

**UNIVERSIDAD PERUANA UNIÓN**  
FACULTAD DE INGENIERIA Y ARQUITECTURA  
Escuela Profesional de Ingeniería de Sistemas



**Optimización Bayesiana de modelos de machine learning para  
mejorar la predicción de clientes e-learning**

Tesis para obtener el Título Profesional de Ingeniero de Sistemas

**Autor:**

Jorge Daniel Maquera Canales

**Asesor:**

Magister Nemias Saboya Rios

Lima, setiembre 2023

## DECLARACIÓN JURADA DE ORIGINALIDAD DE TESIS

Yo Nemias Saboya Rios, docente de la Facultad de Ingeniería y Arquitectura Escuela Profesional de Ingeniería de Sistemas , de la Universidad Peruana Unión.

DECLARO:

Que la presente investigación titulada: **“OPTIMIZACIÓN BAYESIANA DE MODELOS DE MACHINE LEARNING PARA MEJORAR LA PREDICCIÓN DE CLIENTES E-LEARNING”** del autor Jorge Daniel Maquera Canales tiene un índice de similitud de 13% verificable en el informe del programa Turnitin, y fue realizada en la Universidad Peruana Unión bajo mi dirección.

En tal sentido asumo la responsabilidad que corresponde ante cualquier falsedad u omisión de los documentos como de la información aportada, firmo la presente declaración en la ciudad de Lima, a los 28 días del mes de Setiembre del año 2023



---

**NEMIAS SABOYA RIOS**

**ACTA DE SUSTENTACIÓN DE TESIS**

En Lima, Ñaña, Villa Unión, a los **06** día(s) del mes de **setiembre** del año 2023 siendo **las 10:00 horas**, se reunieron en modalidad virtual u online sincrónica, bajo la dirección del Señor Presidente del jurado: **Dra. Erika Inés Acuña Salinas**, el secretario: **Mg. Geraldine Verónica Alvizuri Llerena**, y los demás miembros: **PhD. Javier Linkolk Lopez Gonzales** y **el MSc. Fredy Abel Huanca Torres**, y el asesor, **Mg. Nemias Saboya Rios**, con el propósito de administrar el acto académico de sustentación de la tesis titulada: " Optimización bayesiana de modelos de aprendizaje automático para mejorar la predicción de clientes e-learning "

de el(los)/la(las) bachiller/es: a) **JORGE DANIEL MAQUERA CANALES**

..... b) .....

conducente a la obtención del título profesional de **INGENIERO DE SISTEMAS**

*(Nombre del Título profesional)*

con mención en.....

El Presidente inició el acto académico de sustentación invitando al (los)/a(la)(las) candidato(a)s hacer uso del tiempo determinado para su exposición. Concluida la exposición, el Presidente invitó a los demás miembros del jurado a efectuar las preguntas, y aclaraciones pertinentes, las cuales fueron absueltas por el(los)/la(las) candidato(a)s. Luego, se produjo un receso para las deliberaciones y la emisión del dictamen del jurado.

Posteriormente, el jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:

Candidato (a): ..... **JORGE DANIEL MAQUERA CANALES** .....

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	
<b>Aprobado</b>	<b>19</b>	<b>A</b>	<b>Con nominación de Excelente</b>	<b>Excelencia</b>

Candidato (b): ..... .....

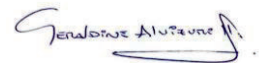
CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	

*(\*) Ver parte posterior*

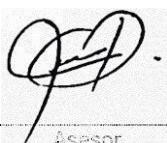
Finalmente, el Presidente del jurado invitó al(los)/a(la)(las) candidato(a)s a ponerse de pie, para recibir la evaluación final y concluir el acto académico de sustentación procediéndose a registrar las firmas respectivas.



Presidente  
Dra. Erika Inés Acuña Salinas



Secretario  
Mg. Geraldine Verónica Alvizuri Llerena



Asesor  
Mg. Nemias Saboya Rios



Miembro  
PhD. Javier Linkolk Lopez Gonzales



Miembro  
MSc. Fredy Abel Huanca Torres



Candidato/a (a)  
Jorge Daniel Maquera Canales

Candidato/a (b)

## Índice

<b>Resumen</b> .....	<b>5</b>
<b>Abstract</b> .....	<b>5</b>
<b>Introducción</b> .....	<b>7</b>
<b>Fundamento teórico</b> .....	<b>8</b>
<b>Enfoque bayesiano</b> .....	<b>8</b>
<b>Optimización bayesiana</b> .....	<b>8</b>
<b>Tree-Structured Parzen Estimator (TPE)</b> .....	<b>8</b>
<b>Técnicas de aprendizaje automático</b> .....	<b>9</b>
<b>Propuesta</b> .....	<b>9</b>
<b>Comprensión de los datos</b> .....	<b>10</b>
Descripción del conjunto de datos .....	10
Dataset preparation .....	12
<b>Bayesian Optimization for each algorithm</b> .....	<b>12</b>
<b>Performance Metrics</b> .....	<b>13</b>
<b>Resultados</b> .....	<b>14</b>
<b>Observaciones finales</b> .....	<b>18</b>
<b>Disponibilidad de los datos</b> .....	<b>19</b>
<b>Referencias</b> .....	<b>19</b>
<b>ANEXOS</b> .....	<b>21</b>

# **Optimización Bayesiana de modelos de machine learning para mejorar la predicción de clientes e-learning**

## **Bayesian optimization of machine learning models to improve e-learning customer prediction**

### **Resumen**

La puntuación de clientes potenciales desempeña un papel crucial en el marketing al evaluar el nivel de interés y compromiso de posibles clientes. Las técnicas de aprendizaje automático ofrecen un medio para automatizar los procesos de puntuación de clientes potenciales, permitiendo a los profesionales del marketing priorizar sus esfuerzos y asignar recursos de manera efectiva. Sin embargo, el rendimiento de los modelos de aprendizaje automático depende en gran medida de la configuración de sus hiperparámetros. Los métodos de optimización tradicionales pueden ser ineficientes y no lograr explorar eficazmente el espacio de hiperparámetros de alta dimensionalidad. En este estudio, proponemos un enfoque novedoso que utiliza la optimización bayesiana de hiperparámetros para mejorar la predicción de conversión de clientes en el ámbito del e-learning. Al aprovechar las estadísticas bayesianas y un modelo probabilístico, nuestro método explora eficientemente el espacio de hiperparámetros para identificar configuraciones óptimas que maximizan el rendimiento del modelo. Consideramos los algoritmos de aprendizaje automático Extreme Gradient Boosting, Support Vector Machine, Random Forest, Logistic Regression y Decision Tree, y comparamos los algoritmos optimizados con sus versiones base en términos de precisión, sensibilidad, puntuación F1 y el área bajo la curva característica de operación del receptor (AUC). Los resultados demostraron que los algoritmos optimizados superaron consistentemente a sus versiones base. Nuestra investigación destaca la importancia de la optimización de hiperparámetros para lograr un rendimiento óptimo del modelo de aprendizaje automático y proporciona información valiosa para los profesionales del marketing en la industria del e-learning. Al aprovechar los algoritmos optimizados, las organizaciones pueden tomar decisiones basadas en datos, maximizar las tasas de conversión y optimizar las estrategias de marketing.

Palabras clave: puntuación de clientes potenciales, e-learning, optimización bayesiana.

### **Abstract**

Lead scoring plays a crucial role in marketing by evaluating the level of interest and commitment of potential customers. Machine learning techniques offer a means to automate lead scoring processes, enabling marketers to prioritize their efforts and allocate resources effectively. However, the performance of machine learning models heavily relies on the configuration of their hyperparameters. Traditional optimization methods can be inefficient and fail to navigate the high-dimensional hyperparameter space effectively. In this study, we propose a novel approach using Bayesian hyperparameter optimization to enhance customer conversion prediction in the e-learning domain. By leveraging Bayesian statistics and a probabilistic model, our method efficiently explores the hyperparameter space to identify optimal configurations that maximize model performance. We considered the machine learning algorithms Extreme Gradient Boosting, Support Vector Machine, Random Forest, Logistic Regression, and Decision Tree, and compared the optimized algorithms against their base versions in terms of accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC). The results demonstrated that the optimized algorithms consistently outperformed their base