presentación de resultados y en el cuarto capítulo se hará la discusión de resultados, y con base a la discusión de resultados, las conclusiones y recomendaciones pertinentes.

2. ANTECEDENTES

En el marco de referencia, se han encontrado investigaciones relacionadas a la construcción de modelos de regresión, el diseño de experimentos de software, la validación de modelos de regresión y marcos de trabajo o metodologías para abordar la problemática. En las siguientes líneas, se puede apreciar cómo cada autor aporta los conocimientos necesarios, para la realización del presente trabajo y la validación de los resultados obtenidos.

Según Sun y Chen (2016), quienes hicieron un análisis de la educación online y las prácticas efectivas, en donde estudian la naturaleza transformativa de la educación y las tendencias a futuro en la modalidad en línea. En el año 2013 se estimaba que el 32 % de la población, haciendo uso de más de seis años de información institucional para la estimación, se había asignado por lo menos a un curso totalmente en línea. Hoy en día la educación virtual ha crecido de forma considerable en países desarrollados e incluso en países como Guatemala.

Este trabajo brinda información importante sobre el crecimiento de la educación en línea y de la importancia de la utilización de dichas herramientas. El crecimiento en estas herramientas implica un crecimiento en la demanda de los recursos de las plataformas y demuestra la importancia de la necesidad de poder estimar dichas demandas.

De acuerdo con Lloyd y Connie (2002), publicaron un artículo en el cual se detalla una serie de pasos para solucionar problemas de rendimiento en software, brindando un marco de trabajo para análisis varios contextos,

relacionados con el proceso de mejorar el rendimiento de las plataformas. Así mismo, describen las siguientes técnicas para estimar los requerimientos de recursos del software: *benchmarks* de prototipos para encontrar modelos de rendimiento y modelos para estimar el objetivo de rendimiento deseado.

Contar con técnicas y configuraciones básicas para el ambiente de la plataforma Moodle, es fundamental para poder aislar las mediciones de rendimiento y los problemas de las configuraciones incorrectas o inadecuadas. El uso de prototipos también ayuda a diseñar de mejor manera las pruebas de rendimiento.

Tal como lo indican Bezemer y Zaidman (2011) publicaron un artículo en donde describen cómo lograron predecir la falla por sobrecarga en un servidor, a través de un modelo de clasificación, para poder detectar patrones de falla o secuencia de estados que pueden producir una falla.

La publicación de Bezemer y Zaidman describe cómo crearon ambientes en estado normal del servidor y ambientes en estado de sobrecarga, para poder entrenar el modelo de clasificación, esto es útil porque brinda un marco de trabajo para la creación de ambientes y los estados que se deben buscar para cuantificar el rendimiento del servidor.

Palmer y O'connell (2009) publicaron un documento en el que hacen una abstracción del proceso del análisis de regresión lineal, el cual describen como “una técnica estadística para estimar la relación entre una variable dependiente (criterio) y una o más variables independientes (predictor)”. En la selección de los predictores, describen que el objetivo de la discriminación del modelo es minimizar el número de predictores, que logren contener la varianza máxima, del criterio o variable a predecir. Para encontrar las variables independientes que

mejor se adaptan, usan el coeficiente de determinación (R^2) y un análisis de correlación múltiple.

La publicación describe criterios y pruebas que pueden ser útiles al momento de seleccionar las variables independientes para la construcción de los modelos matemáticos para predicciones tales como: la selección del criterio o variable de salida, selección de los predictores, la evaluación de la precisión de la predicción y la evaluación de la estabilidad de los modelos.

Praire (1996) publicó un artículo sobre la evaluación de la efectividad de la predicción de un modelo de regresión en donde se plantea la utilización de clases. Las clases son construidas con el rango total y la desviación estándar (se utiliza dos veces la desviación estándar para estimar un intervalo de confianza de aproximadamente 95 %). Las variaciones en la efectividad de la evaluación de la potencia de las regresiones tienen una mayor influencia cuando el coeficiente de determinación es superior al 0.6.

El artículo proporciona información importante sobre la forma en la que se puede evaluar el rendimiento de un modelo. Para este trabajo es de particular importancia porque se estarán produciendo diferentes modelos y evaluando cuál tiene mejor rendimiento o efectividad.

Senthilnathan (2019) publicó un artículo en donde describe la importancia del análisis de correlación y estimación del grado de relación entre dos variables, especialmente en el análisis de relación entre las variables dependientes e independientes. La publicación hace énfasis en evitar la multicolinealidad de las variables, la cual se puede identificar con medidas tales como el factor de la inflación de la varianza (VIF). El autor proporciona una serie de intervalos como referencia para identificar el grado de multicolinealidad.

El estudio de la multicolinealidad de las variables es una etapa importante para y en especial para los modelos lineales y multicolinealidad. El artículo proporciona un marco de trabajo para identificar este tipo de relaciones y las mediciones que se deben de tomar en cuenta al momento de generar los modelos para su evaluación.

En su artículo Rajalaxmi, Natesan, Krishnamoorthy y Ponni (2019) hablan sobre la construcción de un modelo de regresión, para la estimación del desempeño académico de estudiantes en donde se describe que, para la generación y validación del modelo, el conjunto de datos fue dividido en dos grupos, un grupo para la estimación del modelo y el otro grupo para la validación del grupo.

La separación del conjunto de datos para estimar y validar es importante para reducir el sobreajuste y mejorar la tasa de ajuste, para los datos fuera del conjunto de datos de estimación del modelo. La estrategia de la separación de los conjuntos de datos será aplicada en el presente trabajo para reducir el sobreajuste del modelo.

Khan, Algarni, Kumar, Choudhary y Srivastava (2002) publicaron un artículo sobre la evaluación del rendimiento de modelos de regresión, para la predicción de infecciones de COVID, en donde realizaron un conjunto de regresiones tales como regresiones lineales, multilineales y polinomiales. Luego de realizar las regresiones, y sus estimaciones de ajuste y error, los autores realizaron una serie de gráficas, para estimar de forma visual el modelo con mejor ajuste para las predicciones de infecciones y recuperaciones. En el artículo se describe la importancia de utilizar múltiples modelos de regresión, para lograr estimar el modelo correcto y evitar utilizar un modelo tradicional el cual puede producir un ajuste deficiente.

La creación de múltiples modelos para la evaluación del rendimiento y ajuste es importante para escoger el modelo adecuado. Si se utiliza un único modelo, puede ser que los datos se ajusten dentro de los márgenes esperados, pero puede no ser el adecuado. El enfoque de crear múltiples modelos y evaluar su rendimiento será utilizado en el presente trabajo.

De acuerdo con Corrales y Ledezma (2018) presentan un proceso para la limpieza y el aseguramiento de la calidad de los datos para regresiones. Algunos de los problemas que los autores describen son los valores faltantes, valores atípicos y redundancia. Finalmente, el artículo brinda una serie de mecanismos para tratar cada uno de los problemas, en los datos tales como la eliminación de duplicados, el relleno de información con la media de la muestra y la eliminación de datos faltantes cuando no representan una gran proporción de la muestra.

El correcto tratamiento de los datos es fundamental para que los modelos no presenten un bajo rendimiento o un comportamiento sesgado. La limpieza y preparación de los datos es una etapa fundamental y tan importante como la validación del rendimiento de los modelos.

Uyanik y Güler (2013) publican un artículo en el que abordan el análisis de las regresiones multilíneas, y las relaciones entre causa y efecto de dos o más variables independientes. En el artículo los autores describen una serie de pasos para el tratamiento de los datos, comenzando con los datos faltantes, normalidad de los datos, valores extremos y la relación multilínea. Adicionalmente los autores utilizaron ANOVA, coeficiente de inflación de la varianza (VIF) y índice de condición (CI), para estimar la importancia de las variables y verificar cuáles pueden ser eliminadas del modelo para volverlo más eficiente con un menor número de variables regresoras.

Analizar el comportamiento y distribución de los datos a través de mediciones o herramientas como un ANOVA, las cuales brindan un panorama más amplio de la información con la que se cuenta y la forma en la que se deben de tratarlas y utilizarlas. El estudio brinda un marco de trabajo que se puede seguir para el tratamiento adecuado de los datos.

Durante esta revisión bibliográfica se puede concluir que tanto el tratamiento de las variables, la limpieza de los datos y el uso de técnicas para estimar la importancia de las variables, son de mucho valor para poder construir modelos matemáticos adecuados. También se puede concluir que la construcción de múltiples modelos es un enfoque conveniente para encontrar el modelo que tenga mejor rendimiento.

3. PLANTEAMIENTO DEL PROBLEMA

3.1. Contexto general

La plataforma Moodle, ha tomado mucha importancia en la enseñanza a través de medios virtuales y ha tenido un incremento, en la cantidad de usuarios activos y concurrentes.

Dicho aumento de usuarios ha provocado que muchas plataformas colapsen, durante el desarrollo de actividades, como lo son los exámenes o evaluaciones. Para las instituciones que brindan servicios con la plataforma Moodle, es de suma importancia poder estimar la cantidad de recursos necesarios, para poder brindar un servicio de calidad y sin interrupciones.

Existen herramientas y configuraciones que pueden ayudar a mejorar el rendimiento, pero ninguna de ellas permite identificar de forma directa, cuales deberían de ser los recursos requeridos, para brindar un servicio adecuado o determinar la capacidad de los servidores, con los que cuenta las instituciones educativas como la Universidad de San Carlos, en la Facultad de Arquitectura.

3.2. Descripción del problema

La Facultad de Arquitectura no cuenta con un modelo matemático, que le permita predecir la carga sobre las plataformas o el rendimiento total de usuarios concurrentes, que podría soportar un servidor Moodle.

Al no contar con un método para estimar el rendimiento de los servidores Moodle, muchos servicios se han visto interrumpidos o brindado un servicio ineficiente, por la saturación de los servidores y la cantidad de recursos asignados a los servidores.

3.3. Formulación del problema

A continuación, se presenta la pregunta central y las preguntas auxiliares que servirán para orientar el desarrollo de la investigación. Las respuestas a las preguntas auxiliares nos orientan a responder la pregunta central.

- Pregunta central

¿Cuál es el modelo matemático que permite estimar el rendimiento de un servidor Moodle durante exámenes en línea?

- Preguntas auxiliares

- ¿Cuál es el tiempo promedio que le toma a un estudiante realizar una evaluación en línea a nivel universitario?
- ¿Cuál es la carga promedio en un servidor Moodle por cada estudiante que realiza una evaluación en línea?
- ¿Cuál es la relación de los recursos de un servidor Moodle sobre el rendimiento de la cantidad de usuarios activos simultáneos durante una evaluación?

3.4. Delimitación del problema

- Espacial: Universidad de San Carlos de Guatemala.
- Temporal: evaluaciones de estudiantes del año 2020 al 2022.
- Poblacional: Facultad de Arquitectura de la Universidad de San Carlos de Guatemala.

4. JUSTIFICACIÓN

El problema se centra en el área de educación, porque Moodle es una plataforma de educación virtual, y el enfoque principal es el comportamiento de los servidores durante las evaluaciones en línea. El diseño es de tipo experimental y utiliza la línea de investigación de pronósticos para el desarrollo de la investigación.

La construcción de un modelo matemático, para determinar el rendimiento de un servidor Moodle, es de suma importancia para poder administrar de forma efectiva, los recursos y poder cumplir con la demanda de usuarios en la plataforma Moodle.

El interés personal del investigador está relacionado con el trabajo que se realiza día a día por el personal administrativo y técnico de la Facultad de Arquitectura, para poder brindar un servicio adecuado a sus estudiantes en los últimos años, en donde se han desarrollado de forma virtual todos los cursos.

Con el modelo matemático que se desarrollará, el personal administrativo y técnico podrán tener un panorama claro de la capacidad de sus servidores, en donde se encuentra la plataforma Moodle y también podrán realizar proyecciones a futuro.

Los beneficiarios de la investigación son el personal administrativo, personal técnico y los estudiantes en general de la Facultad de Arquitectura.

La relevancia de la construcción de un modelo matemático, para cualquier ambiente social o laboral es bastante grande. Con ello se logra obtener una herramienta que permite facilitar la gestión de los eventos y también permite determinar de forma sencilla las oportunidades de mejora. A través de la estadística se pueden generar modelos de pronósticos, los cuales son de vital importancia en el desarrollo de los procesos.

5. OBJETIVOS

5.1. General

Construir de un modelo matemático para estimar el rendimiento de un servidor Moodle, durante el periodo de exámenes en línea, de la Facultad de Arquitectura de la Universidad de San Carlos de Guatemala.

5.2. Específicos

1. Estimar el tiempo promedio que le toma a un estudiante, a nivel universitario, realizar una prueba en línea, a través de un intervalo de confianza de la media en la plataforma Moodle de la Facultad de Arquitectura para parametrizar la herramienta de simulación de estudiantes concurrentes.
2. Diseñar una herramienta desarrollada en el lenguaje de programación Python de simulación de usuarios concurrentes, en un ambiente controlado de Moodle para sobrecargarlo y estimar la media del consumo de recursos del servidor (memoria RAM, uso de disco duro y uso de CPU).
3. Cuantificar la relación que existe entre los recursos del servidor Moodle (RAM, CPU, velocidad del disco duro, etc.) y el rendimiento de la plataforma a través de pruebas de correlación y regresión para poder descartar las variables que no aportan demasiada información al modelo matemático.

6. NECESIDADES POR CUBRIR Y ESQUEMA DE LA SOLUCIÓN

El personal técnico y administrativo de la Facultad de Arquitectura, no cuenta con una herramienta que le permita determinar la cantidad de usuarios, que puede soportar la plataforma Moodle que utilizan para impartir los cursos de Pregrado y Postgrado.

En algunas oportunidades la capacidad de los servidores Moodle se ha visto sobrepasada y se han producido problemas, como la interrupción del servicio y el deterioro del rendimiento.

Con el modelo matemático que se va a construir con la investigación, el personal administrativo puede realizar proyecciones de la cantidad de estudiantes, que van a utilizar el servicio de forma simultánea y el personal técnico puede determinar la cantidad de recursos, que requiere para que el servicio no se vea interrumpido o presente problemas de rendimiento.

Para la solución del problema se ensayará el siguiente esquema de solución:

Primero se realizará la revisión de literatura, obteniendo el conjunto de conocimientos teóricos y técnicos necesarios para llevar a cabo la investigación. Posteriormente se estará realizando la recolección de datos de la base de datos histórica de las evaluaciones en Moodle, para estimar la cantidad de tiempo que utiliza un estudiante para realizar una evaluación en línea.

Posteriormente se estará construyendo la aplicación que servirá para simular el comportamiento de un estudiante en la plataforma Moodle (ingresando a la aplicación con usuario y contraseña, navegando por los menús, realizando la evaluación en línea de prueba, etc.).

Con la aplicación de simulación de usuarios, se estarán realizando una serie de experimentos con condiciones controladas en donde se recolectará los datos más relevantes del servidor, durante un periodo de tiempo y condiciones específicas.

Luego de recolectar los datos, se construirán un conjunto de modelos matemáticos, para que luego sea evaluada la precisión de las predicciones y así elegir el mejor modelo. Por último, se redactará el informe final.

7. MARCO TEÓRICO

A continuación, se presenta una serie de conceptos que sirven como guía para el desarrollo de la investigación.

7.1. Estadística

La estadística es una disciplina de la ciencia que se ocupa de la recolección, análisis, presentación e interpretación de datos. La estadística permite adquirir conocimientos a través de los datos y es crucial para el conocimiento científico.

7.1.1. Estadística descriptiva

La estadística descriptiva describe una muestra o grupos de datos. La estadística descriptiva proporciona información de los datos, pero no intenta generar una estimación o proyección de los datos, es decir, se limita únicamente a la descripción de la información que se encuentra analizando.

7.1.1.1. Utilidad

La estadística descriptiva proporciona información general de los datos y permite realizar resúmenes de la información y representar gráficamente los datos, para tener una idea de la forma en la que se distribuyen los datos, patrones, datos atípicos, etc.

7.1.1.2. Medidas

Existe un conjunto de medidas que se pueden obtener con estadística descriptiva, pero algunas de las principales y útiles para este trabajo de investigación se encuentran a continuación:

- Tendencia central

Las medidas de tendencia central permiten conocer hacia dónde se dirigen la mayoría de los valores. Regularmente se utiliza la media o la mediana para conocer cuál es el centro del conjunto de datos.

La ecuación para encontrar la media es la siguiente:

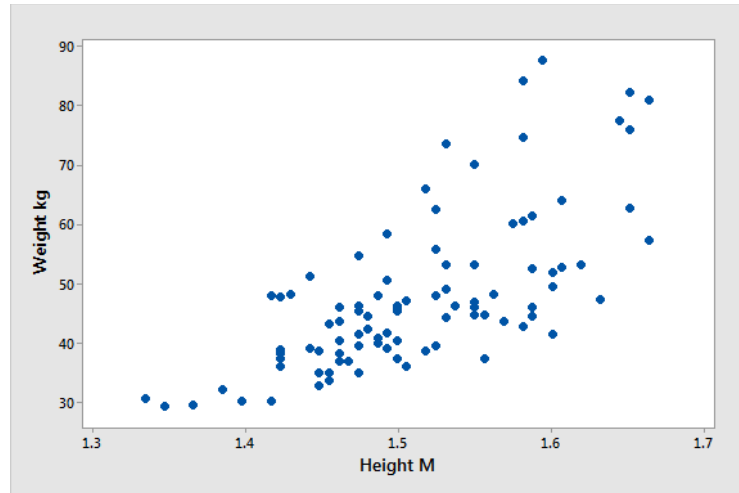
$$\mu = \frac{\sum_{i=1}^N x_i}{N} \quad (\text{EC 1})$$

La mediana es el valor que se encuentra en el medio de la mitad de los grupos de datos, es decir, abajo de la mediana tenemos el 50 % de los datos y arriba de la mediana tenemos el otro 50 % de los datos.

- Dispersión

Con esta medida se puede determinar qué tan dispersos se encuentran los datos. Regularmente se utilizan los rangos o la desviación estándar como referencia. Otra forma útil de entender la dispersión de los datos es a través de visualizaciones o gráficos.

Figura 1. **Scatter Plot – Emisiones de CO₂ vs tamaño del motor**



Fuente: Frost. (2020). *Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries*.

La desviación estándar puede ser calculada con la siguiente ecuación:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N} - \mu^2} \quad (\text{EC } 2)$$

- **Asimetría**

Permite determinar si la distribución de los datos tiene forma simetría o dispersa, y con ello determinar la forma adecuada para tratar los datos.

- **Correlación**

La correlación permite conocer el grado de vinculación o asociación que existe entre dos o más variables. La correlación es importante para saber qué tanta influencia tiene una variable sobre otra variable.

La correlación entre dos variables indica en qué dirección tiende a cambiar una variable, cuando la otra variable relacionada cambia. Este comportamiento es útil porque permite predecir el valor de una variable basado en la otra variable.

Uno de los mecanismos más utilizados para identificar el grado de correlación, es realizar una gráfica de dispersión de ambas variables y visualizar si existe cierto comportamiento o tendencia.

El coeficiente de correlación de Pearson identifica si existe correlación entre dos variables, pero únicamente identifica las correlaciones lineales. El coeficiente puede variar su valor entre el rango de -1 a +1. Entre más cerca se encuentre el valor de -1 o +1, se dice que la correlación es más fuerte.

Si el valor del coeficiente es positivo, se dice que la línea tiene una tendencia hacia arriba o en ascenso, lo cual implica que si el valor de una de las variables incrementa, la variable correlacionada también aumenta. En un caso contrario sucede si el coeficiente es negativo, en donde el incremento en una de las variables significa un decremento en la otra variable correlacionada. Siempre es importante realizar una visualización de los datos, para determinar el tipo de relación que existe en los mismos.

La correlación no implica causalidad, únicamente describe cómo los cambios en una variable están asociados con los cambios en otra variable, pero no son causa directa de esos cambios.

7.1.2. Estadística inferencial

A diferencia de la estadística descriptiva, la estadística inferencial si trata de tomar la data para realizar una inferencia o estimación. La inferencia

estadística toma una muestra de una población, para tratar de inferir o estimar los parámetros de la población, generalizando, y a dichas estimaciones se les conoce como estadísticos.

Por la naturaleza de los estadísticos y por ser una estimación, se necesita un grado de confianza de estos estadísticos calculados reflejan con cierta certeza los parámetros de la población. Asociado a las inferencias, también se puede encontrar un grado de error asociado a las estimaciones.

Por estas razones, es necesario crear intervalos de confianza para cada uno de los valores estimados, con cierto grado de confianza, para determinar los parámetros de la población.

7.1.2.1. Medidas

Las medidas básicas que se pueden obtener y que serán utilizadas en el presente trabajo, son las medidas de la media y la desviación estándar de una muestra.

La ecuación para el cálculo de la media de una muestra es la siguiente:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (\text{EC 4})$$

La ecuación para el cálculo de la desviación estándar de una muestra es la siguiente:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (\text{EC 5})$$

7.1.2.2. Intervalos de confianza

Un intervalo de confianza es un rango de valores, con límite inferior y límite superior, en donde es probable que se encuentre el parámetro estimado de la población. Este intervalo contiene un margen de error alrededor del estadístico, o valor puntual, que se calcula para una muestra.

Los intervalos de confianza están asociados a un nivel de confianza, o nivel de error, el cual permite incorporar la incertidumbre y los márgenes de error del manejo de muestra a nuestro estadístico puntual acotando los posibles valores del parámetro poblacional.

- Nivel de confianza

El nivel de confianza es la probabilidad de que el parámetro poblacional se encuentre dentro de un intervalo dado $1 - \alpha$.

En donde α es la probabilidad de que el parámetro poblacional no se encuentre dentro del intervalo de confianza. También es conocido como el nivel de significancia.

- Intervalo de confianza para la media

Para determinar el intervalo de confianza para la media de una muestra, se utiliza la siguiente ecuación:

$$\bar{x} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} ; \bar{x} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \quad (\text{EC 6})$$

7.1.3. Regresión

Las regresiones son métodos o procesos estadísticos que permiten analizar y comprender la relación entre dos o más variables de interés. Este proceso hace uso de medidas estadísticas y un conjunto de gráficos, para tratar de analizar y comprender mejor los resultados.

La regresión trata de encajar un modelo, para poder explicar la variabilidad de los datos, sin embargo, hay variabilidad que no puede ser explicada por el modelo o el conjunto de datos con los que se cuenta. Debe de existir un límite o equilibrio entre la cantidad de variabilidad que puede ser explicada y el no excederse con el sobre encaje de un modelo.

Regularmente se nombran las variables dependientes como “y” y las variables independientes como “x”.

7.1.3.1. Tipos de variables

Los tipos básicos de variables en los métodos de regresión son los siguientes:

- Variables dependientes

Son las variables que se quieren explicar o describir, a través de las variables de las cuales dependen. Regularmente se ven como las variables, que son descritas por los modelos predictivos que son construidos durante la regresión.

- Variables independientes

Son las variables que permiten explicar las variables dependientes a través de un modelo predictivo. Cuando se utilizan experimentos, estas variables pueden ser fijadas y modificadas por el observador. También se les conoce como variables predictoras.

7.1.3.2. Regresión lineal

La regresión lineal es el caso más simple de regresión. En este caso se pueden tener una sola variable independiente, a lo cual se le denomina regresión lineal simple, o se pueden tener múltiples variables independientes, denominado multilíneal.

El coeficiente de correlación tiene un papel fundamental, en estos modelos de regresión y se le llama frecuentemente coeficiente de correlación de la muestra. El coeficiente de correlación se puede calcular con la siguiente ecuación:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{EC 7})$$

- Regresión lineal simple

En este tipo de regresión, únicamente se cuenta con una variable dependiente y una variable independiente. Es el caso más simple de regresión. La regresión simple está descrita por la siguiente ecuación:

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i \quad (\text{EC } 8)$$

En donde β_0 y β_1 son los coeficientes de correlación y ε_i se denomina error.

- **Regresión multilínea**

Para los modelos de regresión multilínea, se cuenta con una variable dependiente y más de una variable independiente. Para estos casos, la ecuación que describe este modelo es la siguiente:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad (\text{EC } 9)$$

Cuando dos variables independientes se encuentran muy relacionadas, pueden llegar a encontrarse en la misma línea, a lo cual se les denomina colineales o se dice que existe multicolinealidad. La colinealidad es algo que debe de evitarse al construir un modelo de regresión multilínea.

7.1.3.3. Regresión no lineal

Las regresiones no lineales, son aquellas regresiones en donde el modelo no es representado por rectas o planos como en el caso anterior. Estos modelos pueden incluir polinomios, funciones exponenciales, funciones logarítmicas, etc.

7.1.3.4. Selección del modelo

La selección del modelo que represente mejor los datos debe de contener el menor número de variables, debe de ser el más simple, pero que sea capaz de representar los datos según lo esperado. Para ello es necesario construir

varios modelos con la ayuda de gráficos y mediciones, para determinar cuáles modelos tienen el mejor ajuste con la mayor simplicidad en su constitución.

7.1.4. Diseño de experimentos

El diseño del experimento es una secuencia de pasos o etapas ordenadas, que permiten garantizar la obtención de los datos para un análisis y estudio posterior. El diseño del experimento es fundamental para garantizar el éxito del mismo.

7.1.4.1. Requisitos de un diseño experimental

Un diseño de experimentos debe de contar con al menos los siguientes requisitos: proveer estimaciones no sesgadas; los efectos deben de ser estimados al mismo tiempo que las estimaciones mismas y el diseño debe de ser lo más simple posible.

- Diseño del experimento

Se refiere a la característica de un diseño experimental, en donde un proceso puede ser repetido bajo características similares.

Se denomina balanceado, cuando el número de réplicas es igual para todos los tratamientos. Si en dado caso, las réplicas no son las mismas para todos los tratamientos, se dice que el experimento es desbalanceado.

- Replicación

Se refiere a la característica de un diseño experimental en donde un proceso puede ser repetido bajo características similares.

Se denomina balanceado cuando el número de réplicas es igual para todos los tratamientos. Si en dado caso, las réplicas no son las mismas para todos los tratamientos, se dice que el experimento es desbalanceado.

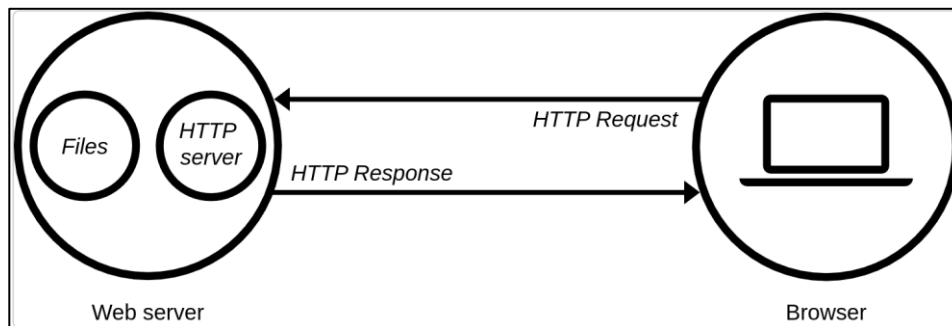
7.2. Plataforma Moodle

La plataforma Moodle es una plataforma de aprendizaje en línea, la cual permite personalizar ambientes de aprendizaje para estudiantes, docentes y administradores. La plataforma Moodle opera sobre un servidor web, con el lenguaje de programación PHP y bajo el protocolo HTTP o HTTPS.

7.2.1. Servidor Web

Un servidor web es un equipo de cómputo y/o software, el cual se encarga de recibir las peticiones de los usuarios, para acceder a ciertos recursos y de brindar el recurso, si se encuentra disponible o si el usuario tiene permisos de acceso a ese recurso. Regularmente están basados en el protocolo HTTP sobre TCP/IP.

Figura 2. **Representación básica de la conexión cliente/servidor**



Fuente: MDN Plus. (s.f.). *What is a web server?* Consultado el 20 de agosto de 2022.
Recuperado de https://developer.mozilla.org/en-US/docs/Learn/Common_questions/What_is_a_web_server

7.2.1.1. Rendimiento

Cada servidor web, está vinculado con cierta expectativa de rendimiento, es decir, se tiene un estimado de cuántos usuarios debería de poder soportar, sin tener problemas de funcionamiento. Por esta razón, es importante analizar el rendimiento del servidor previo a soportar las cargas que se esperan que pueda atender.

Para el propósito de poder medir el rendimiento de un servidor, se recomienda la construcción de un modelo basado en análisis, el cual permite la captura de aspectos relevantes del servidor web, y los servicios que tenga configurados, para luego transformarlos en un modelo predictivo de validación del rendimiento.

7.2.1.2. Modelo de rendimiento

En la práctica, los servidores web manejan una cola de solicitudes, con el objetivo de no quedarse sin recursos y poder brindar el servicio a todos los usuarios, aunque el servidor demore una cantidad considerable de tiempo. Generalmente se asume que la cola de servicio es infinita, para que esta variable dentro de un modelo no sea un factor crítico, ya que la cola de espera depende mucho de la aplicación web con la que cuente el servidor y los recursos del mismo.

Los modelos regularmente se centran en dos tipos específicos de requerimientos: la cantidad de tiempo de respuesta para un cliente y la cantidad de usuarios que puede atender al mismo tiempo.

7.2.1.3. Recursos de un servidor web

Los servidores web, cuentan con un conjunto de recursos limitados y que son los recursos que tienden a limitar los servicios a los usuarios y/o causar interrupciones del servicio. Sobre los recursos se puede mencionar los siguientes:

- Velocidad del CPU

La velocidad del CPU se refiere a la cantidad de ciclos que el CPU ejecuta por segundo. Durante un ciclo el CPU es capaz de abrir y cerrar miles o millones de transistores dentro del CPU. Al abrir y cerrar los transistores, los CPU son capaces de ejecutar tareas a bajo nivel. La velocidad con la que se ejecutan los ciclos se mide en gigahercios.

- Cantidad de hilos por CPU

Los hilos o núcleos que puede trabajar un CPU se refieren a las unidades de procesamiento por cada CPU con el que cuenta el equipo. Básicamente es cada operador que puede tener un procesador, para procesar alguna solicitud o tarea durante cada cambio de ciclo del mismo. Puede verse como un punto de venta de un supermercado, en donde puede haber varios puntos de venta y cada uno de ellos atender a un cliente por separado.

- Memoria RAM

La memoria RAM es la memoria que utiliza un equipo de cómputo, para almacenar de forma temporal la información, que se encuentra utilizando o procesando para poder realizar ciertas tareas de forma eficiente y eficaz. A diferencia de otros dispositivos de almacenamiento, la memoria RAM cuenta con una gran velocidad.

- Velocidad de escritura y lectura del disco duro

La velocidad de escritura de un disco duro se refiere a la velocidad, con la que puede almacenar la información en el disco duro y la velocidad de lectura, se refiere a la velocidad con la que puede recuperarla del disco duro a la memoria RAM. La velocidad se mide en Megabits por segundo.

- Velocidad de la red

La velocidad de la red está compuesta por la velocidad en la red interna de datos entre servidores y la puerta de enlace (por lo regular es bastante alta la

velocidad en la red de servidores), y el ancho de banda externo al Internet (el cual es limitado por lo general).

Para este estudio, la red no es un inconveniente por tratarse de un servicio en la nube y, por lo tanto, se puede eliminar esta limitante para el presente trabajo.

- Otros recursos

Los servidores web, cuentan con otros recursos como lo son los trabajos, la cantidad de conexiones disponibles, la cantidad de conexiones a base de datos, el *buffer* de datos, etc. Para el presente trabajo, se asume que estos están configurados de tal forma que, no son una limitante en el rendimiento del servidor.

7.2.2. Plataforma Moodle

La plataforma Moodle, es una plataforma de aprendizaje en línea, la cual permite personalizar ambientes de aprendizaje para estudiantes, docentes y administradores. La plataforma Moodle funciona sobre un servidor web, el cual puede ser de la elección de los administradores y está escrito en el lenguaje de programación PHP.

Moodle ha sido utilizado por pequeñas y grandes instituciones educativas, alcanzando gran popularidad en los últimos años.

La Universidad de San Carlos se ha unido al uso de la plataforma Moodle, como plataforma principal con la Red de Apoyo Digital Docente -RADD-, al igual que otras Facultades, como la Facultad de Arquitectura.

7.2.2.1. Tipo de licencia

Moodle es un software libre, sin licencias de pago, bajo la *GNU General Public License*. Cualquiera puede modificar, adaptar y publicar la plataforma para usos comerciales o no comerciales.

7.2.2.2. Escalabilidad

Los creadores de Moodle indican que, la plataforma es escalable a cualquier nivel, siempre y cuando la infraestructura lo permita.

7.2.2.3. Rendimiento

La carga de los servidores Moodle, depende mucho de la carga de cada usuario concurrente y activos. Con usuarios concurrentes, se refiere a los usuarios que están activos, pero que se encuentran realizando alguna solicitud de recursos a la plataforma Moodle.

La sección de preguntas frecuentes de Moodle indica que, tomar los usuarios únicamente concurrentes, no es una métrica adecuada y esto se debe a que, un usuario puede encontrarse como activo en la plataforma, pero puede que únicamente esté leyendo un documento o que tenga la sesión abierta sin realizar alguna actividad.

En contraste con los estudiantes que no se encuentran realizando ninguna actividad o que se encuentran leyendo algún material, se encuentran los estudiantes que pueden estar realizando alguna evaluación en línea, los cuales estarán realizando peticiones al servidor de forma constante y demandando recursos.

También se recomienda, la utilización de los *logs* generales del sistema operativo y los *logs* del servidor web, como otro recurso de medición para determinar el rendimiento de los servidores Moodle.

7.2.2.4. Mecanismos para determinar el rendimiento

Entre los mecanismos para determinar el rendimiento de un servidor Moodle se encuentran los *benchmarks*, que no es nada más que un software que se encarga de simular muchos usuarios concurrentes, solicitando recursos del sistema al mismo tiempo y durante un intervalo de tiempo determinado, por la persona que realiza el estudio.

En estos casos, un *benchmark* puede no ser un mecanismo adecuado para determinar el rendimiento de un servidor, ya que este software realiza peticiones a una página en específico del servicio, la cual puede no representar la carga real que soporta una plataforma por usuario. En este orden de ideas, es posible hacer un contraste entre una solicitud de la página de inicio, donde un servidor web únicamente se encarga de brindar la página de inicio, que es relativamente ligera y se puede comparar con una página de un examen, en donde se debe de validar el usuario, la sesión del usuario, los permisos, la cantidad de preguntas disponibles, la cantidad de preguntas restantes, el tiempo utilizado, la cantidad de intentos disponibles, etc.

La cantidad de carga sobre un servidor está vinculada directamente con la cantidad de los usuarios y la actividad de los mismos.

7.2.3. Pruebas de software

La definición de pruebas de software es verificar que el software, hace lo que tiene que hacer o que sigue las especificaciones. Esta definición puede ser ampliada cuando se habla en el contexto de no solo pruebas de funcionamiento, sino también pruebas de rendimiento.

Las pruebas de rendimiento son una buena práctica de todo software y en los últimos años se han ido integrando al paquete de pruebas básicas (pruebas unitarias, de aceptación, etc.).

Hoy en día la automatización de las pruebas es un enfoque vital para todo software, que es creado para cualquier propósito. Las pruebas ayudan a encontrar problemas, inconsistencias, comportamientos inadecuados, características de rendimiento, etc.

7.2.3.1. Tipos de pruebas

- Pruebas unitarias

Verifican que cada unidad mínima del software funcione correctamente, y están relacionadas directamente con el código fuente de las aplicaciones.

- Pruebas funcionales

Recrean escenarios de funcionamiento del negocio y tratan de validar la funcionalidad completa de una función del negocio.

- Pruebas de aceptación

Verifican que todo el sistema o módulo funcione de la forma esperada. Estas pruebas regularmente se encuentran a cargo de los clientes y/o interesados.

- Pruebas de integración

Verifica que los módulos o componentes del software interactúen de forma correcta. Se busca determinar fallas entre la integración, de los diferentes componentes del software.

- Pruebas de usabilidad

Verifica o evalúa que tan bien puede utilizar la aplicación un cliente. Se centra en la experiencia de usuario del cliente con la plataforma.

- Pruebas de rendimiento

Trata de evaluar cómo se comporta el sistema, bajo escenarios reales de carga de peticiones y recursos. Regularmente se realizan pruebas de cargas, para simular un escenario o escenarios de la vida real. Las cargas pueden ser medidas desde diferentes perspectivas, para determinar el rendimiento en diferentes escenarios.

- Pruebas de estrés

Valida que tanta carga, puede soportar una plataforma, antes de llegar al punto de fallo.

7.2.3.2. Pruebas web automatizadas

Las pruebas web automatizadas, utilizan exploradores web automatizados, a través de una aplicación de simulación de navegadores web y su comportamiento.

El fin principal de las pruebas con exploradores web automatizados, es determinar cómo se comporta un sistema, bajo escenarios de la vida real, con las condiciones de los usuarios reales.

Para que una prueba tenga un resultado satisfactorio, es necesario diseñar escenarios de pruebas que contengan un conjunto de interacciones comunes de los usuarios de las plataformas.

- Selenium

Selenium es una sombrilla para un conjunto de herramientas y librerías. Tiene soporte para muchos lenguajes de programación y utiliza *WebDriver* como componente principal.

Permite la creación de componentes programables para la navegación web, con la posibilidad de llenar formularios de ingreso a portales, navegación automatizada y llenado de formularios automatizados.

Con Selenium pueden construirse pruebas completas, de un sistema bajo escenarios reales de navegación y carga. Los componentes construidos en Selenium, pueden ser replicados para realizar pruebas de estrés o de rendimiento.

7.2.3.3. Métricas de rendimiento

Las métricas más importantes para las pruebas de software web son las siguientes:

- Tiempo de respuesta
- Tiempo de espera
- Tasa de error (en la conexión de red)
- Tasa más alta de respuesta
- Peticiones por segundo
- Transacciones satisfactorias e insatisfactorias
- Utilización del CPU
- Usuarios concurrentes
- Memoria utilizada
- Uso del disco duro

8. PROPUESTA DE ÍNDICE DE CONTENIDOS

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

LISTA DE SÍMBOLOS

GLOSARIO

RESUMEN

PLANTEAMIENTO DEL PROBLEMA

OBJETIVOS

RESUMEN DE MARCO METODOLÓGICO

INTRODUCCIÓN

1. MARCO REFERENCIAL

2. MARCO TEÓRICO

2.1. Estadística

2.1.1. Estadística descriptiva

2.1.1.1. Utilidad

2.1.1.2. Medidas

2.1.2. Estadística inferencial

2.1.2.1. Medidas

2.1.2.2. Estadística inferencial

2.1.2.3. Regresión

2.1.2.4. Diseño de experimentos

2.2. Moodle

2.2.1. Servidor Web

2.2.1.1. Rendimiento

- 2.2.1.2. Modelo de rendimiento
- 2.2.1.3. Recursos de un servidor web
- 2.2.2. Plataforma Moodle
 - 2.2.2.1. Tipo de licencia
 - 2.2.2.2. Escalabilidad
 - 2.2.2.3. Rendimiento
 - 2.2.2.4. Mecanismos para determinar el rendimiento
- 2.2.3. Pruebas de software
 - 2.2.3.1. Tipos de pruebas
 - 2.2.3.2. Pruebas web automatizadas
 - 2.2.3.3. Métricas de rendimiento

3. PRESENTACIÓN DE RESULTADOS

4. DISCUSIÓN DE RESULTADOS

CONCLUSIONES

RECOMENDACIONES

REFERENCIAS

APÉNDICES

9. METODOLOGÍA

A continuación, se describen las características, variables y fases que componen la investigación:

9.1. Características del estudio

El enfoque del estudio propuesto es cuantitativo, dado que la naturaleza de las variables es de tipo cuantitativo y no se encuentra ninguna variable de tipo cualitativa en el estudio y la finalidad es la de modelizar.

El diseño adoptado será experimental, dado que la información que será utilizada para la construcción del modelo será generada a través de una serie de experimentos en ambientes controlados y con variables controladas, es decir, se produce bajo circunstancias propicias. El estudio también manipulará variables independientes, tales como, la cantidad de usuarios concurrentes y la cantidad de recursos disponibles. El experimento es puro, ya que existe la manipulación de variables independientes, la medición de variables dependientes, control y validez de los valores obtenidos, se harán grupos de comparación, alterando las características de los servidores y usuarios concurrentes, y los usuarios para la carga al servidor serán generados al azar, con un comportamiento al azar.

El alcance de estudio es de tipo longitudinal de tendencia, ya que es de vital interés los cambios en las variables a través del tiempo y las relaciones entre éstas, y la recolección de los datos se realizará en diferentes momentos, capturando el estado de las variables completas. El componente de tendencia se

refiere a la evolución del consumo de recursos y la tendencia a agotarlos, causando deterioros en el servicio brindado por el servidor Moodle.

El alcance también es de tipo descriptivo correlacional, dado que se estudiará el efecto o relación de las variables independientes sobre las dependientes, correlación, y descriptivo porque busca describir el comportamiento de las variables y sus características.

La técnica de investigación será de campo y de observación directa, esto debido a que el investigador tomará los datos de primera mano, durante el experimento y también participará de forma directa en el desarrollo y control de las variables.

La observación experimental es de tipo *Ex Post Factum*, debido a que se desea verificar la carga que producen los usuarios concurrentes sobre un servidor Moodle, en donde el fenómeno conocido es la saturación de los recursos del servidor.

9.2. Unidades de análisis

La población del estudio estará constituida por el conjunto de datos capturados a través del tiempo del servidor de Moodle, al realizar las cargas de usuarios concurrentes, durante los tiempos controlados de pruebas. Para este estudio, se utilizará el conjunto completo de datos recolectados durante el experimento, por lo que no es necesario realizar muestreos de la población.

9.3. Variables

En la tabla I se presenta la operativización de variables.

Tabla I. **Operativización de variables**

Variable	Definición teórica	Definición operativa
Número de estudiantes concurrentes (n)	Un usuario concurrente es un usuario para el cual un servidor se encuentra realizando una operación o tarea activa. Variable cuantitativa discreta.	Variable que será controlada para simular usuarios concurrentes en el servidor Moodle y con ello poder obtener las variables independientes. Escala de razón. Variable independiente.
Tiempo en segundos (s)	Es el periodo en el cual se hace una captura del estado del servidor y el uso de sus recursos. Variable cuantitativa continua.	Tiempo medido en segundos, controlado, y representa el tiempo en el que se toma una captura del estado del servidor y el uso de sus recursos en un momento dado. Escala de razón. Variable independiente.
Uso del CPU (U_{cpu})	Es una representación de la capacidad y el uso de la capacidad de procesamiento de información de un servidor. Variable cuantitativa continua.	Representa el porcentaje del uso de la capacidad total del CPU. Escala de intervalo con valores entre 0 y 100 %. Variable dependiente.
Uso de la memoria RAM (U_{mem})	Es una representación de la capacidad y el uso de la capacidad de almacenamiento de información en la memoria RAM de un servidor. Variable cuantitativa continua.	Representa el porcentaje del uso de la capacidad total de la memoria RAM. Escala de intervalo con valores entre 0 y 100 %. Variable dependiente.
Uso del disco duro (U_{hdd})	Es una representación de la capacidad y el uso de la capacidad de la velocidad de escritura y lectura de información de un disco duro de un servidor. Variable cuantitativa continua.	Representa el porcentaje del uso de la capacidad total de la velocidad de lectura y escritura del disco duro. Escala de intervalo con valores entre 0 y 100 %. Variable dependiente.
Cantidad de tiempo utilizado para resolver un examen en línea (t_e)	El periodo que le toma a un estudiante realizar una prueba en línea en un servidor Moodle. Variable cuantitativa discreta.	Tiempo estimado en minutos. Es calculado con la base de datos historia de exámenes de la Facultad de Arquitectura. Escala de razón. Variable independiente.

Fuente: elaboración propia.

9.4. Fases del estudio

Para la presente investigación se desarrollarán los siguientes seis capítulos:

- Fase 1: revisión de literatura

En la primera fase se realizará una revisión bibliográfica de temas relacionados con el estudio, los cuales servirán de base para el desarrollo de la investigación.

- Fase 2: recolección y análisis de la base de datos histórica de Moodle

Durante esta fase, se realizará la recolección de la base de datos histórica de Moodle, para la estimación de la media del tiempo utilizado por los estudiantes para realizar una prueba en línea. En esta etapa se utilizan las técnicas de estadística descriptiva e inferencial, para determinar comportamientos y entender mejor la información.

- Fase 3: recolección y análisis de los datos recolectados durante las cargas a Moodle

Durante esta fase, se llevará a cabo el experimento de cargas de usuarios concurrentes al servidor Moodle y se recolectará la información necesaria, para luego construir los modelos matemáticos. Durante esta fase también se aplicarán las técnicas de estadística descriptiva e inferencial, para determinar comportamientos y entender mejor la información.

- Fase 4: construcción de los modelos matemáticos

Durante esta fase, se estarán construyendo una serie de modelos matemáticos, con la base de datos de las pruebas de la fase anterior. Los datos serán divididos en dos grupos: el grupo de entrenamiento y el grupo de pruebas.

Primero serán construidos los modelos lineales, utilizando las variables independientes, que tengan mayor influencia sobre la variable dependiente y de forma independiente, para generar modelos lineales simples. Posteriormente se estará construyendo modelos multilineales con las variables independientes.

Adicional a las regresiones lineales, serán construidos modelos no lineales, con el apoyo de visualizaciones de datos tales como los gráficos de dispersión, para encontrar patrones, enfoque dirigido por datos, y decidir cuál grado del polinomio utilizar. Serán construidos varios modelos, con diferentes grados, para evitar el sobre entrenamiento de los modelos.

- Fase 5: pruebas de desempeño de los modelos matemáticos

Durante esta fase, se examinarán los criterios de aceptación de los modelos, el desempeño de los modelos con la separación de un conjunto de datos, para entrenamiento y un conjunto de datos para pruebas.

Para todos los modelos, se estarán realizando pruebas de precisión con los datos de pruebas. De las pruebas de precisión se generarán los errores cuadráticos medios, RMSE, con el objetivo de verificar cuál de los modelos tiene una mejor precisión en la predicción. El error cuadrático medio será utilizado para verificar la precisión, en el conjunto de datos de prueba y el conjunto de datos de entrenamiento.

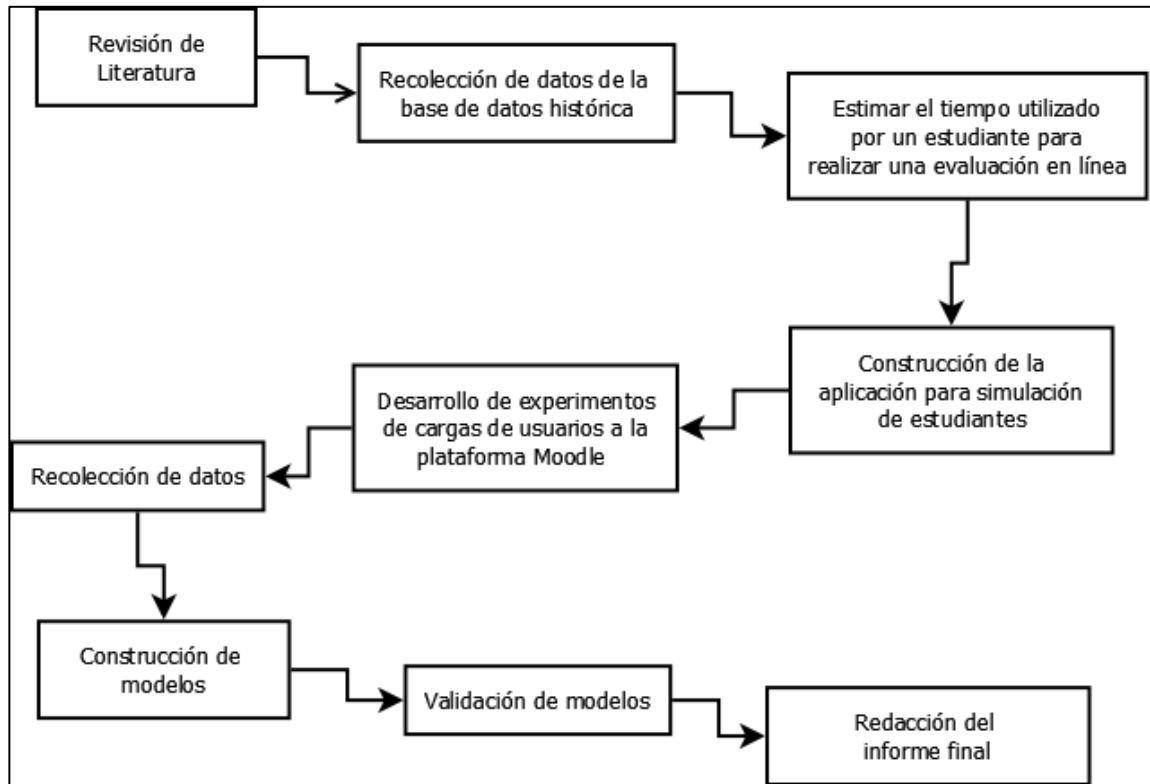
Para los modelos multilineales se usarán como referencia las siguientes suposiciones para determinar el mejor modelo:

- Linealidad entre la variable dependiente y las independientes a través de gráficos de dispersión.

- Normalidad multivariada: los residuos se distribuyen de forma normal. Se utilizarán gráficos de QQplot y pruebas de normalidad como Shapiro-Wilk y Anderson-Darling.
 - Análisis de multicolinealidad: las variables independientes no están altamente correlacionadas. Se calculará el factor de inflación de la varianza (VIF).
 - Homocedasticidad: la varianza de los términos de error es similar en los valores de las variables independientes. En este caso, se utilizarán gráficos de residuos estandarizados frente a valores pronosticados.
- Fase 6: redacción del informe final.

A continuación, se puede observar un diagrama de flujo del esquema de la solución:

Figura 3. Esquema de solución



Fuente: elaboración propia, realizado con Microsoft Project 2019.

10. TÉCNICAS DE ANÁLISIS DE LA INFORMACIÓN

El estudio utilizara una serie de técnicas para su elaboración las cuales se describen a continuación:

10.1. Estadística descriptiva

La estadística descriptiva será utilizada para la construcción de visualizaciones e información general de la información para poder determinar cuáles son los mejores enfoques para la gestión y análisis de la información. Con la estadística descriptiva también podemos determinar información faltante, valores atípicos y la distribución de los datos.

10.2. Estimación de medias

La estimación de medias será utilizada para determinar la media del tiempo, que un estudiante utiliza para resolver un examen en línea. Para ellos se construyen intervalos de tiempo y serán utilizados en la aplicación, para la generación de usuarios concurrentes.

Primero se debe de determinar si los tiempos utilizados por los estudiantes, para realizar una prueba en línea siguen una distribución normal. Si los datos siguen una distribución normal, la estimación de la media se realizará con la ecuación 4, 5 y 6 del marco teórico. Si los datos no siguen una distribución normal, se construirá el intervalo con el valor puntual de la ecuación 4 y dos veces el error estándar, calculado con la ecuación 5.

10.3. Correlación

Para determinar la importancia de las variables independientes en el estudio, se calculará el coeficiente de correlación de Pearson, y de esa forma, poder medir el grado de asociación entre las variables.

10.4. Pruebas de bondad de ajuste

A través de pruebas de bondad de ajuste, se determinará si los datos se ajustan a la distribución normal. Este estudio se realizará a través de R y a través de múltiples pruebas específicas y generales.

Se usará la prueba de Shapiro-Wilk en donde se evalúa si el p-valor es menor al 5 %. Si el p-valor es menor que el 5 %, los datos no siguen una distribución normal.

También se utilizará la prueba de Anderson-Darling, en donde también se examina el p-valor, y si éste es menor al 5 %, los datos no siguen una distribución normal.

10.5. Regresión lineal y multilineal

Las técnicas de regresión lineal y multilineal, serán utilizadas para la construcción de modelos de regresión. Los modelos serán sometidos a pruebas de rendimiento para determinar el mejor modelo.

Para las regresiones lineales se utilizará el coeficiente de correlación de Pearson, para evaluar el grado de relación entre las independientes y la variable dependiente.

Para las regresiones multilineales se utilizarán los estadísticos F y VIF, para determinar el valor explicativo de las variables y la colinealidad de las variables independientes. También serán utilizados los errores cuadráticos medios, para determinar la precisión de los modelos con el conjunto de datos de prueba.

10.6. Regresión no lineal

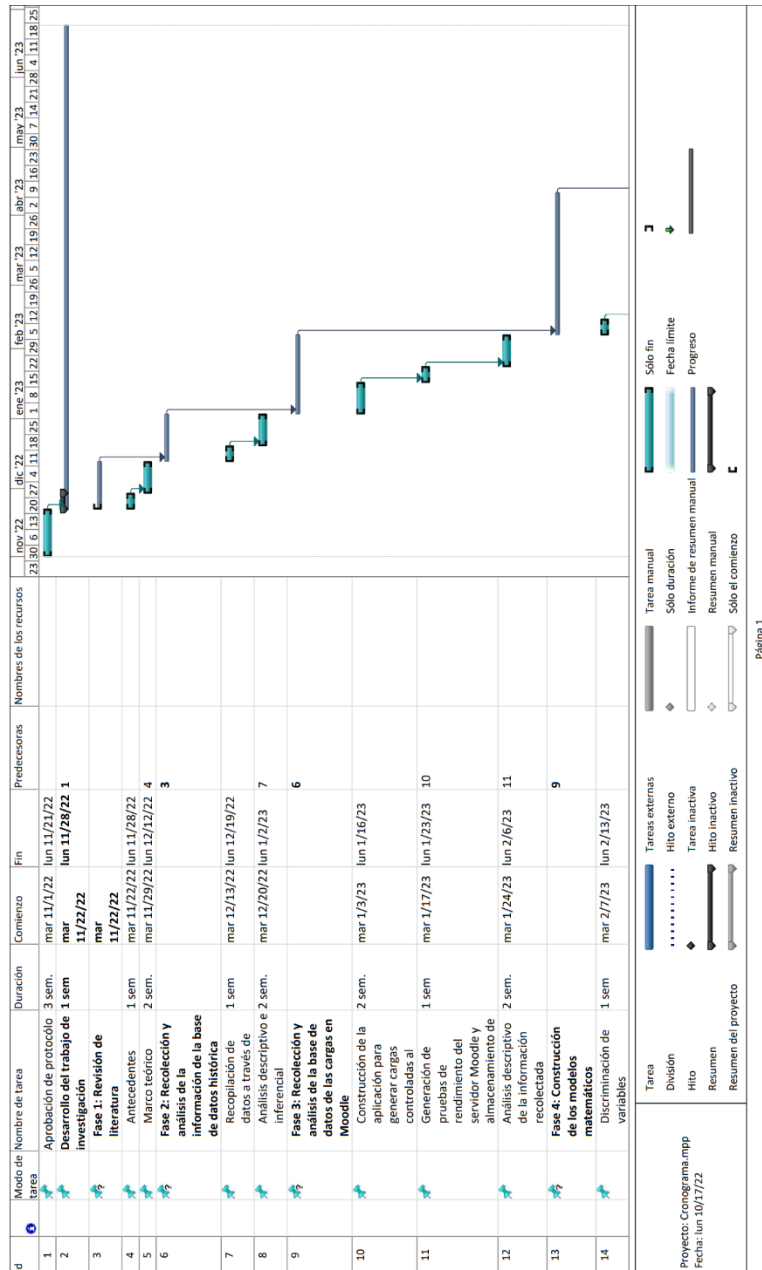
Las técnicas de regresión no lineal serán utilizadas para la construcción de modelos no lineales tales como, los modelos polinomiales, exponenciales, etc.

Para los modelos no lineales, se utilizará el error cuadrático medio para calcular el error en los datos de entrenamiento, como en los datos de prueba. Esto para evitar el sobreajuste y comparar la precisión de cada uno de los modelos que se construyan durante esta etapa. Adicionalmente se utilizará el coeficiente de determinación R^2 .

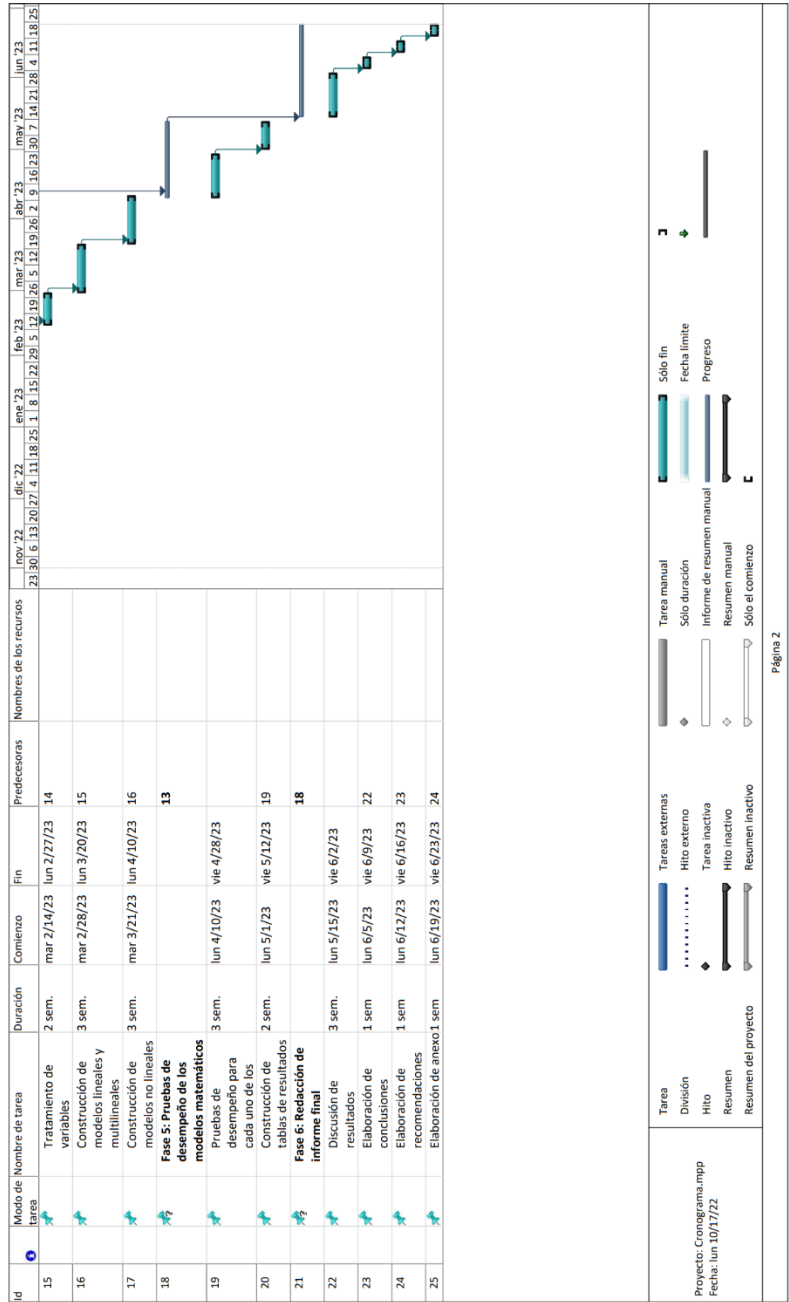
También se utilizará la estadística descriptiva para realizar gráficos y verificar que los gráficos tengan valores lógicos en los pronósticos.

11. CRONOGRAMA

Figura 4. Cronograma de actividades



Continuación de la figura 4.



Fuente: elaboración propia, empleando Microsoft Project 2019.

12. FACTIBILIDAD DEL ESTUDIO

La investigación es factible porque se cuenta con todos los recursos necesarios para poder llevarse a cabo.

Los recursos necesarios para el desarrollo de la investigación se describen a continuación:

12.1. Recurso humano

Se refiere a todas las personas que llevarán a cabo las tareas necesarias para el desarrollo de la investigación.

12.2. Recursos tecnológicos

Se cuenta con todos los recursos tecnológicos necesarios para poder desarrollar la investigación. El software que se utilizaran para encontrar los diferentes estadísticos de la investigación, gráficos y pruebas son: R Studio, Microsoft Excel, Moodle y Python.

12.3. Acceso a información y permisos

Para la investigación, se cuentan con los permisos necesarios para acceder a la información de la base de datos histórica de Moodle y los permisos necesarios para el acceso a la infraestructura.

12.4. Equipo e infraestructura

Se cuenta con el equipo e infraestructura necesaria para realizar la investigación. El equipo para el desarrollo de la investigación lo estará proporcionando el investigador. La infraestructura de la plataforma Moodle, la estará brindando la Facultad de Arquitectura, de la Universidad de San Carlos de Guatemala.

12.5. Recursos financieros

En esta sección se describen los recursos financieros, que son necesarios para desarrollar la investigación. La totalidad de los costos financieros serán cubiertos por el investigador.

Tabla II. **Detalle de los recursos financieros**

No.	Recurso	Descripción	Costo	Cantidad	Total
1	Humano	Tiempo invertido por el investigador (en meses)	Q 8,000.00	6	Q 48,000.00
2	Servicios	Internet	Q 250.00	6	Q 1,500.00
3	Servicios	Luz	Q 300.00	6	Q 1,800.00
4	Humano	Asesor	Q 0.0	6	Q0.0
				Total	Q 51,300.00

Fuente: elaboración propia

13. REFERENCIAS

1. Bezemer, C. y Zaidman A. (junio, 2011). Server overload detection and prediction using pattern classification. *8th International Conference on Autonomic Computing*, 1-2. Recuperado de: <https://repository.tudelft.nl/islandora/object/uuid%3Ab7914e96-73be-456c-9f7e-a07eef65d25a>.
2. Black, R., Veenendaal, E. y Graham, D. (2012). *Foundations of Software Testing ISTQB Certification*. Hampshire, Reino Unido: Cengage Learning EMEA.
3. Bondrenko, A. (1 de febrero de 2021). Python Load Testing Best Practices. [Mensaje de blog]. Recuperado de: <https://stackify.com/python-load-testing-best-practices/>.
4. Corrales, D. y Ledezma, A. (marzo, 2018). How to Address the Data Quality Issues in Regression Models: A Guided Process for Data Cleaning. MDP. *Symmetry*, 10(99), 1-20. Recuperado de <https://www.mdpi.com/2073-8994/10/4/99>.
5. Frost, J. (2020). *Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries*. Pensilvania, Estados Unidos: Statistics By Jim Publishing.

6. IBM. (s.f.). What is Software Testing and How Does it Work? | IBM. [Mensaje de blog]. Recuperado de: <https://www.ibm.com/topics/software-testing>.
7. James, G, Witten, D., Hastie, T. y Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Nueva York, Estados Unidos: Springer.
8. Khan, M., Algarni, F., Kumar, I., Choudhary, A. y Srivastava, A. (enero, 2022). Performance evaluation of regression models for COVID-19: A statistical and predictive perspective. *Ain Shams Engineering Journal*, 13(2022), 1-18. Recuperado de: <https://www.sciencedirect.com/science/article/pii/S2090447921003385?via%3Dihub>
9. Lloyd, G. y Connie, U. (2002). *Five Steps to Solving Software Performance Problems*. Nuevo México, Estados Unidos: Performance Engineering Services. Recuperado de: <http://www.perfeng.com/papers/step5.pdf>
10. Lu, J. y Gokhale, S. (2006). Web server performance analysis. *International Journal of Information Technology and Web Engineering (IJITWE)*, 3(3), 1-16. Recuperado de <https://www.igi-global.com/article/performance-analysis-web-server/2653>.
11. McElreath, R. (2020). *Statistical Rethinking A Bayesian Course with Examples in R and Stan*. Florida, Estados Unidos: CRC Press.

12. Melo, O., López, L. y Melo, S. (2020). *Diseño de Experimentos Métodos y Aplicaciones*. Bogotá, Colombia: Universidad Nacional de Colombia Facultad de Ciencias.
13. Moodle. (15 de septiembre de 2019). Performance FAQ. [Mensaje de blog]. Recuperado de: https://docs.moodle.org/400/en/Performance_FAQ
14. Moodle. (15 de septiembre de 2022). About Moodle. [Mensaje de blog]. Recuperado de: https://docs.moodle.org/400/en/About_Moodle
15. Palmer, P. y O'Connell D. (2009). Regression Analysis for Prediction: Understanding the Process. *Cardiopulmonary Physical Therapy Journal*, 20(3), 23-6. Recuperado de: https://www.researchgate.net/publication/44596956_Regression_Analysis_for_Prediction_Understanding_the_Process.
16. Praire, Y. (marzo, 1996). Evaluating the predictive power of regression models. *Canadian Journal of Fisheries and Aquatic Sciences*, 53(3), 490-492. Recuperado de <https://doi.org/10.1139/cjfas-53-3-490>.
17. Rajalaxmi, R., Natesan, P., Krishnamoorthy, N. y Ponni, S. (abril, 2019). Regression Model for Predicting Engineering Students Academic Performance. *International Journal of Recent Technology and Engineering (IJRTE)*, 7, 71-75. Recuperado de: https://www.researchgate.net/publication/333102911_Regression_Model_for_Predicting_Engineering_Students_Academic_Performance.

18. Selenium. (3 de marzo de 2022). About Selenium | Selenium. [Mensaje de blog]. Recuperado de: <https://www.selenium.dev/about/>.
19. Selenium (3 de marzo de 2022). The Selenium Browser Automation Project | Selenium. [Mensaje de blog]. Recuperado de: <https://www.selenium.dev/documentation/>.
20. Selvamuthu, D. y Das, D. (2018). *Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control*. Singapore, Singapore: Springer.
21. Senthilnathan, S. (julio, 2019). Usefulness of Correlation Analysis. *International Training Institute*, 1-9. Recuperado de: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3416918.
22. Sole, M. (noviembre, 2013). A Primer for Mathematical Modeling. *Journal of Mathematics Education at Teachers College*, 4(2), 44-49. Recuperado de <http://dx.doi.org/10.7916/jmetc.v4i2.629>.
23. Sun, Y. y Chen, X. (mayo, 2016). Online Education and Its Effective Practice: A Research Review. *Journal of Information Technology Education: Research*, 15, 157-190. Recuperado de <https://www.informingscience.org/Publications/3502>.
24. TestingXperts. (10 de mayo de 2022). Performance Testing Metrics – A Detailed Guide for Businesses. [Mensaje de blog]. Recuperado de <https://www.testingxperts.com/blog/performance-testing-metrics>.

25. Uyanik, G. y Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia-Social and Behavioral Sciences*, 106, 234-240.
Recuperado de <https://doi.org/10.1016/j.sbspro.2013.12.027>

14. APÉNDICE

Apéndice 1. Matriz de coherencia

Elemento	Problema estadístico	Preguntas de investigación	Objetivos	Fundamento	Metodología
GENERAL	Se desconoce el rendimiento de los servidores Moodle cuando los estudiantes se encuentran en exámenes en línea.	¿Cuál es el modelo matemático que permite estimar el rendimiento de un servidor Moodle durante exámenes en línea?	Construir de un modelo matemático para estimar el rendimiento de un servidor Moodle durante el periodo de exámenes en línea de la Facultad de Arquitectura de la Universidad de San Carlos de Guatemala.	Se evaluarán varios modelos y dependiendo del comportamiento de los datos y el ajuste de cada uno de los modelos, se usará el modelo con mejor rendimiento.	El enfoque de la investigación es cuantitativo porque se analizan recursos y cantidad de usuarios, con variables de tipo cuantitativo.
ESPECÍFICO	Se desconoce el tiempo promedio que utilizan los usuarios para realizar una evaluación en línea en Moodle.	¿Cuál es el tiempo promedio que le toma a un estudiante realizar una evaluación en línea a nivel universitario?	Estimar el tiempo promedio que le toma a un estudiante, a nivel universitario, realizar una prueba en línea, a través de un intervalo de confianza de la media en la plataforma Moodle de la Facultad de Arquitectura para parametrizar la herramienta de simulación de estudiantes concurrentes.	Se estimará la media de la información de la base de datos histórica de Moodle con un intervalo de confianza del 95%.	Con un diseño experimental, porque se controla la cantidad de variables, los estados iniciales y el experimento es repetible. De tipo descriptivo porque estudia las variaciones, características y propiedades de las variables dependientes e independientes.
	No se cuenta con herramientas especializadas para simular la carga de un usuario en la plataforma Moodle durante una evaluación.	¿Cuál es la carga promedio, en un servidor Moodle, por cada estudiante que realiza una evaluación en línea?	Diseñar una herramienta desarrollada en el lenguaje de programación Python de simulación de usuarios concurrentes en un ambiente controlado de Moodle para sobrecargarlo y estimar la media del consumo de recursos del servidor (RAM, uso de disco duro y uso de CPU).	Con un software que simule el comportamiento de un usuario en un escenario real, podemos obtener mejores mediciones de rendimiento y producir un modelo mucho más real.	El estudio también tiene un alcance descriptivo correlacional, porque se estudia la importancia de la relación entre las variables independientes y dependientes.
	No se conoce cuál es el impacto de los recursos del servidor de una plataforma Moodle en el rendimiento de la misma.	¿Cuál es la relación de los recursos de un servidor Moodle sobre el rendimiento de la cantidad de usuarios activos simultáneos durante una evaluación?	Cuantificar la relación que existe entre los recursos del servidor Moodle (RAM, CPU, velocidad del disco duro, etc.) y el rendimiento de la plataforma a través de pruebas de correlación y regresión para poder descartar las variables que no aportan demasiada información al modelo matemático.	Medir el impacto que tiene cada usuario sobre un servidor es de vital importancia para poder determinar la capacidad de un servidor Moodle basado en la cantidad de usuarios concurrentes.	

Fuente: elaboración propia.