

EVALUACIÓN COMPARATIVA DE MODELOS DE REGRESIÓN SUPERVISADA EN MACHINE LEARNING PARA LA TOMA DE DECISIONES GERENCIALES

Pedro Pablo Chambi-Condori^{1*} y Miriam Chambi-Vásquez

2

¹Universidad Nacional Jorge Basadre Grohmann; ²Universidad Nacional Mayor de San Marcos

pchambic@unjbg.edu.pe

Resumen

La regresión multivariante es una herramienta fundamental en la toma de decisiones gerenciales, permitiendo a los líderes empresariales comprender y predecir comportamientos complejos mediante el análisis de múltiples variables simultáneamente. A continuación, se presenta un resumen de cómo esta técnica se integra con recursos de Machine Learning para la toma de decisiones gerenciales. El objetivo del estudio fue evaluar los diversos modelos regresionales ejecutados en Python bajo el enfoque de machine learning teniendo en cuenta las métricas de precisión que conduzcan a seleccionar el mejor modelo que permita estimar el impacto de las variables que tienen incidencia en el precio de una vivienda, información muy importante tanto para las empresas que desenvuelven en el mercado mobiliario y para las familias que enfrentan situaciones de compra de una vivienda. Para llevar a cabo ese propósito se seleccionó un caso de ventas de una empresa corporativa impactada por factores internos como costos variables, costos fijos y variables exógenas como la volatilidad de entorno, caso multivariante organizando en un dataset con 506 registros diarios. Esta data ha sido sometida al procesamiento bajo el enfoque de machine learning con diversos modelos, dividiendo la data, 70% para fines de entrenamiento del modelo y 30% para fines de evaluación. Los modelos utilizados para el análisis comparativo fueron: modelos de regresión múltiple Ridge Lasso, Árbol de decisiones, Random forest y Redes neuronales, con los que se han generado métricas de evaluación de los resultados obtenidos con los modelos, siendo

ellos: R^2 , MAE, MSE y RMSE para evaluar la precisión y fiabilidad del modelo. Obteniendo como resultados en orden de importancia, las mejores evaluaciones con modelos de Gradient Boosting y Random Forest.

Palabras clave: Información gerencial, machine learning, decisiones gerenciales.

Introducción

En un entorno empresarial cada vez más complejo y competitivo, la capacidad para tomar decisiones informadas y basadas en datos es crucial para el éxito y la sostenibilidad de las organizaciones. La toma de decisiones gerenciales efectivas se sustenta en el análisis preciso de información y la identificación de patrones y tendencias significativas. En este contexto, los modelos de regresión supervisada en machine learning emergen como herramientas poderosas para mejorar la calidad de las decisiones al proporcionar predicciones y análisis basados en datos históricos. Los modelos de regresión supervisada son algoritmos de machine learning diseñados para prever valores continuos a partir de datos de entrada. Estos modelos se entrenan utilizando un conjunto de datos etiquetados, donde las relaciones entre las variables independientes y la variable dependiente se aprenden para hacer predicciones precisas sobre nuevos datos. La aplicación de estos modelos en el ámbito gerencial puede transformar la forma en que las empresas abordan problemas como la estimación de ventas, la previsión de demanda, la identificación de riesgos y la optimización de recursos.

La presente evaluación comparativa se enfoca en analizar y comparar la eficacia de diferentes modelos de regresión supervisada en el contexto de la toma de decisiones gerenciales. A través de una revisión exhaustiva de técnicas como la regresión lineal, la regresión polinómica, la regresión de soporte vectorial y los árboles de decisión, entre otros, se busca identificar cuál de estos modelos ofrece el mejor rendimiento en términos de precisión, robustez y aplicabilidad a casos reales de toma de decisiones. Esta investigación no solo evalúa el desempeño de los modelos en función de métricas estadísticas y técnicas, sino que también considera aspectos prácticos,

como la interpretabilidad de los resultados y la facilidad de implementación en entornos empresariales. Al proporcionar una visión integral sobre las ventajas y limitaciones de cada modelo, esta evaluación pretende ofrecer recomendaciones prácticas para gerentes y profesionales que buscan integrar machine learning en sus procesos de toma de decisiones.

En relación con los antecedentes que tiene el tema, en la Tabla 1 se presenta el estado del arte que resume la evaluación comparativa de modelos de regresión supervisada en machine learning para la toma de decisiones gerenciales. Esta tabla incluye el título del artículo, los autores, el año de publicación y el medio de publicación, además, proporciona una visión general de los estudios relevantes en el campo de la evaluación comparativa de modelos de regresión supervisada, destacando investigaciones que pueden ser útiles para comprender las tendencias actuales y las mejores prácticas en el uso de machine learning para la toma de decisiones gerenciales.

El uso de métricas estadísticas, se evidenciaron en los estudios de Tatacher (2021) calificando a los modelos de regresión como técnicas poderosas de aprendizaje automático para hacer estimaciones y pronósticos de relaciones entre las variables mediante las métricas de evaluación de MSE, MAE y R^2 . Por otro lado, Kumar y Singh (2024) y Sihombing *et al.* (2024) también valoran positivamente el modelamiento de aprendizaje automático con modelos de regresión lineal, bosques aleatorios, regresión de vectores de soporte (SVR) entre otras. Asimismo, Aco *et al.* (2023) utilizaron las métricas MSE, MAE y R^2 , para los modelos regresión, redes neuronales, árbol de decisiones y SVR para modelar el caso de deserción de estudiantes universitarios.

Tabla 1. Estado del arte de modelamiento regresión lineal múltiple (Elaboración propia).

Título del Documento	Autores	Año	Medio de Publicación
Econometría.	Gujarati, D., Porter, C.	2010	Mc Graw Hill
Introduction to Machine Learning with Python.	Muller, A., Guido, S.	2016	O'reilly Inc.
Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción.	Valero <i>et al.</i>	2021	Revista de Ciencias Sociales (Ve)
Comparative Assessment of Regression Models Based On Model Evaluation Metrics.	Tatachar, A.	2021	International Research Journal of Engineering and Technology
Análisis comparativo de Técnicas de Machine Learning para la predicción de casos de deserción universitaria.	Aco <i>et al.</i>	2023	Revista Ibérica de Sistemas e Tecnologias de Informação
Introduction to the Use of Linear and Nonlinear Regression Analysis in Quantitative Biological Assays.	Jarantow <i>et al.</i>	2023	Current Protocols
An Introduction to Statistical Learning wirh Python.	James <i>et al.</i>	2023	Springer
A Comparative Analysis of Machine Learning Algorithms for Predictive Analytics in Healthcare.	Pal, S.	2024	Heritage Research Journal
Machine learning in business process management: A systematic literature review.	Weinzier <i>et al.</i>	2024	Expert Systems With Applications
Comparison of Regression Analysis with Machine Learning Supervised Predictive Model Techniques.	Sihombing <i>et al.</i>	2024	Jurnal Ekonomi Dan Statistik Indonesia
Analyzing the Impact of Machine Learning Algorithms on Risk Management and Fraud Detection in Financial Institution	Kumar, D., Singh, S.	2024	International Journal of Research Publication and Reviews

Objetivo

El objetivo del estudio fue evaluar los diversos modelos regresionales ejecutadas en Python bajo el enfoque de machine learning teniendo en cuenta las métricas de precisión que conduzcan a seleccionar el mejor modelo que permita estimar el impacto de las variables que tienen incidencia en el precio de una vivienda.

Materiales y métodos

La dataset utilizado para el análisis comparativo se obtuvo de scikit-learn.org con el nombre de “housing_data_for_regression.csv” que tiene 506 registros, 13 variables independientes que tienen incidencia en la variable dependiente de precio de ventas de viviendas.

Para el procesamiento de datos se utilizó el modelo de regresión lineal múltiple de la forma expresada en Gujarati & Porter (2010), el modelo de regresión múltiple que incluye a $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$

En donde Y: es la variable impactada por los factores que tienen en su comportamiento, que en este caso es el precio de una vivienda.

β : es el coeficiente que denota el grado de impacto en Y que expresa una variable explicando.

i : 1, ...n: número de factores que intervienen en el modelo.

ε : ruido blanco. (Gujarati y Porter, 188).

De acuerdo al procedimiento trazado en Muller y Guido (2016) y James *et al.* (2023), ya estando en el editor de Python se procedió a cargar las librerías, para luego se procedió a cargar la dataset al Jupyter Notebook o Google Colab, paso seguido se procedió a ejecutar el preprocesamiento de datos que incluye la limpieza haciendo que todos los datos sean del tipo numérico y que estén exentas de campos vacíos, para luego llevar a cabo la normalización de los datos, seguidamente se dividió la dataset en dos grupos: el grupo entrenamiento de 70% y 30% para el test, ejecutando las líneas de código en Python para para generar las métricas con diferentes modelos para evaluar y discriminar de entre todos los mejores indicadores obtenidos. Se utilizaron los modelos: regresión lineal, regresión Ridge, regresión Lasso, árboles de

decisiones, gradiente Boosting, regresión de soporte vectorial y redes Neuronales para obtener las métricas MSE (error cuadrático medio, MAE (diferencia entre el valor estimado y el valor real en cada punto estimado) y R^2 (coeficiente de determinación), para calificar la precisión del mejor modelo para el caso específico en estudio.

Resultados

Con la aplicación de los diversos modelos bajo el enfoque de machine learning se obtuvieron los resultados que se aprecian en la Tabla 2 y en la Figura 1 que a partir del análisis comparativo de los tres indicadores: MSE, MAE y R^2 , se concluye, en este caso específico el modelo que dio mayores valores de precisión en las tres métricas utilizadas, la tiene el modelo de Gradient Boosting seguido del modelo de Random Forest y teniendo en cuenta el coeficiente de determinación evidencia al 81.75% de ventas explicada por las variables independientes del modelo.

Tabla 2. Métrica de los modelos aplicados obtenida con los datos de la muestra.

MODELO	MSE	MAE	R^2
Linear Regression	16.4954	3.0559	0.7439
Ridge Regression	15.8978	3.0537	0.7532
Lasso Regression	23.1552	3.7513	0.6405
Decisión Tree	17.9383	3.1127	0.7215
Random Forest	13.5971	2.4489	0.7889
Gradient Boosting	11.7538	2.2844	0.8175
Support Vector Regression	48.1511	4.7470	0.2524
Neural Network	57.8383	6.6356	0.1020

Nota: MSE : error cuadrático medio. MAE: error absoluto medio.

R^2 ; coeficiente de determinación, medida de la precisión del modelo.

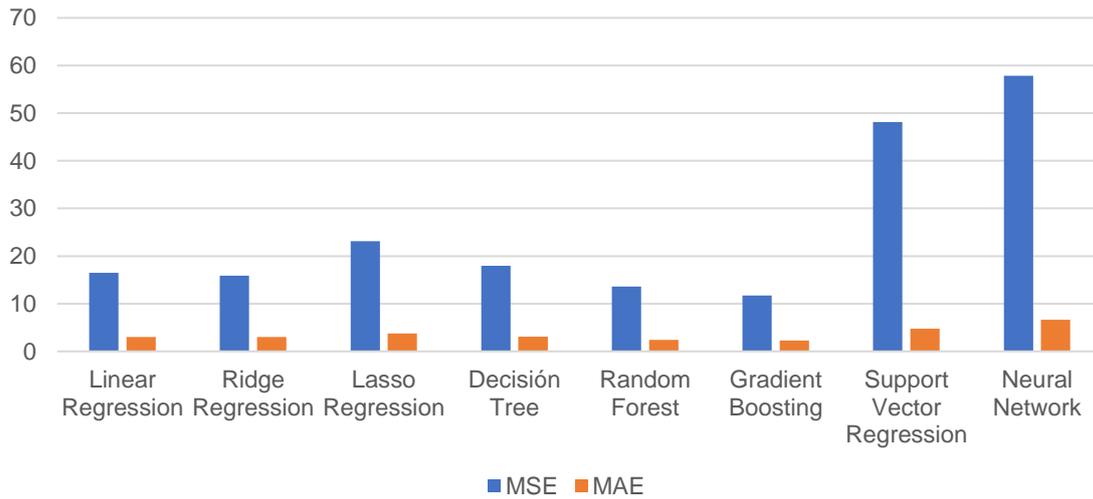


Figura 1. Evaluación comparativa del modelo de regresión obtenida con los datos de los resultados de la Tabla 2.

Discusión

En el contexto actual en el contexto de los negocios, la toma de decisiones basada en datos ha pasado de ser una ventaja competitiva a una necesidad fundamental. La creciente complejidad y la alta competencia en el mercado han incrementado la importancia de herramientas analíticas avanzadas que permitan a las organizaciones no solo interpretar datos, sino también hacer predicciones que puedan guiar la estrategia y las operaciones. Así la tecnología computacional alrededor de minería de datos y big data y el enfoque de machine learning y Deep learning ofrecen soluciones robustas que contribuyen en la calidad en la toma de decisiones gerenciales. En ese sentido los resultados encontrados en el estudio van en concordancia con los resultados de los estudios de Aco *et al.* (2023), Sihombing *et al.* (2023) y Alí *et al.* (2024) en sus estudios utilizaron las métricas cuantitativas de MSE, MAE y R^2 para expresar la precisión evidenciando la capacidad de ajuste que tienen cada uno de ellos. En la evaluación realizada en el estudio, se ha seleccionado como el mejor modelo a Gradient Boosting evaluando las tres métricas de precisión superando considerablemente a otros modelos, sin embargo, es preciso señalar que los resultados que se puedan obtener con cada modelo dependen del caso específico en estudio.

La capacidad de elegir el modelo adecuado de regresión supervisada puede tener un impacto significativo en la calidad de las decisiones gerenciales. La elección debe basarse no solo en la precisión del modelo, sino también en su alineación con los objetivos estratégicos de la empresa, la facilidad con que los resultados pueden ser interpretados y comunicados, y la viabilidad de su implementación en los sistemas existentes. Las recomendaciones derivadas de esta evaluación comparativa pretenden guiar a los gerentes en la selección de modelos que optimicen la toma de decisiones y se alineen con las necesidades y capacidades específicas de su organización.

Conclusión

La integración de modelos de regresión supervisada en los procesos de toma de decisiones gerenciales puede ofrecer beneficios significativos, siempre y cuando se seleccionen y apliquen de manera adecuada. La presente evaluación comparativa proporciona una base para comprender las fortalezas y limitaciones de cada modelo, facilitando una toma de decisiones más informada y efectiva en un entorno empresarial cada vez más cambiante.

Agradecimientos

Mi gratitud a la Universidad Nacional Jorge Basadre Grohmann por concederme todas las facilidades para realizar el trabajo de investigación.

Referencias

Aco, A., Hanco, B., Pérez, Y. (2023). Análisis comparativo de Técnicas de Machine Learning para la predicción de casos de deserción universitaria. *RISTI - Revista Ibérica de Sistemas e Tecnologías de Informação*.
<https://doi.org/10.17013/risti.51.84-98>

Ali, R., Xu, J., Baig, M.H., Rehman, H.S.U., Waqas Aslam, M., Qasim, K.U. (2024). From data to decisions: enhancing financial forecasts with LSTM for AI token prices. *Journal of Economic Studies*, <https://doi.org/10.1108/JES-01-2024-0022>

Gujarati, D., Porter, C. (2010). *Econometría*. Quinta edición. Mc Graw Hill, México DF.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2023). *An Introduction to Statistical Learning with Python*. Springer.

Jarantow, S.W., Pisors, E.D., Chiu, M.L. (2023). Introduction to the use of linear and nonlinear regression analysis in quantitative biological assays. *Current Protocols*, 3, e801. doi: 10.1002/cpz1.801

Kumar, D., Singh, S. (2024). Analyzing the Impact of Machine Learning Algorithms on Risk Management and Fraud Detection in Financial Institution. *International Journal of Research Publication and Reviews*.
<http://dx.doi.org/10.55248/gengpi.5.0524.1135>

Mohd.Ariffin, N. Mohamed.Iqbal, M, Yusoff, M., Mohd, N. (2024). A Study on the Best Classification Method for an Intelligent Phishing Website Detection System. *Journal of Advanced Research in Applied Sciences and Engineering Technology*. <https://doi.org/10.37934/araset.48.2.197210>

Muller, A., Guido, S. (2016). *Introduction to Machine Learning with Python*. O'Reilly Books.

Pal, S. (2024). A Comparative Analysis of Machine Learning Algorithms for Predictive Analytics in Healthcare. *Heritage Research Journal* ISSN No: 0474-9030.

Sihombing, P.R., Budiantono, S., Arsani, A.M., Aritonang, T.M., Kurniawan, M.A. (2023). Comparison of Regression Analysis with Machine Learning Supervised Predictive Model Techniques . *Jurnal Ekonomi Dan Statistik Indonesia*, 3(2): 113-118. <https://doi.org/10.1594/jesi.03.02.03>

Tatachar, A. (2021). *Comparative Assessment of Regression Models Based On Model Evaluation Metrics*. *International Research Journal of Engineering and Technology (IRJET)*.

Valero, J., Navarro, A., Larios, A., Julca, J. (2021). Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción. *Revista de Ciencias Sociales (Ve)*.

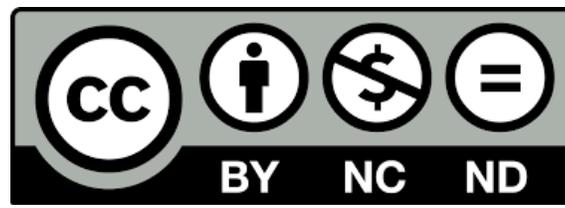
Weinzierl, S., Zilker, S., Dunzer, S., Matzner, M, (2024). Machine learning in business process management: A systematic literature review. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2024.124181>



D. R. © UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO.

Excepto donde se indique lo contrario esta obra está bajo una licencia Creative Commons Atribución No comercial, No derivada, 4.0 Internacional (CC BY NC ND 4.0 INTERNACIONAL). <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>

Fecha de asignación de la licencia 2024-10-28, para un uso diferente consultar al responsable jurídico del repositorio por medio del correo electrónico unidadjuridicafesc@cuautitlan.unam.mx



ENTIDAD EDITORA

Facultad de Estudios Superiores Cuautitlán.

Av. Universidad 3000, Universidad Nacional Autónoma de México, C.U., Delegación Coyoacán, C.P. 04510, Ciudad de México.

FORMA SUGERIDA DE CITAR:

Chambi-Condori, P. P., y Chambi-Vásquez, M. (2024). **EVALUACIÓN COMPARATIVA DE MODELOS DE REGRESIÓN SUPERVISADA EN MACHINE LEARNING PARA LA TOMA DE DECISIONES GERENCIALES**. MEMORIAS DEL CONGRESO NACIONAL DE TECNOLOGÍA (CONATEC), Año 7, No. 7, septiembre 2024 - agosto 2025. Facultad de Estudios Superiores Cuautitlán. UNAM.

https://tecnicosacademicos.cuautitlan.unam.mx/CongresoTA/memorias2024/Mem2024_Paper11.html