

# Modelos y métricas.

MALVADOS Y ASOCIADOS:

Docente : Mg.Ing. HUGO ESPETIA HUAMANGA

Presentado por:

- ANAYA RODRIGUEZ SEBASTIAN RODRIGO
- DEL CASTILLO MOLINA GINO GALOIS
- HUANCA PILARES, JEAN CARLOS
- ZAMATA SOTEC JAMIL
- ZA VALETA HANDA FERNADO HIROSHI

# Presentación

El objetivo principal del presente estudio es implementar un sistema de predicción de precios de viviendas en Cusco utilizando diferentes modelos de inteligencia artificial. Para ello, se cuenta con un dataset que contiene diversas características de las viviendas como área, número de dormitorios, baños, locación, entre otras, que se utilizarán como variables predictoras del precio de las propiedades. La propuesta incluye el desarrollo de varios modelos de IA para identificar el que proporcione los resultados más precisos en función de las características del mercado inmobiliario en Cusco.

# Descripción del Dataset

El dataset contiene las siguientes columnas: Precio, Locación, Área, Dormitorios, Baños, Piso y Zona. Cada una de estas variables será utilizada para entrenar los modelos y predecir el precio de las propiedades.

# Modelos de IA Utilizados

- **Regresión Lineal Múltiple:** Este modelo busca establecer una relación lineal entre las características del inmueble y el precio de venta. Es un enfoque clásico que permite obtener un modelo interpretable.
- **Árboles de Decisión y Random Forest:** Los árboles de decisión y su extensión Random Forest permiten capturar interacciones no lineales entre las variables. Random Forest, en particular, genera múltiples árboles y promedia sus predicciones para mejorar la precisión y reducir el sobreajuste.
- **K-Means Clustering:** Este modelo fue utilizado para agrupar las propiedades según sus características, identificando diferentes segmentos de viviendas en el mercado de Cusco. Este modelo fue utilizado para agrupar las propiedades según sus características, identificando diferentes segmentos de viviendas en el mercado de Cusco.

# Evaluación de los Modelos

Se utilizaron varias métricas de rendimiento para evaluar y comparar los modelos, tales como:

- **$R^2$  (Coeficiente de determinación):** Para evaluar la calidad de la predicción en modelos de regresión.
- **MSE (Error Cuadrático Medio) y RMSE:** Para medir la magnitud de los errores en las predicciones de los precios.

**Método del codo:** En el caso del K-Means para determinar el número óptimo de clusters.

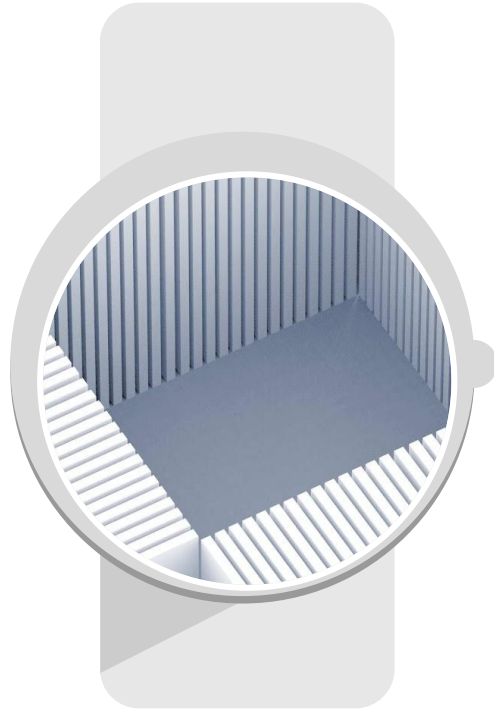
# Propuesta de Modelo no visto en la Asignatura

Un modelo interesante para proponer sería el XGBoost (Extreme Gradient Boosting), un modelo que supera a menudo a otros modelos de regresión y clasificación, y es eficiente tanto en términos de tiempo como de precisión. Las ventajas de este modelo serían:

- Maneja tanto características categóricas como numéricas sin necesidad de transformación extensa.
- Tiene la capacidad de capturar relaciones no lineales y complejas entre las variables.
- Tiende a obtener mejor rendimiento que otros modelos clásicos como la regresión o los árboles de decisión simples.

# Resultados

- El modelo de Regresión Lineal Múltiple proporcionó una predicción razonable, pero fue limitado por su suposición de linealidad.
- Random Forest mostró un rendimiento superior debido a su capacidad para manejar no linealidades en los datos, destacando por su interpretabilidad y robustez.



## Conclusiones

El estudio permitió identificar que los modelos de inteligencia artificial ofrecen una herramienta eficaz para la predicción de precios de viviendas en el mercado inmobiliario de Cusco. El uso de modelos como Random Forest, que capturan relaciones complejas y no lineales, resultó ser más eficiente que los enfoques clásicos de regresión lineal. Los resultados muestran que las características de la propiedad como el área, el número de dormitorios y la locación tienen un fuerte impacto en la predicción de precios.