

UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR



Grado en Ingeniería de Tecnologías y Servicios de la Telecomunicación

TRABAJO FIN DE GRADO

Estudio del estado actual y futuro del Mar Menor usando algoritmos de Machine Learning

Autor: Ángel Allepuz Conesa

Tutor: Gonzalo Martínez Muñoz

junio 2024

Todos los derechos reservados.

Queda prohibida, salvo excepción prevista en la Ley, cualquier forma de reproducción, distribución comunicación pública y transformación de esta obra sin contar con la autorización de los titulares de la propiedad intelectual. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (arts. 270 y sgts. del Código Penal).

DERECHOS RESERVADOS

© Junio 2023 por UNIVERSIDAD AUTÓNOMA DE MADRID

Francisco Tomás y Valiente, nº1

Madrid, 28049

Spain

Ángel Allepuz Conesa

Título: Mar Menor...

Ángel Allepuz Conesa

C\ Francisco Tomás y Valiente N.º 11

IMPRESO EN ESPAÑA – PRINTED IN SPAIN

Agradecimientos

Agradecimientos

Resumen

El Mar Menor, un ecosistema acuático de gran valor, ha enfrentado desafíos significativos debido a la actividad humana y los cambios ambientales, amenazando su biodiversidad y estabilidad ecológica.

Este Trabajo de Fin de Grado **CON TITULO**, presenta el desarrollo de un modelo predictivo basado en técnicas de Machine Learning, diseñado para estimar y predecir variables críticas que influyen en el estado del Mar Menor. Utilizando un conjunto de datos compuesto por parámetros físico-químicos, biológicos y ambientales recogidos a lo largo de varios años.

El objetivo principal del proyecto es desarrollar un algoritmo capaz de predecir con alta precisión variables importantes que, según el Real Decreto 817/2015, indican la calidad de un cuerpo de agua, como son la clorofila α y la cantidad de nitratos y fosfatos.

Los resultados obtenidos intentan demostrar la capacidad de este tipo modelo para ofrecer predicciones precisas que pueden facilitar la toma de decisiones proactivas en la gestión y conservación del Mar Menor.

Este estudio no solo pretende subrayar la importancia de las tecnologías avanzadas en la gestión ambiental, sino que también trata de proponer direcciones futuras para la investigación continua que podrían incluir la integración de modelos más complejos y el análisis de nuevas variables emergentes.

Palabras clave

Mar Menor, Machine Learning, Modelo Predictivo, Conservación Ambiental, Gestión de Ecosistemas.

Abstract

The Mar Menor, an aquatic ecosystem of great value, has faced significant challenges due to human activity and environmental changes, threatening its biodiversity and ecological stability.

This final degree project, titled [Insert Title Here], presents the development of a predictive model based on Machine Learning techniques, designed to estimate and predict critical variables that influence the state of the Mar Menor. It utilizes a dataset comprised of physicochemical, biological, and environmental parameters collected over several years.

The main objective of the project is to develop an algorithm capable of predicting with high accuracy important variables that, according to Real Decreto 817/2015, indicate the quality of a body of water, such as chlorophyll α and the levels of nitrates and phosphates.

The results obtained attempt to demonstrate the capability of this model type to provide accurate predictions that can facilitate proactive decision-making in the management and conservation of the Mar Menor or water bodies alike.

This study not only aims to underscore the importance of advanced technologies in environmental management but also seeks to propose future directions for ongoing research that could include the integration of more complex models and the analysis of new emerging variables.

Keywords

Mar Menor, Machine Learning, Predictive Modeling, Environmental Conservation, Ecosystem Management

Índice

1	Introducción	1
1.1.	Motivación.....	1
1.2.	Objetivos.....	2
1.3.	Organización de la memoria	1
2	Estado del arte	5
2.1.	Subsección	 Error! Marcador no definido.
2.2.	Subsección	 Error! Marcador no definido.
2.3.	Subsección	 Error! Marcador no definido.
2.3.1.	Subsubsección	 Error! Marcador no definido.
	Bibliografía	16
	Apéndices.....	18
	Apéndice A	20

Índice de figuras

X.X Figura 1

Índice de tablas

X.X Tabla 1

Índice de cuadros

X.X Cuadro 1

Introducción

En la era actual, donde el impacto ambiental de las actividades humanas se hace cada vez más palpable, la preservación de nuestros ecosistemas acuáticos emerge como una prioridad crítica. Los cuerpos de agua, especialmente aquellos de pequeña escala como el Mar Menor, son hábitats cruciales para una diversidad biológica significativa. No obstante, estos sistemas están frecuentemente expuestos a presiones antropogénicas que amenazan su equilibrio y la supervivencia de las especies que albergan.

El Mar Menor, un humedal de importancia internacional y el mayor lago de agua salada de Europa, ha sido testigo de dramáticos cambios ecológicos que comprometen su salud y biodiversidad. La eutrofización, el cambio en la salinidad, y la contaminación son solo algunos de los desafíos que enfrenta este valioso ecosistema. En este contexto, la necesidad de herramientas eficaces para la gestión y conservación del Mar Menor es más urgente que nunca.

Este Trabajo de Fin de Grado se centra en la aplicación de técnicas de Machine Learning para desarrollar un modelo predictivo capaz de estimar y predecir variables clave que reflejan el estado actual del Mar Menor. A través de este enfoque, buscamos ofrecer una solución innovadora que permita a los gestores ambientales tomar decisiones informadas y oportunas para la protección y restauración de este cuerpo de agua.

Adoptar un enfoque basado en datos para la gestión ambiental no solo mejora la precisión de las intervenciones, sino que también facilita un monitoreo continuo y adaptativo del estado del ecosistema. Mediante la integración de la ciencia de datos en la ecología, este proyecto aspira a contribuir a la conservación del Mar Menor, garantizando su sostenibilidad y la biodiversidad que sostiene para las generaciones futuras.

1.1. Motivación

Habiendo nacido en Cartagena, mi conexión con el Mar Menor es algo personal. Con el tiempo, experimenté de primera mano su deterioro, viendo cómo el lugar que una vez disfruté se convertía en un ambiente cada vez menos agradable para estar. Esta transformación personal del Mar Menor ha sido un poderoso catalizador en mi decisión de enfocar mi Trabajo de Fin de Grado en este ecosistema.

Este proyecto busca explorar cómo las técnicas de Machine Learning pueden aplicarse para entender mejor el estado de este cuerpo de agua y quizás mitigar algunos de los problemas que enfrenta el Mar Menor. La elección de este tema se inspira en la idea de que, aunque como individuo no puedo revertir los daños, a través de mi formación en ingeniería y ciencias de la computación, puedo contribuir a un mejor entendimiento de su estado actual y el posible estado que tendría si se continúan aplicando ciertas presiones.

Con este Trabajo de Final de Grado, mi intención es aportar modestamente al cuerpo de conocimiento que los gestores ambientales pueden utilizar para tomar decisiones informadas. Aunque soy consciente de que los desafíos son grandes y las soluciones complejas, espero que mi estudio pueda servir como un pequeño paso hacia la comprensión y eventual mejora del Mar Menor.

Y, sobre todo, intentar empezar a trazar el camino del uso de este tipo de tecnologías en temas relacionados con el Mar Menor.

1.2. Objetivos

Los objetivos de este Trabajo Final de Grado son:

O-1.- Plasmar la capacidad de este tipo de modelo de estimar correctamente valores actuales de características importantes en la evaluación de cuerpos de agua. Por lo tanto, construir un modelo fiable a estimar posibles valores futuros de estas características

O-2.- Evaluar el estado actual de Mar Menor, en correspondencia con los umbrales de calidad de los cuerpos de agua, asignados en el BOE.

1.3. Organización de la memoria

ESTADO DEL ARTE

2.1. Mar Menor

El Mar Menor es la laguna salada más grande de Europa, con una superficie aproximada de 135 km^2 , situada en el litoral de la Región de Murcia.

Una parte importante de su estado actual ha dependido y sigue dependiendo de factores naturales como sus dimensiones, características como su alta salinidad y posición geológica, como la falta de lluvia y temperatura ambiente de la Región de Murcia [1].

Sin embargo, en las últimas décadas, el Mar Menor ha enfrentado a unos crecientes factores antropogénicos que han afectado su calidad ambiental y su estabilidad ecológica. Entre estos factores, la entrada y salida de agua en la laguna juegan un papel crucial. El Mar Menor tiene dos principales entradas de agua, cuya regulación está controlada por la Confederación Hidrográfica del Segura (CHS) y la administración autonómica de Murcia. Esta gestión incluye el control de recursos hídricos y la prevención de la contaminación por nutrientes, principalmente nitratos y fosfatos, provenientes de prácticas agrícolas intensivas y otros vertidos.



Figura 2.1: Ilustración de litoral del Mar Menor y las diversas áreas protegidas. Extraído de: [1]

2.2. Eutrofización

Uno de los principales problemas que enfrenta el Mar Menor es la eutrofización, un proceso causado por el exceso de nutrientes, principalmente nitratos y fosfatos. Estos nutrientes promueven el crecimiento excesivo de fitoplancton y algas, lo que conduce a varios efectos adversos en el ecosistema acuático.

El crecimiento extraordinario de fitoplancton y algas provoca un enturbiamiento de las aguas, impidiendo la llegada de la luz a la vegetación del fondo marino. Como consecuencia, esta vegetación muere al no poder realizar la fotosíntesis, lo que a su vez alimenta a bacterias y microorganismos con la materia orgánica muerta. Este proceso incrementa la demanda biológica de oxígeno, reduciendo significativamente los niveles de oxígeno disuelto en el agua, lo que priva de oxígeno a los peces y moluscos autóctonos del Mar Menor [2].

Las causas de la eutrofización pueden ser naturales; sin embargo, en el caso del Mar Menor, existe un consenso sobre la naturaleza antropogénica de los motivos. Las

principales fuentes de estos nutrientes son de carácter humano, provenientes tanto del sector agrícola como del urbano.

Históricamente, se ha sostenido que la principal fuente de nutrientes era la agricultura intensiva en la cuenca del Mar Menor. Sin embargo, un estudio reciente realizado por el experto Brian Lapointe desafía esta creencia. Según los hallazgos de Lapointe, el 50% de estos nutrientes son introducidos por las aguas residuales de origen urbano, mientras que solo un 8% provienen del sector agrícola [4].

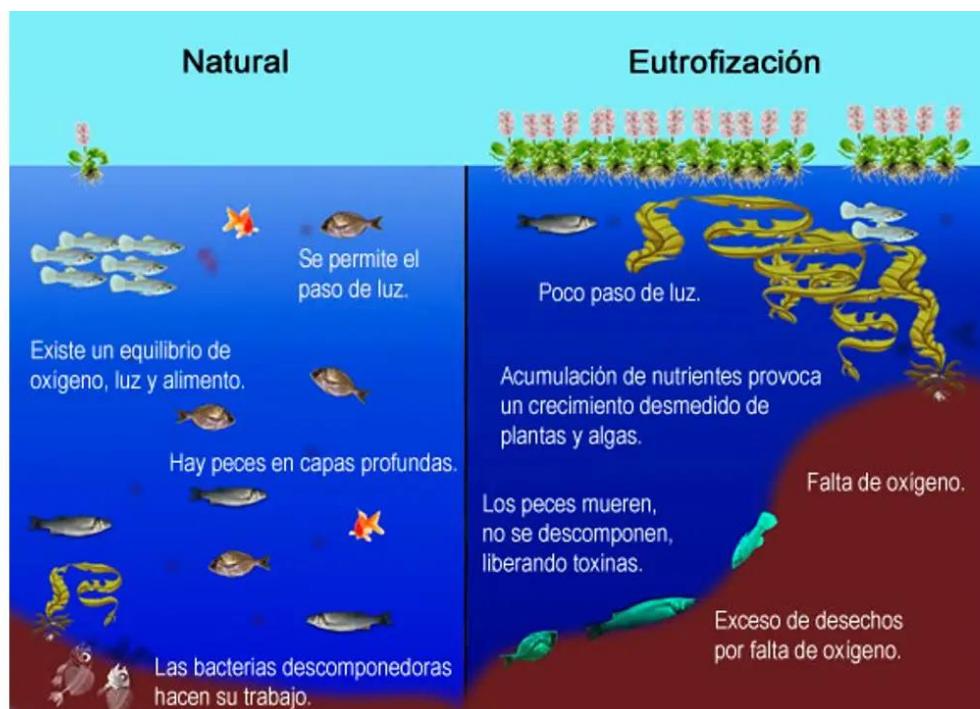


Figura 2.2: Ilustración descriptiva del proceso de eutrofización. Extraído de: [3]

2.3. Legislación sobre los cuerpos de agua

Para estimar la calidad de las aguas del Mar Menor en este trabajo tendré en cuenta los criterios de seguimiento y evaluación del estado de las aguas superficiales de España del Real Decreto 817/2015. Este marco normativo establece valores umbral para diversas variables que indican la calidad del agua.

Este modelo en específico, intentará predecir las variables de Clorofila-a y las concentraciones de Nitratos y Fosfatos (Chl-a, Nitratos y Fosfatos en la tabla respectivamente), ya que son los datos a los que he tenido acceso.

TIPOS AGUAS COSTERAS	INDICADOR	UNIDADES	VALOR ABSOLUTO	Indicadores biológicos e hidromorfológicos: RCE Indicadores químicos y biológicos (ChIA): CONCENTRACIÓN			
				Condición de referencia/ Condición específica del tipo	Límite muy bueno/ bueno	Límite bueno/ moderado	Límite moderado/ deficiente
AC-T11	Chl-a	µg/L	0,9	1,1	1,8		
AC-T11	CARLIT		Valor establecido para cada una de las situaciones ambientales definidas según tipo de costa y sustrato	0,75	0,60	0,40	0,25
AC-T11	BOPA		Fauna únicamente compuesta por especies sensibles (anfípodos excepto género Jassa) y ausencia de poliquetos oportunistas. BOPA: 0	0,95	0,54		
AC-T11	Amonio	µmol NH ₄ /L			4,60 (CP) 2,30 (CM)		
AC-T11	Nitritos	µmol NO ₂ /L			0,92 (CP) 0,46 (CM)		
AC-T11	Nitratos	µmol NO ₃ /L			12,90 (CP) 6,45 (CM)		
AC-T11	Fosfatos	µmol PO ₄ /L			0,76 (CP) 0,38 (CM)		

Figura 2.3: Tabla de los valores umbral de calidad del Mar Menor (AC-T11). Extraído de: [6]

CP se refiere a campo próximo, esto es, mediciones de 0 a 200 metros de la costa y CM a campo medio, a más de 200 metros de la costa. Entendemos que a la laguna del Mar Menor se le aplica los umbrales de CP por su tamaño.

2.3.1. Clorofila-a (Clorofila α)

La clorofila α es un pigmento fotosintético fundamental que se encuentra en todo organismo capaz de hacer la fotosíntesis ya que posibilita la transformación de energía lumínica en energía química [5].

En el contexto de la eutrofización, la clorofila α es un indicador crucial del crecimiento de fitoplancton, lo cual puede desencadenar una serie de eventos negativos, como la eutrofización en los ecosistemas acuáticos [6].

La cantidad de clorofila α en un cuerpo de agua está directamente relacionada con la biomasa del fitoplancton, ya que cada célula de fitoplancton contiene clorofila α. Medir la concentración de este pigmento en el agua proporciona una estimación de la biomasa del fitoplancton presente.

Durante la eutrofización, los niveles elevados de nutrientes llevan a un crecimiento

acelerado del fitoplancton, lo que resulta en un aumento de la concentración de clorofila a. En grandes cantidades, puede tener otras consecuencias ya que reduce la transparencia al incrementar la turbidez, dos variables que también tendrá en cuenta en la base de datos.

Esta disminución de la claridad del agua afecta a las plantas de los fondos al reducirse la cantidad de luz penetrante, requerida para hacer la fotosíntesis.

2.3.2. Nitratos y fosfatos

Los nitratos, junto con los fosfatos, son responsables del aumento de nutrientes en el agua, lo que favorece el crecimiento descontrolado de algas. Este fenómeno reduce la transparencia del agua y afecta la fotosíntesis de las plantas acuáticas.

Aunque los nitratos no interfieren directamente en la adsorción de fosfatos en materiales como dolomita e hidroxiapatita, su presencia es indicativa de una alta carga de nutrientes (Boeykens). Esta alta concentración de nitratos facilita la creación de condiciones de hipoxia, debido al consumo de oxígeno disuelto durante la descomposición de las algas muertas, afectando negativamente a la biodiversidad acuática.

2.4. Aprendizaje automático

El Aprendizaje Automático, también conocido como Machine Learning (ML), es una subdisciplina de la Inteligencia Artificial que se enfoca en el desarrollo de algoritmos y técnicas que permiten a las computadoras aprender y mejorar automáticamente a partir de la experiencia.

En lugar de ser explícitamente programadas para realizar una tarea, las máquinas utilizan datos y ejemplos para identificar patrones y tomar decisiones con mínima intervención humana. Esto se logra mediante la construcción de modelos matemáticos que pueden hacer predicciones o tomar decisiones basadas en datos de entrada.

El Machine Learning utilizado en este trabajo se diferencia de otras categorías de aprendizaje automático como el Deep Learning en la manera de aprender de los datos que

se le proporcionan. El ML no “profundo” es dependiente, en mayor medida, de la intervención humana. Esta intervención viene en forma de su necesidad de trabajar con datos estructurados, es decir, necesita de la determinación inicial del conjunto de características para entender las segregaciones de los datos. (<https://www.ibm.com/es-es>)

es/topics/machine-learning).

Este tipo de modelo se conoce como ML supervisado, definido por su uso de un conjunto de datos etiquetados previamente. En cada iteración, el modelo a su vez, optimiza sus pesos hasta llegar a un punto que considera correctamente ajustado.

2.4.1. Validación cruzada

<https://datascientest.com/es/cross-validation-definicion-e-importancia>

El objetivo de esta técnica es evaluar la capacidad de generalización de un modelo, es decir, su rendimiento en datos no vistos durante el entrenamiento. Esta técnica es fundamental para prevenir el sobreajuste, donde en esencia, el modelo funciona bien en los datos de entrenamiento, pero falla en datos nuevos.

La validación cruzada divide el conjunto de datos en múltiples subconjuntos (folds) y utiliza algunos de ellos para entrenar el modelo y otros para probarlo, repitiendo este proceso varias veces para obtener una evaluación robusta del rendimiento del modelo. Esto permite una estimación más precisa y confiable del desempeño del modelo, comparado con la simple partición de datos en un conjunto de entrenamiento y otro de prueba.

Se pueden encontrar muchos métodos de aplicación de esta técnica. Entre las más utilizadas encontramos, la validación cruzada k-fold en el cual el conjunto de datos se divide aleatoriamente en k grupos de mismo tamaño aproximado. (https://es.wikipedia.org/wiki/Validaci%C3%B3n_cruzada).

El modelo se entrena k veces, cada vez utilizando $k-1$ folds como conjunto de entrenamiento y el fold restante como conjunto de prueba. Este proceso se repite k veces, asegurando que cada fold se utilice exactamente una vez como conjunto de prueba. Los resultados de cada iteración se promedian para obtener una estimación final del rendimiento del modelo.

Sin embargo, para el caso de este caso del Mar Menor, las mediciones son dependientes del tiempo, a ser sensibles a variables meteorológicas, estacionales y eventos específicos, por lo que, el método elegido, es uno cuya separación en subconjuntos no es aleatoria, llamada validación cruzada con series temporales.

2.4.2. Validación cruzada con series temporales

La validación cruzada con series temporales, a diferencia de la validación cruzada tradicional, enfrenta retos únicos debido a la naturaleza secuencial y dependiente del

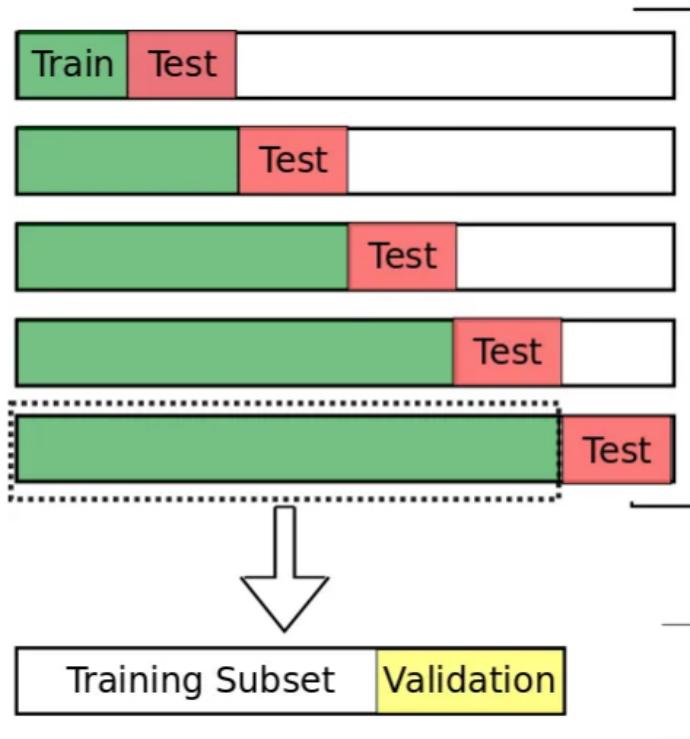
tiempo de los datos. En un entorno de series temporales, es crucial preservar el orden temporal de las observaciones para evitar la introducción de sesgos y garantizar que el modelo pueda generalizar correctamente a datos futuros no vistos. Este proceso implica dividir el conjunto de datos en varios subconjuntos de entrenamiento y prueba.

Una de las técnicas más comunes para realizar la validación cruzada en series temporales es la validación cruzada expansiva ("expanding window cross-validation"). En este método, el tamaño del conjunto de entrenamiento crece con cada iteración mientras que el conjunto de prueba se desplaza hacia adelante. Aquí, en lugar de una ventana deslizante fija, se empieza con un pequeño conjunto de entrenamiento y se va añadiendo más datos de entrenamiento en cada iteración, manteniendo siempre el orden temporal.

<https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9>

https://www.uv.es/lapeva/Thesis/TFM_2019_Hector_Mirete.pdf

<https://forecastegy.com/posts/time-series-cross-validation-python/#sliding-window-validation>



2.5. Random Forest

Se podrían haber elegido muchos algoritmos con el que entrenar este modelo. Sin embargo, por capacidad de entrenamiento y facilidad de instalación se escogió el algoritmo Random Forest.

Utiliza una técnica de aprendizaje automático que usa un enfoque de conjunto

(ensemble learning) para mejorar la precisión de las predicciones. Este método implica la

creación de múltiples árboles de decisión, cada uno entrenado con diferentes subconjuntos de los datos originales, generados mediante bagging (bootstrap aggregating). El bagging crea subconjuntos de datos mediante muestreo con reemplazo, lo que introduce variabilidad y reduce la correlación entre los árboles, mejorando así la robustez y precisión del modelo. Los nodos en un árbol de decisión representan puntos donde se evalúa una característica específica, y las ramas son las rutas que derivan de esos puntos de decisión, llevando a otros nodos o a hojas finales con predicciones.

En contraste, el boosting es otra técnica de ensemble learning que crea modelos secuenciales, cada uno corrigiendo los errores de los anteriores. Mientras que el bagging en Random Forest entrena árboles en paralelo e independientes entre sí, el boosting ajusta cada modelo de manera secuencial, enfocándose en los errores cometidos por los modelos previos. Esta técnica acumulativa permite que el modelo final sea más preciso al adaptarse iterativamente y mejorar continuamente las predicciones. Así, ambos métodos utilizan conjuntos de modelos para mejorar la precisión, pero difieren en cómo manejan la interdependencia y el enfoque en los errores.

<https://link.springer.com/article/10.1023/A:1010933404324>

<https://www.ibm.com/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems.>

2.6. Regresión Lineal

2.7. Métricas

Para evaluar el rendimiento de los modelos predictivos en este Trabajo de Fin de Grado, se han utilizado varias métricas, cada una con sus propias características y aplicaciones.

Mean Squared Error (MSE). Es una métrica que cuantifica la diferencia promedio entre los valores predichos por el modelo y los valores reales observados. Se calcula como el promedio de los cuadrados de los errores, donde los errores son las diferencias entre los valores predichos (\hat{y}) y los valores observados (y_i), n es el número de observaciones:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Mean Absolute Error (MAE), o Error Absoluto Medio, es otra métrica utilizada para evaluar el rendimiento de los modelos predictivos. Se calcula como el promedio de los valores absolutos de los errores:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Coeficiente de Determinación (R^2) es una métrica que indica la proporción de la variación en la variable dependiente que es explicada por el modelo, \bar{y} es el valor medio de las observaciones:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

https://databitai.com/machine-learning/metricas-de-evaluacion-en-machine-learning/#Error_Absoluto_Medio_MAE

| 3

Diseño

Bibliografía

- [1] Fuente1
- [2] Fuente2

Apéndices

A

Apéndice A

Apéndice A

