

Predicción de Intervenciones en Playas Españolas Mediante Árboles de Decisión

Forecasting Interventions on Spanish Beaches through Decision Trees

Íñigo L. Ansorena

*e-mail: ilopezans@gmail.com

Universidad Internacional de La Rioja, España

Keywords: Decision analysis, political/public decision making, Spanish beaches, supervised machine learning, data-driven approach..

Abstract

While large-scale data utilization can be a crucial element in coastal studies, few works have addressed this issue. This paper utilizes a refined database from the beach catalogue (Ministry for Ecological Transition and Demographic Challenge of Spain) to develop a Decision Tree that sheds light on the criteria guiding actions on Spanish beaches. The Decision Tree is constructed through a supervised machine learning technique that learns from 40 features and over 100,000 descriptive data points from the 3,554 Spanish beaches. This work reveals the importance of each variable when making a decision (to act or not) on a specific beach. The model allows for a better understanding of the criteria used by the Ministry to make the decision with statistically significant levels of certainty.

Submitted: July, 2024

Accepted: April, 2025

Associate Editor: Eleonora Verón

Early knowledge of this critical decision can be used by all social, economic, and political agents to make contributions that complement the action proposed by the Coastal Directorate.

Resumen

Aunque el uso de grandes volúmenes de datos puede ser un elemento crucial en los estudios de costas, pocos trabajos han abordado esta cuestión. El presente artículo emplea una base de datos depurada del catálogo de playas (del Ministerio para la Transición Ecológica y Reto Demográfico de España) para elaborar un Árbol de Decisión que arroje luz sobre los criterios que guían las actuaciones en las playas de España. El Árbol de Decisión se elabora a través de una técnica de aprendizaje automático supervisado que aprende de 40 características y más de 100.000 puntos de datos descriptivos de las 3.554 playas españolas. Este trabajo revela la importancia de cada variable a la hora de tomar la decisión de actuar o no sobre una playa determinada. El modelo permite entender mejor los criterios con los que el Ministerio toma la decisión con unos niveles de certeza estadísticamente significativos. El conocimiento con antelación de esa decisión crítica puede ser utilizado por todos los agentes sociales, económicos y políticos para realizar aportaciones que complementen a la actuación planteada por la Dirección General de Costas.

Palabras clave: Análisis de decisiones; toma de decisiones políticas/públicas; playas españolas; aprendizaje automático supervisado; enfoque basado en datos

1. Introducción

A pesar de la amplia base de conocimiento existente en materia de Gestión Integrada de las Zonas Costeras (GIZC), escasean las publicaciones que analicen el valor de los datos y las experiencias previas como herramientas para optimizar las políticas costeras. Algunos autores ya han subrayado que la recopilación de datos pertinentes, precisos y actualizados (espaciales y no espaciales) sobre las zonas marinas y la costa es uno de los retos vitales a los que se enfrentan los investigadores en la actualidad (Tailor *et al.*, 2021). Hoy en día, el proceso de digitalización ha producido enormes cantidades de datos para ayudar en el análisis de decisiones. La cuestión ahora es cómo aprovechar todos esos datos. Fabri (1998) sugirió que, dada la complejidad de los sistemas costeros y la multidisciplinariedad necesaria para un desarrollo costero sostenible, los sistemas informáticos y el denominado Machine Learning (ML) son una buena herramienta para la integración y distribución de esa ingente cantidad de datos y conocimientos especializados.

Siguiendo el principio de que la recopilación de diversos casos y situaciones reales fomenta la inspira-

ción y amplía el conocimiento, se ha desarrollado la presente investigación. Su objetivo es recopilar, analizar y evaluar experiencias previas en España, y hacerlas accesibles mediante un Árbol de Decisiones (AD). El objetivo principal es construir un modelo AD para predecir dónde actuará la Administración española en el futuro. Esta información podría ser utilizada posteriormente, por ejemplo, por la Administración para aumentar la visibilidad y transparencia de su gestión. Además, se plantean dos objetivos secundarios. El primero es evaluar el modelo, es decir, determinar hasta qué punto nuestro modelo es válido para predecir el futuro; y el segundo es identificar las variables que tienen más peso en la predicción, es decir, los atributos de las playas que son más útiles en nuestro modelo.

El resto del artículo se organiza de la siguiente manera. El contexto de estudio y una breve revisión de la literatura, con especial atención a los AD y al ML, se presentan en las secciones 2 y 3 respectivamente. La metodología se explica en la sección 4, que incluye la subsección 4.1, Recogida de datos y marco

general, la subsección 4.2, Construcción del AD con el algoritmo de Árbol de Clasificación y Regresión (CART), y la subsección 4.3. Evaluación del modelo.

En la sección 5 se presentan y discuten los principales resultados, y finalmente la sección 6 se exponen las conclusiones.

2. Contexto de estudio

España cuenta con aproximadamente 8.000 km de costa y una gran variedad de playas repartidas por todo su territorio, lo que la posiciona como el segundo país más visitado del mundo con más de 85 millones de turistas durante el año 2023 (INE, 2024). Dada la relevancia del turismo en las playas españolas, este estudio busca profundizar en las acciones desarrolladas por la Administración española en materia de gestión playas. Se analizará el enfoque histórico adoptado por España en la aplicación de los

principios de la GIZC, establecidos en el año 1992 durante la Cumbre de la Tierra de Río de Janeiro. Cabe destacar que, si bien España es un referente en el ámbito turístico, no ha contado con una política específica de GIZC en el pasado. En este sentido, el presente trabajo no persigue evaluar el cumplimiento estricto de dichos principios por parte de la Administración española, sino más bien analizar las acciones emprendidas en el marco de la gestión costera.

3. Revisión de literatura

En un contexto de creciente preocupación por el cambio climático y su impacto en los sistemas costeros, la adaptación se erige como una tarea fundamental. Como bien lo señaló el Panel Intergubernamental sobre Cambio Climático (IPCC) en su Sexto Informe de Evaluación, las zonas costeras se encuentran entre las áreas más vulnerables a los efectos del cambio climático (véase el capítulo 3). Entre sus principales problemas se encuentran el aumento del nivel del mar, la intensificación de eventos climáticos extremos y la erosión costera (Pörtner *et al.*, 2021). Esta situación exige la implementación de medidas urgentes de adaptación para proteger las comunidades costeras y sus medios de vida.

En el caso de España, la modificación de la Ley de Costas en el año 2013 marcó un hito significativo en el desarrollo de una Estrategia Española de Adaptación Costera al Cambio Climático. Según la Dirección General de Sostenibilidad de la Costa y

del Mar (2016), el instrumento jurídico más importante para abordar el problema del cambio climático en las zonas costeras españolas es la anteriormente mencionada Ley (Ley 2/2013, de 29 de mayo, 2013). Recientemente, Losada *et al.* (2019) han examinado cómo se gestiona la incertidumbre del cambio climático y se superan las barreras en España, dando algunas recomendaciones generales. Aún hoy la gestión costera presenta ciertos vacíos de conocimiento en áreas como la utilización efectiva de la información por parte de los gestores, el impacto de los extensos procesos de consulta en la toma de decisiones y el óptimo aprovechamiento de los datos disponibles.

En este contexto, los AD pueden ser una valiosa herramienta para extraer conocimiento de los datos disponibles. Un AD es una representación de un conjunto de alternativas posibles en forma de red, que muestra los caminos alternativos de que dispone el responsable de la toma de decisiones (Simm &

Cruickshank, 1998). Esta metodología se basa, por tanto, en la representación del modelo en forma de árbol para transmitir una gran cantidad de información en un formato compacto.

Numerosas investigaciones anteriores (den Bieman *et al.*, 2020; Bertrand *et al.*, 2022; Siders y Keenan, 2020; Demetriou *et al.*, 2021), han subrayado algunas de las ventajas de los AD como herramienta de clasificación, ya que los mismos tienen la capacidad de representar cualquier clasificador de valores discretos. Además, los AD son autoexplicativos y exhaustivos; pueden manejar campos de entrada tanto categóricos como numéricos, e incluso conjuntos de datos con valores perdidos o faltantes, además de errores. Gracias a todas estas características y ventajas, los AD pueden introducirse como herramientas de decisión en la gestión costera y emplearse para explorar y comprender los datos del pasado con el fin de predecir el futuro o tomar mejores decisiones. Por ejemplo, Braud y Hill (1988) desarrollaron un conjunto de AD para permitir actividades de construcción en la zona costera. El Center for Coastal Resources Management, Virginia Institute of Marine Science (2010) propuso un AD para la gestión integrada de las costas vírgenes y no urbanizadas. Massey *et al.* (2013) presentaron una guía práctica para la toma de decisiones más sostenibles y el buen manejo de los riesgos costeros, y Zhang y Yang (2020) propusieron un AD para determinar las variables óptimas que mejoran la clasificación del suelo o cubierta terrestre en una zona costera urbanizada compleja.

También hay varios estudios interesantes relacionados con cuestiones costeras más específicas. Por ejemplo, Bae *et al.*, (2010) utilizaron un AD para predecir la concentración bacteriana en una zona costera de California, Reis *et al.* (2012) propusieron un AD para analizar las condiciones de crecimiento de los bivalvos, y Triki *et al.* (2017) y Carretero *et al.* (2019) exploraron el rendimiento de los AD como herramienta para la gestión de acuíferos costeros (ex-

tracción de agua subterránea). Li *et al.* (2022) han demostrado más recientemente que los AD también son aplicables para desagregar e identificar los factores clave que influyen en la adopción de tecnologías de limpieza para la posterior gestión sostenible.

El uso de AD se ha visto mejorado por las técnicas de aprendizaje automático desde que Breiman *et al.* (1984) propusieron el algoritmo CART, que se incluyó en la lista de los 10 algoritmos más influyentes en la comunidad de minería de datos hace una década (Wu & Kumar, 2009). Hoy en día, las aplicaciones de ML ya han llegado a la costa y a los puertos españoles, por ejemplo, se ha aplicado ML para analizar terminales de ferry (Ansorena, 2020) o para valorar proyectos emblemáticos del frente marítimo (Ansorena, 2023). También se ha aplicado ML para analizar la morfodinámica costera y examinar el transporte de sedimentos. Goldstein *et al.* (2019) ha esbozado un conjunto de mejores prácticas para los investigadores que utilizan ML en ese tipo de estudios y sugieren posibles áreas de investigación futura, incluido el uso de nuevas técnicas de aprendizaje automático.

Desde un punto de vista técnico, los gestores del litoral en España deberían mejorar el proceso de toma de decisiones en relación con los problemas costeros. Más aún, en España no existe una política pública de gestión de la zona costera ni se evidencia voluntad de las autoridades para la formulación de una política explícita en este sentido (de Andrés *et al.*, 2020). En este sentido, existen múltiples problemas en el campo del análisis de decisiones que requieren un modelo para predecir o clasificar una variable de salida (objetivo). Por un lado, cuando el objetivo es una variable categórica tenemos un problema de clasificación. En este caso utilizamos el ML para construir un modelo que prediga una categoría (clase) para una nueva entrada. Es decir, ML “clasifica” las salidas en categorías, por ejemplo, “verdadero o falso”, “alto o bajo riesgo”, etc. Por otro lado, cuando el objetivo es una variable numérica tenemos un problema de

regresión. El ML se utiliza aquí para construir un modelo que prediga un valor continuo, por ejemplo, “el número de turistas”, “el ratio de utilización”, etc. Tanto los problemas de clasificación como los de regresión pueden resolverse mediante técnicas supervisadas basadas en el análisis bayesiano. Se denominan supervisadas en el sentido de que los valores de salida los proporciona un experto humano o un proceso de-terminista automatizado, por ejemplo, el algoritmo CART, véase (Breiman & Cutler, 2004).

De esta breve revisión de la bibliografía pueden extraerse dos conclusiones. La primera es que España está dando sus primeros pasos hacia una verdadera adaptación de sus costas al cambio climático. Estos pasos se dieron con la Ley 2/2013 y en el año 2016 con la publicación de la Estrategia de Adaptación al Cambio Climático de la Costa Española. Sin embargo, todavía existe un largo camino por recorrer y algunos autores incluso han denunciado falta de voluntad de las autoridades en este sentido.

4. Metodología

Recogida de datos y marco general

La fuente de datos para este estudio proviene de ‘La Guía de Playas’ publicada por el MITECO (2018), que es un servicio que tiene como objetivo informar a los ciudadanos sobre el patrimonio de las playas españolas utilizando herramientas cartográficas. Se ofrece una descripción completa de cada playa, incluyendo condiciones físicas (como longitud, anchura, ubicación, tipo de arena, composición, existencia de paseo marítimo, etc.), aspectos ambientales (como espacios protegidos, vegetación, etc.), urbanísticos y socioeconómicos (como grado de urbanización, grado de utilización, población del municipio, distancias al hospital y puertos deportivos más cercanos, nudismo, surf, etc.) y servicios (como aseos, lavapiés, duchas,

En segundo lugar, de la revisión de literatura también se deduce un desequilibrio en el número y calidad de las publicaciones que usan inteligencia artificial y ML para tratar de identificar y resolver problemas en la costa. Mientras que a nivel internacional no es difícil encontrar casos de estudios con esta doble temática (ML y costas), en el ámbito nacional es más difícil encontrar este tipo de trabajos. Todavía hay gran cantidad de datos sobre la costa española disponibles y sin explotar suficientemente.

En este contexto, los AD y el ML han despertado el interés de los científicos debido a su potencial para gestionar los problemas costeros. Sin embargo, dada la gran variedad de problemas y el relativamente reciente desarrollo de estas técnicas, pocos estudios han explorado el uso del ML desde la perspectiva de la GIZC, y este artículo está en la línea de empezar a cerrar esta laguna. En resumen, el uso del ML y en particular de AD puede ser una buena estrategia para tomar mejores decisiones, darles más transparencia y planificar las inversiones en el entorno costero.

teléfonos, papeleras, servicios de limpieza, servicios de alquiler, zonas infantiles, oficinas de turismo, etc.) Por último, se añade una variable objetivo, que en esta investigación es una variable binaria que indica si la playa recibió una actuación por parte de la Administración en el pasado. La base de datos se presenta en un formato compacto en el Apéndice I.

El marco completo del estudio se presenta en la figura 1. En primer lugar, se limpia la fuente de datos MITECO (2018) para evitar valores perdidos en el campo (variable) objetivo. A continuación, con el conjunto de datos filtrado se elabora una imagen real (diagrama solar) de las playas españolas utilizando el algoritmo CART. Este algoritmo también permite elaborar un mapa de decisiones que se puede utilizar

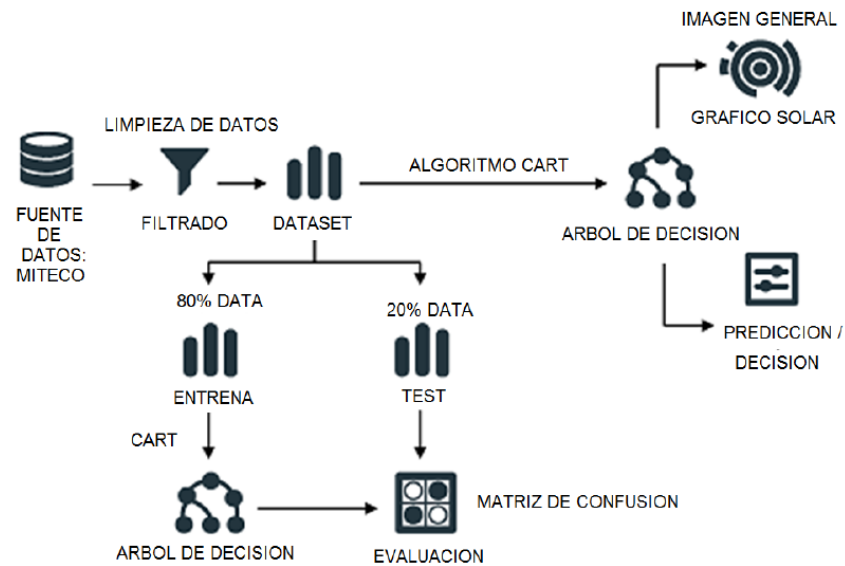


Figura 1. Esquema general de trabajo.

Figure 1. General workflow.

para hacer predicciones o tomar decisiones futuras. Pero ¿es éste un buen modelo AD para tomar decisiones? Es decir ¿se ajusta a la realidad? Para responder a esta pregunta, el conjunto de datos inicial se divide en 2 subconjuntos disjuntos con un 80% y un 20% de datos (Figura 1). Se entrena primero un modelo AD con el primer subconjunto de datos y luego se prueba con información que ese árbol no ha visto antes. Si el árbol de decisión supera la evaluación, el modelo entonces puede utilizarse como herramienta de decisión en el futuro.

Construcción del árbol de decisión (AD)

Como se ha mencionado antes, la base de datos, que describe las principales características de la costa española, se utiliza para construir un AD. El objetivo de este árbol es predecir si una playa necesita una actuación o no. Por lo tanto, el árbol no da ningún detalle sobre la acción en sí (por ejemplo, regenera-

ción de la playa, protección de la costa, mejora de los servicios, etc.). Sólo indica que una playa con características similares recibió atención en el pasado. En este sentido, la figura 2 muestra una visión general de las playas en las que la Administración española ha actuado en el pasado. Suponemos que las actuaciones pasadas se inspiraron en la política española de GIZC y, en consecuencia, las actuaciones futuras seguirán el mismo camino.

El AD elaborado se adapta al conjunto de datos, que es una representación exacta de las playas y de las acciones que se aplicaron en ellas en el pasado. En el AD se distinguen nodos intermedios y finales (hojas) y ramas. Por un lado, los nodos representan los predictores que influyen en la trayectoria predictiva hasta la decisión final (hoja). Cada nodo tiene su propio color que identifica el campo de entrada asociado a él, es decir, el campo que proporciona la mejor división hasta los nodos finales (u hojas) que

representan las decisiones (acción o no acción). Por otro lado, las ramas representan las reglas seguidas para tomar una decisión. La figura 3 presenta dos vistas teóricas del mismo AD, una vista típica de árbol (a) y una vista de sol (b) que también puede verse como una representación de árbol en la que la raíz es el centro del diagrama.

Cada paso (nodo intermedio) está definido por una pregunta (campo) que influye en la predicción.

El algoritmo CART entrena el modelo utilizando un conjunto de datos que se divide recursivamente en dos partes y luego se vuelve a dividir y así sucesivamente. Cada ruta de predicción termina en un nodo de salida (hoja) que se caracteriza por una decisión (clase: sí/no) caracterizada por la certeza del modelo en la predicción (es decir, la confianza y la probabilidad de acertar la clase).



Figura 2. Actuaciones de la Dirección General de Costas en playas de España. No: izquierda, Sí: derecha.

Figure 2. Actions of the Coastal Directorate on Spanish beaches. No: left, Yes: right.

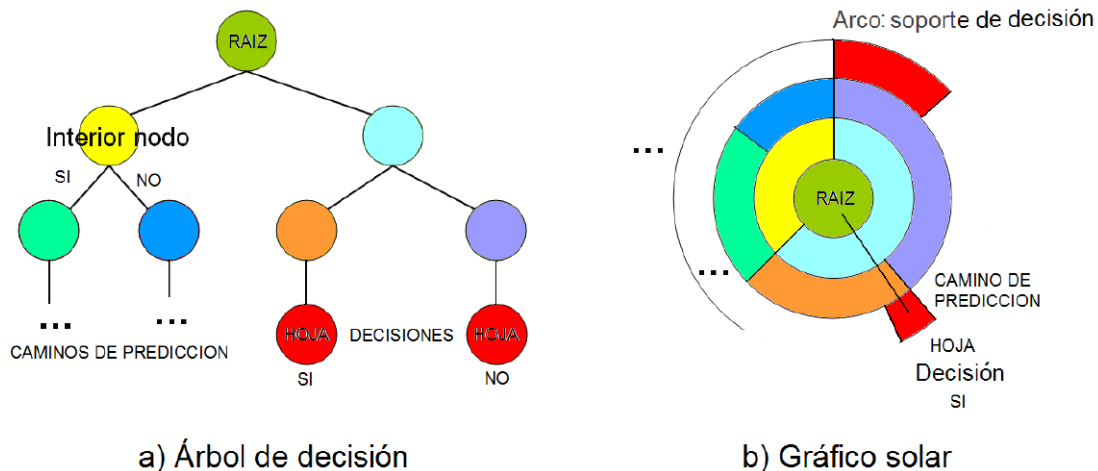


Figura 3. Vista del árbol de decisión (a) y vista en forma de gráfico solar (b).

Figure 3. Decision tree view (a) and sunburst view (b).

Evaluación del modelo

En esta etapa se evalúa el desempeño del modelo AD empleando un conjunto de datos que no ha sido utilizado previamente para entrenar el AD. Para ello, dividimos la base de datos en dos subconjuntos disjuntos: un subconjunto aleatorio que contiene el 80% de la base de datos para crear un nuevo AD, y el 20% restante de la base de datos para probarlo. En la evaluación se compararán las predicciones obtenidas mediante el modelo con los valores reales (subconjunto del 20 %).

Tenemos un problema de clasificación binaria en el que se pretende determinar si una acción es necesaria en una playa concreta (Sí: 1 / No: 0). En este tipo de problemas la clase positiva es la minoritaria y la clase negativa es la mayoritaria, porque a menudo es más interesante predecir los casos raros que los más evidentes y comunes. La clase positiva es la más importante para predecir con exactitud. Se supone que la clase positiva (situación normal) en este estudio es “No (0)”. Por lo tanto:

H_0 : La acción no es necesaria en una playa específica (situación normal)

H_1 : La acción es necesaria en una playa concreta

Como resultado se obtienen los 4 elementos de la matriz de confusión. Estos elementos informan del número de Verdaderos Positivos (*True Positives-TP*), Verdaderos Negativos (*True Negatives-TN*), Falsos Positivos (*False Positives-FP*) y Falsos Negativos (*False Negatives-FN*) (Tabla 1). Todas las predicciones correctas se encuentran en la diagonal de la tabla 1, por lo que es fácil inspeccionar visualmente la misma en busca de errores de predicción (error de tipo I y error de tipo II), ya que estarán representados por valores fuera de la diagonal. Además, se calculan varias métricas descriptivas para ofrecer una imagen completa del modelo. Estas métricas son:

- La precisión (*precision*) es el porcentaje de casos predichos correctamente sobre el total de casos predichos para la clase positiva.

$$\text{PRECISIÓN} = TP/(TP+FP) \quad \text{Ec.1}$$

- La recuperación (*recall*) es el porcentaje de casos clasificados correctamente sobre el total de casos reales de la clase positiva.

$$\text{Recall} = TP/(TP+FN) \quad \text{Ec. 2}$$

- La exactitud (*accuracy*) se calcula como el número de instancias clasificadas correctamente sobre el total de instancias evaluadas.

$$\text{Accuracy} = TP/(TP+FN) \text{ Total instance} \quad \text{Ec. 3}$$

- La medida F (*F-measure*) es la media armónica equilibrada entre *Precision* y *Recall*. La medida F suele ser una métrica más útil que la precisión, ya que un mal rendimiento en *Precision* o *Recall* se traducirá en un valor de medida F bajo.

$$\text{F-measure} = (2 \text{ Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad \text{Ec. 4}$$

- El coeficiente Phi es el coeficiente de correlación entre los valores predichos y los reales. Devuelve un valor entre -1 y 1. Un coeficiente de -1 o negativo implica una correlación negativa entre las predicciones y los valores reales; un 0 indica que la predicción no es mejor que el azar, y un coeficiente de 1 indica una predicción perfecta.

$$\text{Phi Coefficient} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \quad \text{Ec. 5}$$

El último paso es obtener la curva ROC (*Receiver Operating Characteristics*), que es una de las métricas de evaluación más importantes para comprobar los modelos de clasificación (Figura 4). La curva ROC se crea trazando la tasa de verdaderos positivos (también conocida como sensibilidad o recuerdo) frente

a la tasa de falsos positivos (también conocida como probabilidad de falsa alarma o “1 - especificidad”). Por tanto, cada punto de la curva ROC representa un par de sensibilidad/especificidad correspondiente a un umbral de decisión determinado.

El área bajo la curva (AUC) indica la capacidad del modelo para distinguir entre clases. Este parámetro toma valores entre 0,5 y 1; donde 0,5 es un modelo

sin capacidad de discriminación y 1 es un modelo con una capacidad de discriminación perfecta. Como guía para interpretar las curvas ROC se han establecido los siguientes intervalos para los valores AUC: [0,5, 0,6]: Prueba mala, [0,6, 0,75): Prueba regular, [0,75, 0,9): Prueba buena, [0,9, 0,97): Prueba muy buena, [0,97, 1): Prueba excelente.

Tabla 1. Evaluación del modelo. Matriz de confusión. Clase positiva: 0.
Table 1. Model evaluation. Confusion matrix. Positive class: 0.

Pred. \ Act.	0	1	Actual	Recall (d)	F	Phi
0	TP	FN ^(b)	sum	$TP/(TP+FN)$	Eq. 6	Eq. 7
1	FP ^(a)	TN	sum	$TN/(FP+FN)$	Eq. 6	=
Predicted	Sum	Sum	SUM	<i>Avg. Recall*</i>	Avg.	Avg.
Precision ^(c)	$TP/(TP+FP)$	$TN/(TN+FN)$	Avg. P	Accuracy ^(e)		=

Nota: ^(a) Error Tipo I, ^(b) Error Tipo II, ^(c) Eq. 3, ^(d) Eq. 4, ^(e) Eq. 5

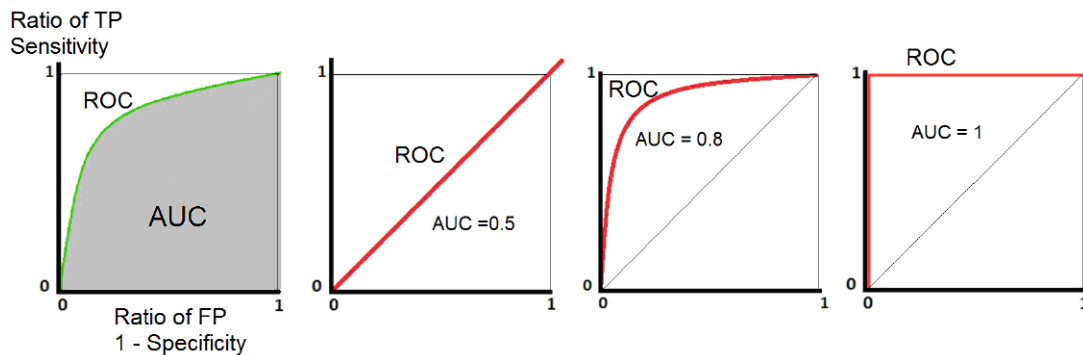


Figura 4. Conceptos de ROC y AUC.
Figure 4. ROC and AUC concepts.

5. Resultados y discusión

El AD resultante se presenta en forma de un diagrama solar (figura 5) que reúne toda la información codificada en el modelo. El diagrama permite visualizar todo el modelo sin filtros en una sola vista. Por un lado, las zonas con más capas (anillos) indican más descendientes, es decir, rutas de predicción más profundas y complejas. Por otro lado, la longitud del arco de cada anillo indica el porcentaje del conjunto de datos cubierto por ese descendiente (datos apoyados), es decir, arcos más pequeños indican menos apoyo para ese descendiente.

El primer gráfico en forma de sol de la izquierda (figura 5a) indica los valores predichos, el segundo (figura 5b) muestra la confianza para cada caso y, por último, el tercero (Figura 5c) presenta los campos incluidos en las rutas de predicción. Aquí cada color indica un campo diferente en el contexto de la predicción.

Ahora se puede utilizar el modelo para explorar lo que ocurrió en el pasado. En otras palabras, se puede

responder a lo siguiente: ¿Es necesario actuar en una playa? En el Apéndice II se incluyen dos ejemplos que dan respuesta a la pregunta. En el primer caso (a), se observa que en el pasado había 72 playas con los mismos atributos y ninguna de ellas recibió una acción. En el segundo caso (b), había 9 playas con los mismos atributos, ocho de ellas no recibieron ninguna acción y sólo una sí. En consecuencia, la respuesta es “No” para ambos escenarios (tabla 2).

Otro aspecto interesante es determinar la contribución relativa de cada campo a las predicciones (es decir, la importancia del campo). Cuanto mayor sea la importancia de campo, mayor será el impacto en las predicciones y, en consecuencia, en las decisiones futuras. La importancia de campo se calcula promediando el error que cada campo contribuye a reducir en cada división del árbol. Como resultado, en la figura 6 se presenta el histograma de la importancia de los campos con mayor peso en el modelo. Aquí se puede ver que los campos más importantes son

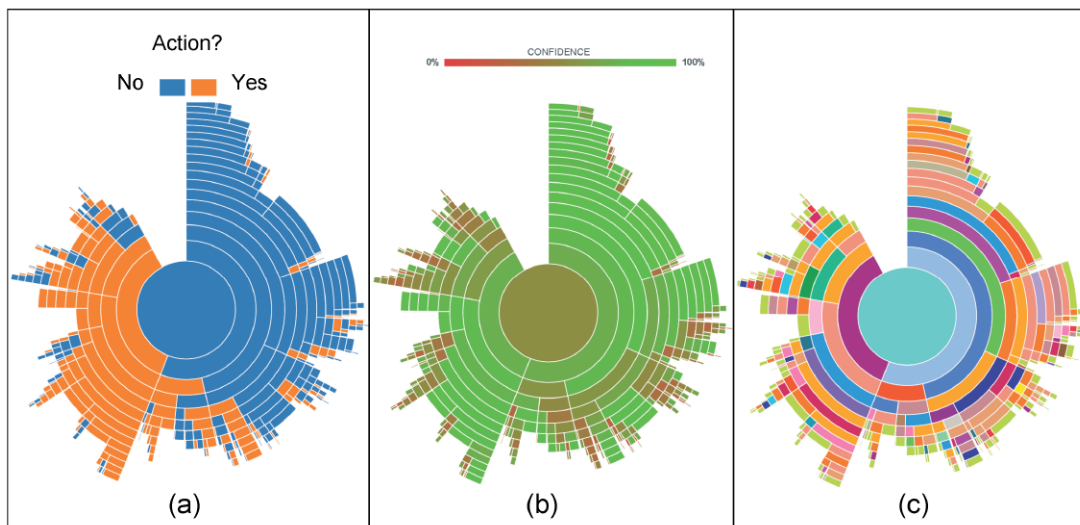


Figura 5. Color por predicción (a), color por confianza (b) y color por campo (c).
Figure 5. Colour by prediction (a), colour by confidence (b) and colour by field (c).

Tabla 2. ¿Es necesario actuar en la playa? (salidas).
Table 2. Is action on the beach necessary? (outputs).

Playa	Decisión	Confianza estadística	Probabilidad	Total playas	Playas con acción	Playas sin acción
a	0 (No)	94.94%	99.35%	72 (3.99%)	0	72
b	0 (No)	56.50%	85.27%	9 (0.50%)	1	8

Campos con más peso en la decisión

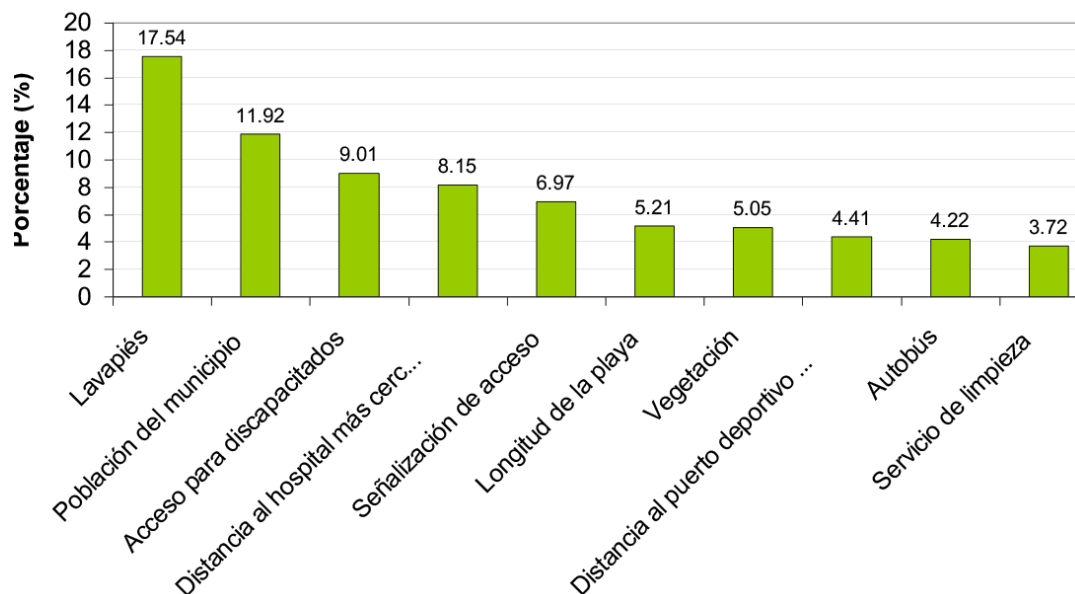


Figura 6. Importancia de las variables (atributos).
Figure 6. Field importance.

la existencia de un “Lavapiés” en la playa y la “Población” (número de habitantes) de su entorno. Teniendo en cuenta las otras variables con mayor peso (Figura 6), se deduce que la decisión de actuar en una playa en España está directamente vinculada a un carácter urbano y de calidad de servicios. Otras variables que indiquen las características del tipo de arena o entorno natural tienen claramente menos peso en

la decisión final de actuar (mejorarla) o no. Cabe señalar que los campos que no aparecen en la figura 6 (es decir, los de menor peso) también se correlacionan con el campo objetivo. Esto sólo significa que el modelo prefiere los campos de la Figura 6 a la hora de elegir divisiones y por tanto realizar la predicción final. La lista completa está en el Apéndice III.

Por último, hay que evaluar el AD. Este es el paso más importante, ya que un modelo que no sea lo suficientemente preciso no puede utilizarse para predecir el futuro. En otras palabras, no puede utilizarse para respaldar decisiones futuras. El estudio parte del supuesto de que la clase positiva es “Sin acción (0)” (situación normal) y, en consecuencia, se obtienen la matriz de confusión y las medidas de clasificación de la tabla 3.

Además, se obtiene el ROC con un valor AUC superior a 0,8 (figura 7). Esto significa que el modelo es capaz de distinguir entre clases, es decir, predice “clase 0” como “clase 0” y “clase 1” como “clase 1”. Un

AUC <0,9 puede considerarse una prueba regular en términos de separación de clases, pero estamos cerca de una buena prueba. En otras palabras, en general cuanto mayor es el AUC mejor es el modelo. Por lo tanto, la brecha puede cerrarse fácilmente refinando el modelo. Una posibilidad es filtrar los valores que faltan en el campo más importante (obsérvese que hay una laguna en el primer anillo). Esto debería mejorar el rendimiento del AD. Dejando a un lado las estrategias para mejorar el modelo, podemos concluir que los resultados de la evaluación ya son satisfactorios.

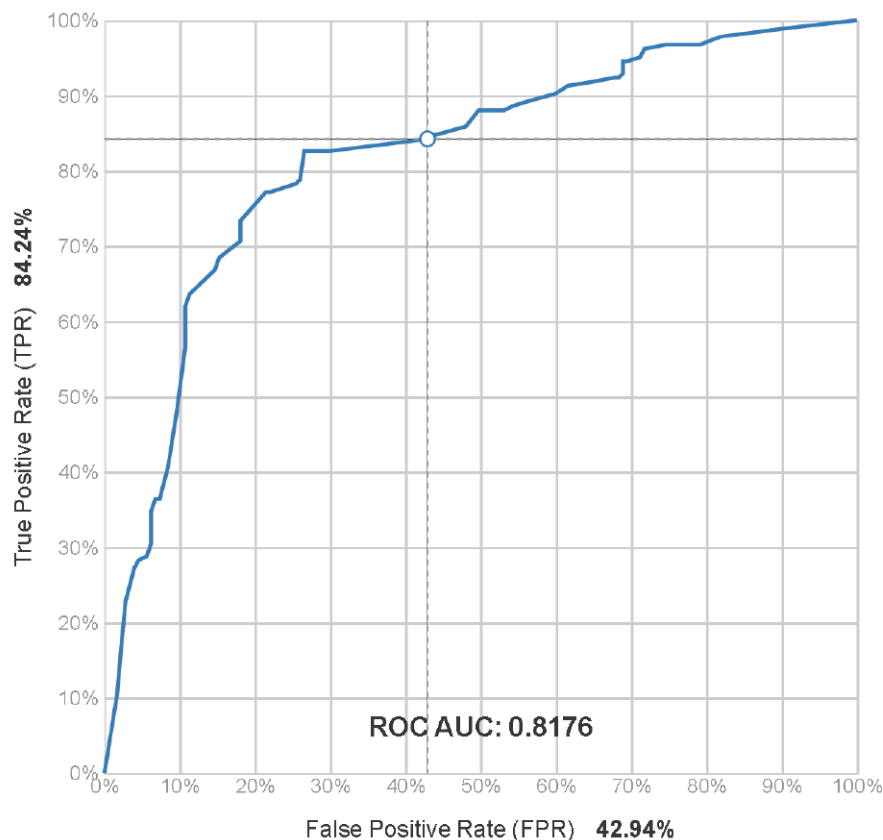


Figura 7. ROC y AUC.
Figure 7. ROC and AUC.

6. Conclusiones

Hoy en día, al menos en España, todavía no se ha explorado a fondo toda la potencia de los datos (muchos en abierto y de uso libre) con herramientas y métodos adecuados. Esto tiene implicaciones para una óptima GIZC. En este estudio se ha demostrado la capacidad del aprendizaje automático para comprender cuando se actúa en una playa en España. En este sentido, se ha entrenado y probado un modelo AD que indica cuando el Ministerio realizó una actuación en una playa (con independencia del tipo de actuación concreta) y ahora contamos con una herramienta que puede proporcionar información crucial para apoyar y facilitar las decisiones de la Administración española y otros agentes en el futuro.

El AD representa un excelente paso inicial hacia la introducción del ML en el estudio de la costa española y en particular de las actuaciones de la Administración sobre las playas. Este es el principal logro

del estudio. Además, el método puede generalizarse a cuestiones más específicas relacionadas con la ingeniería costera, por ejemplo, ¿es necesario regenerar una playa de arena? ¿es necesario reforzar las estructuras de defensa costera?, etc., o incluso a otras preguntas que puedan resultar especialmente apropiadas para comprender el futuro de nuestras costas, el uso de las playas, la sostenibilidad, etc. Se están llevando a cabo nuevas investigaciones sobre estas cuestiones y es muy probable que confirmen que el marco general presentado en este documento funciona con otras preguntas y ámbitos relacionados. No deja de ser curioso e interesante observar que una mejora tan sencilla como un lavapiés en la playa sea un indicador de futuras actuaciones en esa playa. Algo que también podría entenderse como una buena señal para las empresas turísticas interesadas en la zona.

7. Referencias

- Ansorena, I.L. (2020). Managing uncertainty in ferry terminals: a machine learning approach. *International Journal of Business Information Systems*, 33(2): 285-297.
- Ansorena, I.L. (2023). K-means analysis of construction projects in port waterfronts. *International Journal of Applied Decision Sciences*, 16(5): 525-544
- Bae, H.K., Olson, B.H., Hsu, K.L., Sorooshian, S. (2010). Classification and regression tree (CART) analysis for indicator bacterial concentration prediction for a Californian coastal area. *Water science and technology*, 61(2): 545-553.
- Bertrand, G., Petelet-Giraud, E., Cary, L., Hirata, R., Montenegro, S., Paiva, A., Almeida, C. (2022). Delineating groundwater contamination risks in southern coastal metropolises through implementation of geochemical and socio-environmental data in decision-tree and geographical information system. *Water Research*, 209, 117877.
- Braud D.H., Hill J.M. (1988). A decision tree for Coastal Management permit guidelines. *Environmental Management*, 12(1): 29-36.
- Breiman, L., Cutler, A. (2004). *Random Forests*. (Accessed 16 July 2016). http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#giniimp.
- Breiman, L., Friedman, J., Olshen, R. y Stone, C. (1984). *Classification and regression trees*. Wadsworth Int. Group, Belmont, California.
- Carretero, S.C., Capítulo, L.R., Kruse, E.E. (2019). Decision tree as a tool for the management of coastal aquifers of limited saturated thickness. *Quarterly Journal of Engineering Geology and Hydrogeology*, qjgh2018-136.
- Center for Coastal Resources Management, Virginia Institute of Marine Science. (2010). Integrated Shoreline Management Decision Tree for Untreated Shorelines. *Rivers & Coast*, Summer 2010 issue. v.5, no.2. Virginia Institute of Marine Science, College of William

- and Mary. <http://dx.doi.org/doi:10.21220/m2-pc86-8355>
- de Andrés, M., Barragán, J.M., Arenas Granados, P., García Sanabria, J., García Onetti, J. (2020). Gestión de las Zonas Costeras y Marinas en España. *Revista Costas* (1): 117-132. doi:10.26359/costas.e106
- Demetriou, D., Michailides, C., Papanastasiou, G., Onoufriou, T. (2021). Coastal zone significant wave height prediction by supervised machine learning classification algorithms. *Ocean Engineering*, 221, 108592.
- den Bieman, J.P., Wilms, J.M., van den Boogaard, H.F., van Gent, M.R. (2020). Prediction of mean wave overtopping discharge using gradient boosting decision trees. *Water*, 12(6): 1703.
- Dirección general de sostenibilidad de la costa y del mar (2016). Estrategia de adaptación al cambio climático de la costa español. Consultado en Mayo de 2024. https://www.miteco.gob.es/content/dam/miteco/es/costas/temas/proteccion-costa/estrategiaadaptacionccaprobada_tcm30-420088.pdf
- Fabri, K.P. (1998). A methodology for supporting decision making in integrated coastal zone management. *Ocean & Coastal Management*, 39(1-2): 51-62.
- Goldstein, E.B., Coco, G., Plant N.G. (2019). A review of machine learning applications to coastal sediment transport and morphodynamics. *Earth-Science Reviews*, 194, 97-108.
- Hunt, E.B., Marin, J., Stone, P.J. (1966) Experiments in induction. Ac. Press, New York.
- INE (Instituto Nacional de Estadística). (2024). Número de turistas según país de residencia. Movimientos Turísticos en Frontera. <https://www.ine.es/jaxiT3/Datos.htm?t=10822>
- Ley 2/2013 de 29 de mayo de 2013. Ley de protección y uso sostenible del litoral y de modificación de la Ley 22/1988, de 28 de julio, de Costas. Boletín Oficial del Estado, 129, 40691-40736. <https://www.boe.es/eli/es/l/2013/05/29/2>
- Li, X., Yi, S., Cundy, A.B., Chen, W. (2022). Sustainable decision-making for contaminated site risk management: A decision tree model using machine learning algorithms. *Journal of Cleaner Production*, 371: 133612.
- Losada, I.J., Toimil, A., Muñoz, A., García-Fletcher, A.P., Díaz-Simal, P. (2019). A planning strategy for the adaptation of coastal areas to climate change: The Spanish case. *Ocean & Coastal Management*, 182: 104983.
- Massey, J., Lynch, K., de Lucia, A. (2013). Decision Making and Coastal Risks: A Good Practice Guide. Atlantic Network for Coastal Risk Management Project.
- Mehta, M., Agrawal, R., Rissanen, J. (1996). SLIQ: A fast scalable classifier for data mining. in *EDBT 1996: Advances in Database Technology*, Lecture Notes in Computer Science. Springer, Berlin Heidelberg. 18-32 pp.
- MITECO (2018). The Beaches Guide. Guía de playas de España. Documentación asociada a la Guía de playas. Ed.: Ministerio para la Transición Ecológica (MITECO), Madrid, España. <https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/costas-medio-marino/guia-playas-descargas.aspx>
- Pörtner, H.O., Roberts, D.C., Adams, H., Adler, C., Aldunce, P., Ali, E.,, Ibrahim, Z.Z. (2022). Climate change 2022: Impacts, adaptation and vulnerability. IPCC.
- Reis, J.P., Pereira, A., Reis, L.P. (2012). Coastal ecosystems simulation: a decision tree analysis for Bivalve's growth conditions. In *Proceedings 26th European Conference on Modelling and Simulation ECMS May 29th - June 1st, 2012*. Koblenz, Germany.
- Siders, A.R., Keenan, J.M. (2020). Variables shaping coastal adaptation decisions to armor, nourish, and retreat in North Carolina. *Ocean & Coastal Management*, 183: 105023.
- Simm, J., Cruickshank, I. (1998). *Construction risk in coastal engineering*. Thomas Telford. London, UK.
- Tailor, F., Shukla, A., Trumbic, I. (2021). Relevance of Marine Spatial Planning in coastal zone management planning: opportunities and challenges in Indian context—case study of Odisha. *Journal of Earth System Science*, 130(2): 97.
- Triki, C., Zekri, S., Al-Maktoumi, A., Fallahnia, M. (2017). An artificial intelligence approach for the stochastic management of coastal aquifers. *Water resources management*, 31(15): 4925-4939.
- Wu, X., Kumar, V. (2009). The top 10 algorithms in data mining. Ed. Taylor & Francis Group, LLC, London.
- Zhang, F., Yang, X. (2020) Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sensing of Environment*, 251: 112105.