

Clasificación temprana de ciclones tropicales del Atlántico Norte mediante modelos de machine learning y variables atmosféricas basadas en el modelo de Gray

María Fernanda Cordero Cascante¹ and José Eduardo Picado Madrigal²

¹Costa Rica, fer.14cc@gmail.com

²Costa Rica, josepicadom02@gmail.com

20 de febrero de 2026

Resumen

La intensificación de ciclones tropicales en el Atlántico Norte constituye un proceso de alta relevancia para la gestión del riesgo y la predicción operativa. Sin embargo, anticipar desde etapas tempranas si un ciclón alcanzará o no la categoría de huracán sigue siendo un desafío debido a la complejidad de las interacciones oceánicas y atmosféricas que lo gobiernan. Este estudio desarrolla un clasificador binario supervisado para anticipar, con 24 horas de antelación, dicha transición, integrando variables ambientales basadas en el modelo conceptual de Gray con técnicas modernas de aprendizaje automático. Se construyó un dataset histórico que combina trayectorias oficiales de huracanes del HURDAT2 con variables atmosféricas y oceánicas del reanálisis ERA5, incluyendo temperatura superficial del mar, cizalladura vertical del viento, humedad en niveles medios y vorticidad, todas ellas basadas en los parámetros del Modelo de Gray. Adicionalmente, se incorporan índices de oscilaciones climáticas como ENOS, MJO, AMO, NAO y QBO. Se evaluaron múltiples algoritmos lineales y no lineales bajo validación cruzada estratificada, dando prioridad a las métricas centradas en la clase positiva (recall, F2 y PR-AUC) dado el desbalance del problema. Los resultados obtenidos muestran un compromiso entre la reducción de omisiones de eventos y el control de falsas alarmas. El modelo GLMNET obtuvo el mejor equilibrio operativo, destaca la influencia de variables locales. Esta propuesta constituye un herramienta interpretable y adaptable para la vigilancia temprana.

Palabras clave: análisis predictivo, visualización, meteorología, ciclones tropicales

1. Introducción

Los ciclones tropicales (CT) pueden comprenderse como sistemas termodinámicos complejos, su intensidad dependen de la interacción no lineal entre procesos oceánicos y atmosféricos [4]. De acuerdo con Goldenberg et al. en el Atlántico Norte, la actividad ciclónica ha presentando una variabilidad importante en escalas interanuales y multidecadales, la cual está también asociada a condiciones ambientales locales y patrones de gran escala. Sin embargo, a pesar de contar con entornos favorables, lograr anticipar con completa certeza cuáles sistemas tendrán una intensificación puede considerarse una limitante operativa. Algunas investigaciones centralizadas en la intensificación rápida de los CT han evidenciado que, aunque se conoce que bajo ciertas condiciones ambientales las probabilidades de que una tormenta suba de categoría, esta transición no se puede identificar de manera determinista a partir de una sola variable ambiental [17].

Por otro lado, se conoce que los CT son uno de los fenómenos ambientales y meteorológicos que mayor impacto tienen en la región del Atlántico y del Caribe, dejando afectaciones socio-económicas considerables y aumentando la vulnerabilidad de regiones que históricamente se han visto más expuestas a estos eventos. Lograr anticipar los cambios en la intensidad se traduce en una mejor administración del tiempo de reacción y la gestión del riesgo, además de mejorar la planificación de las evacuaciones y brindar información más precisa a la población.

Estimar la intensidad máxima que van a alcanzar los eventos no es el único desafío que se identifica, sino lograr determinar con suficiente tiempo de anticipación si una tormenta tropical evolucionará hasta alcanzar la categoría de huracán. Los enfoques tradicionales han demostrado utilidad operativa, sin embargo, su capacidad para capturar interacciones y relaciones no lineales y efectos combinados entre diversas variables ambientales puede verse limitada.

En paralelo, el crecimiento en el aprendizaje automático ha abierto nuevas oportunidades para modelar, explorar y estudiar sistemas más complejos que se caracterizan por alta dimensionalidad y relaciones no lineales, lo que permite contrastar distintos supuestos de aprendizaje dentro de un mismo marco predictivo. En este contexto, el presente trabajo tiene como objetivo construir un clasificador binario supervisado orientado a anticipar la transición de tormenta tropical a huracán en el Atlántico Norte.

Para lograrlo, se toma la base de datos de huracanes del Atlántico (HURDAT2) con información desde 1851 al 2024. Se integran variables atmosféricas derivadas del reanálisis ERA5, incluyendo temperatura superficial del mar, cizalladura del viento, humedad en niveles medios, que forman parte fundamental del modelo conceptual de Gray. Adicional a esto, se trabaja también con índices de variabilidad climática como el fenómeno el Niño-Oscilación del Sur (ENOS), la Oscilación Madden-Julian (MJO), la Oscilación Multidecadal del Atlántico (AMO), la Oscilación del Atlántico Norte (NAO) y la Oscilación Cuasi-Bienal (QBO), las cuales modulan el entorno ambiental en distintas escalas temporales.

Para evaluar la robustez del enfoque, se implementaron varios algoritmos de clasificación tanto lineales como no lineales, lo que permite analizar si la transición a huracán y evidenciar las interacciones complejas que se dan entre las variables ambientales.

La principal contribución del trabajo está en la integración de un marco físico y conceptual consolidado con técnicas modernas de aprendizaje automático para abordar un problema de clasificación temprana. El estudio propone una aproximación enfocada a la toma de decisiones, priorizando la detección temprana de eventos potencialmente peligrosos que puedan trancisionar de tormenta tropical a huracán.

2. Revisión de la literatura (Estado del arte)

2.1. Ciclones tropicales y procesos de intensificación

Los CTs se originan sobre los océanos tropicales y se alimenta principalmente del calor transferido desde el océano. Se categorizan de acuerdo a su velocidad máxima del viento, la cual se define como la velocidad máxima del viento a una altitud de 10 metros promediada durante 10 minutos. Los CTs generalmente se originan sobre superficies oceánicas que superan los 26°C y siempre se desarrollan sobre a partir de preexistentes perturbaciones, especialmente de ondas con escalas entre los 1000-3000 km. Una vez desarrolladas, las tormentas usualmente se mueven hacia el oeste y ligeramente hacia los polos a velocidades comprendidas entre los 2 y 10 m/s [4].

Inicialmente la intensificación rápida (RI) en los CT se definió aproximadamente como el percentil 95 de los cambios de intensidad en 24 horas en la cuenca del océano Atlántico para eventos que se desarrollaron entre 1989 al 2000, lo cual equivale a un aumento sostenido de la velocidad del viento en superficie de 15.4 m/s en 24 horas [17]. Además se ha demostrado que al menos el 31 % de todos los CT y los huracanes de categoría 4 y 5 sufrieron de RI al menos una vez durante su ciclo de vida, lo que respalda la importancia de determinar la identificar de manera temprana la posibilidad de que un CT escale a un huracán mayor.

2.2. Variables atmosféricas y oceánicas asociadas a la intensificación

La intensificación de los CT se da como resultado de la interacción entre procesos termodinámicos y dinámicos que ocurren en distintas escalas espaciales y temporales. Múltiples investigaciones han identificado un conjunto de variables ambientales que condicionan la probabilidad de desarrollo y el fortalecimiento de estos sistemas. Dentro de las variables más importantes se encuentran la temperatura superficial del mar, la cizalladura vertical del viento, la humedad en niveles medios y la vorticidad en bajos niveles, las cuales a su vez forman parte del marco conceptual propuesto de por Gray para explicar la formación y evolución de estos sistemas.

2.2.1. Energía oceánica y temperatura superficial del mar

La temperatura superficial del mar (SST) es la principal fuente de energía para los CT. Emanuel (1986, 1988) desarrolló una teoría de interacción aire-mar en la que el ciclón puede entenderse como una máquina térmica y su intensidad depende del gradiente energético entre la superficie oceánica y la atmósfera superior. Así valores elevados de SST van a favorecer el aumento del flujo de calor hacia la atmósfera, lo que aumenta la convección profunda y fortalece el sistema.

Se ha considerado como umbral mínimo para el desarrollo de ciclones valores de temperatura de 26-27 ° C. Sin embargo, se ha evidenciado que no solamente la temperatura superficial es importante, sino también la estructura térmica del océano superior. La presencia de anomalías cálidas o capas oceánicas profundas con alto contenido de calor pueden alimentar los procesos de intensificación rápida al reducir el enfriamiento inducido por la mezcla vertical [42].

2.2.2. Cizalladura vertical del viento

La cizalladura vertical del viento se define como la variación de la velocidad y dirección del viento entre niveles bajos y altos de la atmósfera, y es uno de los principales factores con la capacidad de inhibir la intensificación de los ciclones. Los valores elevados de cizalladura permiten desorganizar la estructura vertical. Se ha demostrado que la intensificación de CT es significativamente más probable en ambientes de baja cizalladura, y además, mediante simulaciones numéricas, se ha evidenciado que la presencia de cizalladura moderada a fuerte puede debilitar o incluso disipar los sistemas en desarrollo [2, 8].

Por lo tanto, la cizalladura se identifica como una variable crítica para la predicción temprana, ya que se puede observar desde etapas tempranas en los disturbios atmosféricos, así permite capturar una dimensión dinámica fundamental del entorno.

2.2.3. Humedad en niveles medios

Los ambientes secos favorecen la intrusión de aire seco hacia el núcleo de los CT, lo que incrementa la evaporación y reduce la convección. En niveles medios (600-700 hPa) la humedad relativa tiene un papel importante en el mantenimiento de la convección profunda. Por lo tanto el entorno termodinámico controla en gran medida el potencial de intensificación, además que partiendo del concepto de índice de ventilación es posible integrar las variables de humedad y cizalladura de manera combinada como factores que limitan el desarrollo de los eventos [5, 43].

2.2.4. Vorticidad y condiciones dinámicas de bajo nivel

La vorticidad en niveles bajos de la atmósfera está relacionada con el grado de rotación y se encuentra directamente vinculada con la capacidad del sistema de generar una circulación ciclónica. Esta variable es uno de los factores fundamentales para el desarrollo de un ciclón, así que disturbios con mayor rotación inicial presentan mayor probabilidad de incrementar su intensidad, interactuando con el resto de las condiciones ambientales [13].

2.3. Modelo de Gray como marco conceptual

El modelo de Gray es uno de los marcos teóricos y conceptuales más influyentes en el estudio de la generación e intensificación de ciclones. Este modelo plantea que la formación y evolución de los CT dependen de la integración de seis factores ambientales: (1) temperatura superficial del mar suficientemente elevada, (2) inestabilidad atmosférica, (3) humedad en niveles medios, (4) baja cizalladura del viento, (5) vorticidad ciclónica en niveles bajos y (6) divergencia en niveles altos. La intención conceptual del modelo es identificar el marco ambiental que hace posible el desarrollo de los CT, así su contribución radica en poder sistematizar las variables físicas relevantes y demostrar que la actividad ciclónica se puede explicar mediante la convergencia adecuada de condiciones ambientales favorables [11, 13, 12].

2.4. Influencia de los modos de variabilidad climática

2.4.1. ENOS

El fenómeno el Niño-Oscilación del Sur (ENOS) es uno de los principales moduladores interanuales de los CT en el Atlántico Norte. Es un fenómeno cuasiperiódico que tiene una recurrencia promedio cuatrianual que varía entre dos y siete años, se presenta una perturbación del transporte oceáno-atmósfera de energía y de momento angular desde el ecuador a los polos. Las anomalías en la precipitación tienen mucho que ver en ello y se convierten en un elemento activo de las perturbaciones climáticas [29].

La fase conocida como el El Niño (del ENOS fortalece la circulación de Walker generando un incremento en la cizalladura del viento sobre el Atlántico tropical, inhibiendo la formación e intensificación de ciclones. Por otro lado, la fase de La Niña se presenta en entornos donde la cizalladura es menor y favorece el desarrollo ciclónico [44, 9]. La inclusión del índice ENSO en modelos predictivos puede capturar parte de la variabilidad interanual del entorno.

2.4.2. MJO

La Oscilación Madden-Julian (MJO) representa un modo dominante de variabilidad intraestacional que se caracteriza por la propagación de anomalías convectivas en escalas de 30 a 60 días. Es un modo de variabilidad atmosférica que influye en la ubicación e intensidad de las

precipitaciones tropicales [25]. De acuerdo con Maloney y Hartmann Se ha evidenciado también que la MJO favorece a la formación de CT ya que intensifica la convección profunda y reduce la cizalladura en determinadas regiones del Atlántico.

Debido a que la MJO se desarrolla en escalas temporales que son compatibles con la formación temprana de ciclones, la incorporación de esta oscilación permite tomar en cuenta condiciones ambientales más transitorias.

2.4.3. AMO y NAO

La Oscilación Multidecadal del Atlántico (AMO) se ha identificado como un modo de variabilidad natural que ocurre en el Atlántico Norte en un período estimado de 60-80 años. Se basa en el promedio de las anomalías de la temperatura superficial del mar (SST) en la cuenca del Atlántico Norte, típicamente sobre los 0-80N [22]. Las fases cálidas de la AMO se han asociado a aumentos significativos de CT, así como con fluctuaciones e intensidad de huracanes en el Atlántico Norte, lo que permite

Por otro lado la Oscilación del Atlántico Norte (NAO) tiene influencia en la circulación atmosférica sobre el Atlántico Norte. Esta oscilación también presenta dos fases y están asociadas con cambios en toda la cuenca, así como en la intensidad y la ubicación de la corriente en chorro del Atlántico Norte y la trayectoria de las tormentas, mostrando una influencia local [16].

2.4.4. QBO

La Oscilación Cuasi-Bienal (QBO) es una variación del viento zonal tropical que se propaga hacia la estratosfera inferior, con un período de 28 meses aproximadamente. Su importancia radica en que domina la variabilidad en las capas de la estratosfera inferior, además es relevante para las previsiones estacionales, ya que controla la variabilidad del ozono estratosférico y del agua [21]. Su impacto es menos dominante que el ENSO, sin embargo, puede alterar condiciones climáticas favorables para la convección profunda y por lo tanto para el desarrollo de CT.

2.5. Enfoques de clasificación supervisada

Las técnicas estadísticas para la predicción en la intensidad de eventos ciclónicos tiene antecedentes en modelos estadísticos como el Statistical Hurricane Intensity Prediction Scheme (SHIPS, por sus siglas en inglés), el cual combina variables ambientales mediante regresión lineal simple. Sin embargo, estos enfoques tradicionales asumen relaciones que son principalmente lineales y pueden presentar deficiencias al momento de modelar interacciones más complejas [3].

Para lograr capturar supuestos de aprendizaje y contrastar modelos lineales y no lineales en un entorno meteorológico multivariable, se evaluaron los siguientes algoritmos de clasificación:

- Regresión logística (GLM): Modelo lineal probabilístico base.
- Regresión Logística con penalización Elastic Net (GLMNET): Combina regularización L1 y L2 para manejar colinealidad y selección de variables.
- Naive Bayes Gaussiano: Basado en supuestos de independencia condicional.
- k-Vecinos Más Cercano (KNN): Modelo no paramétrico basado en proximidad.
- Soport Vector Machine (SVM): Maximizan el margen entre clases en espacios potencialmente no lineales.
- Árboles de Decisión: Particionan el espacio de características mediante reglas jerárquicas.
- Bosques aleatorios: Método de ensamble basado en agregación de múltiples árboles.

- Potenciación por Gradiente: Construye modelos secuenciales minimizando errores residuales.
- Redes Neuronales Artificiales: Capaces de aproximar funciones altamente no lineales.

2.5.1. Evaluación del clasificador y matriz de confusión

A partir de la matriz de confusión es posible representar parte del desempeño del modelo, ya que permite realizar comparaciones de las predicciones. Definiendo como clase positiva la transición a huracán y como clase negativa la no transición, se distinguen 4 posibles resultados:

- Verdadero Positivo (TP): el modelo indica que la tormenta se convertirá en huracán y efectivamente ocurre.
- Verdadero Negativo (TN): el modelo indica que la tormenta no se convertirá en huracán y efectivamente no ocurre.
- Falso Positivo (FP): el modelo predice que la tormenta se convertirá en huracán, pero no alcanza dicha categoría.
- Falso Negativo (FN): el modelo predice que la tormenta no se convertirá en huracán, pero si llega alcanzar la categoría.

3. Metodología

3.1. Datos

Como parte de este proyecto se tomaron diversas fuentes para la construcción de la base final. Inicialmente se toma la base de datos de huracanes del Atlántico (HURDAT2 por sus siglas en inglés), que es generada por el Centro Nacional de Huracanes (NHC por sus siglas en inglés); esta base se compone de datos de ciclones desde 1851 al 2024 [18] y con base a la clasificación del huracán se creó la variable "label", que es una columna binaria donde 1 es huracán y 0 no lo es. Además, incluye información de fecha, latitud y longitud del desplazamiento del evento.

A su vez con el reanálisis ERA5 del European Centre for Medium-Range Weather Forecasts (ECMWF), se determinan las variables atmosféricas asociadas a los sistemas anteriores. Este reanálisis combina datos del modelo con observaciones de todo el mundo en un conjunto de datos globalmente completo y consistente utilizando las leyes de la física [15]. Los datos de ERA5 están disponibles de 1940 en adelante. Por último se obtuvieron datos sobre algunas oscilaciones atmosféricas como lo son: El Niño-Oscilación del Sur (ENOS) [24], la Oscilación Madden-Julian (MJO) [27], Oscilación Multidecadal del Atlántico (AMO), Oscilación del Atlántico Norte (NAO) [23] y la Oscilación Cuasi-Bienal (QBO) [26]. La información que se extrae de cada conjunto de datos se encuentra en el cuadro 1, que a su vez es la composición de la base final utilizada en el proyecto.

Variable	Símbolo	Base (fuente)	Tipo de dato
Identificador del registro	id	HURDAT2	Texto (string)
Año	year	HURDAT2	Numérico
Mes	month	HURDAT2	Numérico
Día	day	HURDAT2	Numérico
Hora (UTC)	hour	HURDAT2	Numérico
Latitud	lat	HURDAT2	Numérico
Longitud	lon	HURDAT2	Numérico
Etiqueta/clase objetivo*	label	Derivada (HURDAT2)	Numérico
Viento zonal a 200 hPa*	u200	ERA5	Numérico
Viento meridional a 200 hPa*	v200	ERA5	Numérico
Viento zonal a 850 hPa*	u850	ERA5	Numérico
Viento meridional a 850 hPa*	v850	ERA5	Numérico
Cizalladura vertical 200–850 hPa*	shear_200_850	Derivada (ERA5)	Numérico
Humedad relativa 600–700 hPa*	rh600_700	ERA5	Numérico
Temperatura a 500 hPa*	t500	ERA5	Numérico
Agua precipitable total (TCWV)*	tcwv	ERA5	Numérico
Temperatura superficial del mar (SST)*	sst_c	ERA5	Numérico
Presión a nivel del mar*	mslp_hpa	ERA5	Numérico
Anomalía Niño 3.4 (ENOS)*	nino34_anom	ENOS	Numérico
Índice MJO ROMI PC1*	romi1	MJO	Numérico
Índice MJO ROMI PC2*	romi2	MJO	Numérico
Índice AMO*	amo_ssta	AMO	Numérico
Índice NAO*	nao_index	NAO	Numérico
QBO 30 mb*	qbo_30mb	QBO	Numérico

Cuadro 1: Variables utilizadas, símbolo, fuente de datos y tipo.

Variables con * implica que se utilizan en el modelo para su entrenamiento

La base final se constituye de 563 observaciones, debido al estado de la Oscilación Madden-Julian que fue representada mediante el índice ROMI diario. Dado que este índice está disponible de forma consistente a partir de 1991, el análisis se restringió a dicho periodo con el fin de garantizar homogeneidad observacional y evitar sesgos asociados a la calidad de los datos en décadas previas. A su vez cuenta con 24 variables de las cuales solo 17 se utilizan para el entrenamiento de los modelos. Los datos utilizados en este estudio fueron construidos con el

objetivo de representar, de manera consistente y físicamente interpretable, el entorno ambiental asociado a CTs del Atlántico y su posible intensificación hasta alcanzar la categoría de H. Para ello, se integraron variables atmosféricas locales, condiciones oceánicas y diversos índices climáticos de gran escala. El resumen de las variables se encuentra en el cuadro ???. En general, los vientos presentan alta variabilidad (especialmente a 200 hPa), con una cizalladura media de 7.76 m/s. La humedad relativa en niveles medios se concentra en valores altos (mediana 75.02 %), mientras que la temperatura a 500 hPa es estable alrededor de 267.73 K. En superficie, la SST tiene una mediana de 28.15 °C y la presión a nivel del mar se sitúa cerca de 1011 hPa. Los índices climáticos muestran valores promedio cercanos a cero, con variabilidad moderada, destacando una mayor dispersión en QBO (30 mb).

Variable (unidad)	Promedio	Desviación Estándar	Mínimo	Q25	Q50	Q75	Máximo
u200 (m/s)	0.66	8.03	-19.97	-5.06	-0.74	5.33	35.63
v200 (m/s)	3.35	6.46	-18.71	-0.40	2.37	6.86	30.91
u850 (m/s)	-2.18	4.25	-12.43	-4.93	-2.50	-0.03	15.15
v850 (m/s)	2.54	2.96	-11.84	0.86	2.46	4.34	15.20
shear_200_850 (m/s)	7.76	5.58	0.35	3.92	6.50	9.68	43.42
rh600_700 (%)	73.67	9.58	39.46	67.52	75.02	80.95	92.12
t500 (K)	267.11	2.49	247.84	266.71	267.73	268.38	270.98
tcwv (kg/m ²)	52.22	7.79	13.30	48.50	54.43	57.37	67.36
sst_c (°C)	27.54	2.37	12.38	27.00	28.15	28.93	30.69
mslp_hpa (hPa)	1010.70	4.62	974.83	1008.78	1011.01	1013.18	1024.03
nino34_anom (°C)	-0.08	0.72	-1.63	-0.55	-0.14	0.34	2.72
romi1 (adim.)	0.08	0.88	-2.27	-0.49	0.02	0.69	3.17
romi2 (adim.)	0.98	0.53	0.08	0.58	0.91	1.29	3.17
amo_ssta (°C)	0.54	0.34	-0.38	0.39	0.56	0.74	1.45
nao_index (adim.)	-0.24	0.95	-3.18	-0.93	-0.21	0.39	2.54
qbo_30mb (m/s)	-4.96	14.57	-29.55	-17.71	-2.30	9.53	15.62

Cuadro 2: Estadísticos descriptivos de las variables ambientales utilizadas en el modelo

Adicionalmente, se evaluó la dependencia lineal entre predictores mediante una matriz de correlación (Figura 1). La figura muestra que las variables termodinámicas locales presentan asociaciones moderadas a fuertes; por ejemplo, tcwv, sst_c y t500 exhiben correlaciones positivas elevadas, lo cual sugiere que parte de la información que aportan es compartida. En contraste, los índices climáticos de gran escala (nino34_anom, romi1, romi2, amo_ssta, nao_index, qbo_30mb) muestran correlaciones cercanas a cero con la mayoría de variables locales, lo que sugiere que capturan señales de variabilidad de fondo relativamente independientes. Esta inspección se utilizó para identificar posibles redundancias entre predictores y contextualizar el análisis posterior. Como análisis complementario, se incluyen en el **Material Suplementario** (Figura 6) las curvas de densidad por clase para cada variable. Estas figuras se agregan como referencia adicional para el lector interesado, ya que permiten inspeccionar visualmente cómo cambia la distribución de cada predictor entre “No huracán” y “Huracán”.

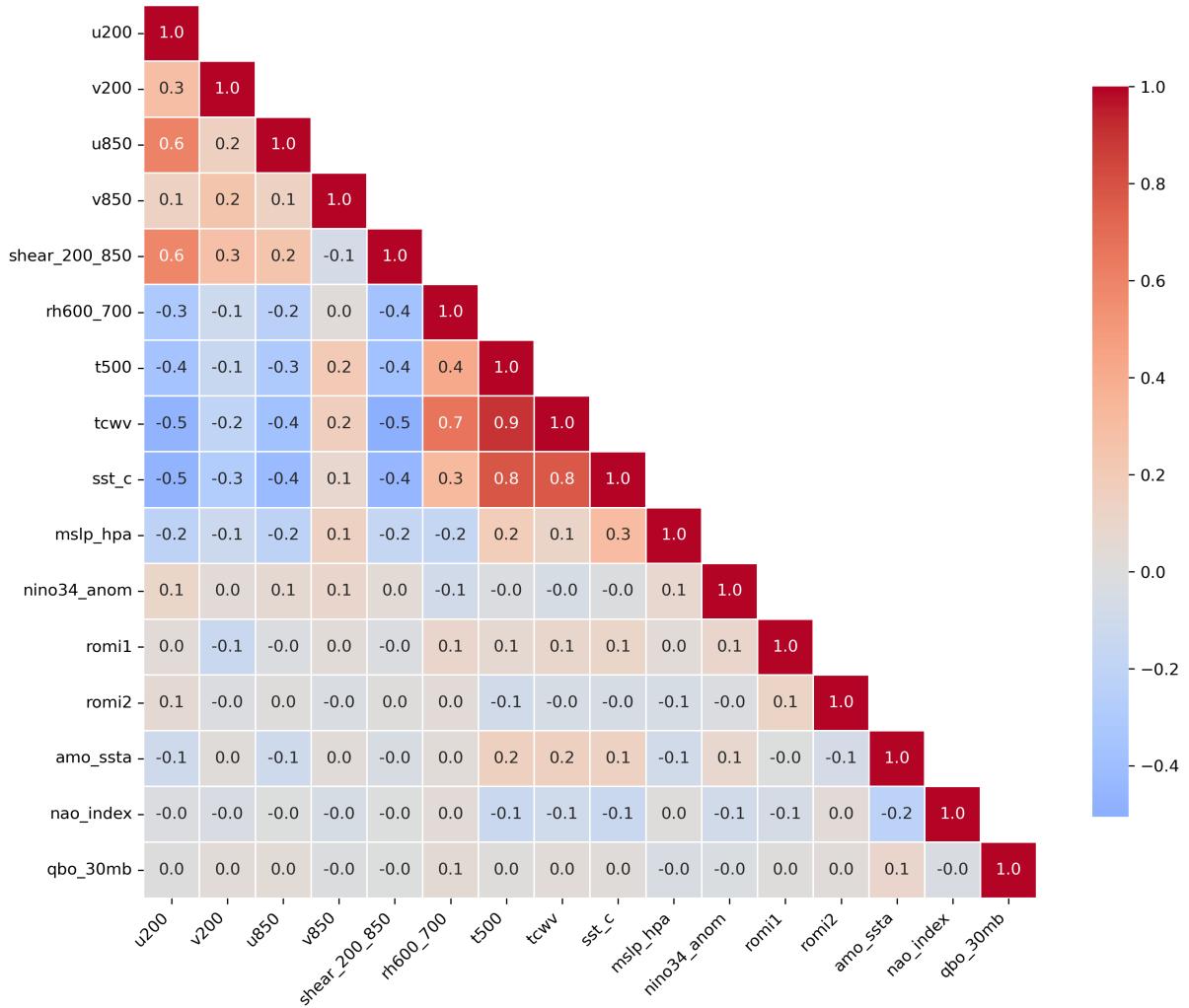


Figura 1: Matriz de correlación entre las variables ambientales. Los tonos rojos indican correlación positiva y los tonos azules correlación negativa, con intensidad proporcional a la magnitud.

Cada observación representa el entorno ambiental asociado a un CT en un instante específico. En el caso de los CTs que alcanzaron categoría de huracán, corresponde a 24 horas antes de la primera clasificación como H, con el fin de capturar condiciones previas a la intensificación. Para los ciclones que no alcanzaron dicha categoría, se seleccionó el instante correspondiente a su máxima intensidad registrada. Con esto se garantiza utilizar solo información previa a la intensificación del sistema.

Por medio de la información de HURDAT2, se restringió el análisis al Atlántico Norte. Para cada CT se almacenaron el identificador del ciclón, la fecha y hora, latitud, longitud y el label que se mencionó anteriormente. Estas variables se utilizaron como llaves para integrar el resto de la información.

Las variables atmosféricas y oceánicas locales fueron extraídas del reanálisis ERA5, utilizando tanto datos de niveles de presión como de niveles simples. Para cada ciclón se descargaron los campos correspondientes a la fecha y hora del instante de referencia. Con el fin de representar el entorno ambiental inmediato del sistema, se calculó un promedio espacial dentro de una grilla de $\pm 3^\circ$ de latitud y longitud centrada en la posición del ciclón. A partir de esto se derivaron variables ambientales clave (cizalladura 200–850 hPa, humedad en niveles medios (rh600_700), SST, MSLP y TCWV) representativas del entorno dinámico y termodinámico asociado a la intensificación ciclónica. La unión a la base se hace por medio del año, mes, día y hora.

Continuando con el objetivo de caracterizar el estado de fondo del sistema climático, se

incorporaron los índices de las oscilaciones atmosféricas, ENOS fue representado mediante la anomalía de temperatura superficial del mar en la región Niño 3.4, utilizando datos mensuales y realizando la unión por año y mes. La MJO fue representada mediante el índice ROMI diario, incorporando sus dos componentes principales, realizando la unión, año, mes y día. Adicionalmente, se incorporaron la AMO, el índice NAO y QBO, esta última representada mediante el viento zonal ecuatorial a 30 hPa. Estos índices fueron integrados al conjunto de datos mediante uniones mensuales por año y mes, al igual que el ENOS, ya que se consideran moduladores de la actividad ciclónica a escalas estacionales e interanuales.

Durante el proceso de integración se identificó un único valor faltante en la variable de temperatura superficial del mar. Este valor fue imputado utilizando la mediana climatológica mensual correspondiente, calculada a partir de todos los eventos disponibles para dicho mes. Esta forma de imputar permite conservar la estacionalidad oceánica y reducir el impacto de valores atípicos, siendo apropiada al tratarse de un valor faltante puntual. Adicionalmente, es importante mencionar que la normalización y escalado de las variables no se realizó durante la construcción del conjunto de datos, sino que se dejó para la etapa de modelado.

3.2. Modelos y métodos

Como se mencionó anteriormente, el objetivo del proyecto fue construir un clasificador binario para anticipar la transición un TS a H en el Atlántico Norte. Desde una perspectiva operativa, se priorizó minimizar los falsos negativos (FN), bajo el supuesto de que no detectar un evento de categoría H tiene un costo mayor que emitir una alerta falsa. En consecuencia, el diseño de evaluación y selección del modelo se centró en métricas orientadas a la clase positiva y en el análisis mediante curvas Precision–Recall (PR), más informativas que ROC en escenarios desbalanceados [31, 1].

Con la base construida se observa un desbalance de clases (Figura 2), por lo que se adoptó una estrategia de evaluación enfocada en: (i) sensibilidad/recall de la clase positiva, (ii) costo operativo de FN y FP, y (iii) medidas resumen sobre la curva PR (PR-AUC / Average Precision).

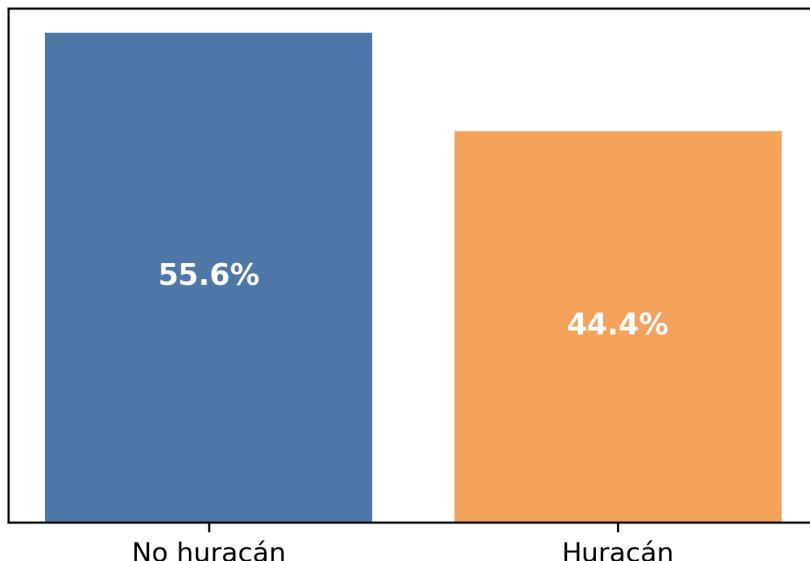


Figura 2: Distribución de la variable `label`.

3.3. Preparación de datos y preprocesamiento

El flujo de trabajo se implementó con un Pipeline, es decir, una secuencia ordenada de pasos que primero prepara los datos y luego entrena el modelo, todo dentro de un mismo objeto. Esto

facilita la reproducibilidad y reduce errores al aplicar exactamente las mismas transformaciones durante el entrenamiento y al momento de predecir nuevos casos [37]. Un aspecto clave fue evitar los sesgos sobre los resultados fue: en cada partición de validación cruzada, las transformaciones se ajustaron solo con los datos de entrenamiento y luego se aplicaron a los datos de validación. De esta forma, el modelo no “ve” información del conjunto de evaluación durante su preparación, lo que mantiene la comparación entre modelos justa y realista.

Las transformaciones utilizadas fueron:

- Estandarización de variables numéricas (`StandardScaler`): se reescalaron las variables para que quedaran en una escala comparable (centradas alrededor de 0 y con dispersión similar). Esto es especialmente útil para modelos sensibles a la magnitud de las variables, como SVM, KNN y redes neuronales [39].
- Codificación de variables categóricas (`OneHotEncoder`) cuando fue necesario: las categorías se transformaron en columnas binarias (0/1) para que el modelo pueda procesarlas correctamente [36].
- Conversión de formato a matriz densa en los casos donde el algoritmo lo requirió, manteniendo compatibilidad entre modelos sin alterar el contenido de la información.

3.4. Modelos evaluados

Se evaluaron varios tipos de modelos de clasificación, con el fin de comparar enfoques con distintos supuestos de aprendizaje y contrastar métodos lineales y no lineales en un entorno meteorológico con múltiples variables.

- Máquinas de Vectores de Soporte (SVM)
- k-Vecinos Más Cercanos (KNN)
- Naive Bayes Gaussiano (BAYES)
- Árbol de Decisión (ÁRBOL)
- Bosque Aleatorio (BOSQUES)
- Potenciación por Gradiente (POTENCIACIÓN)
- Redes Neuronales (RED)
- Regresión Logística (GLM):
 - Regresión Logística con Elastic Net (GLMNET)

3.5. Optimización de hiperparámetros

Para cada modelo se optimizaron hiperparámetros con distintos esquemas de búsqueda: GridSearchCV [34], RandomizedSearchCV [38], una estrategia coarse-to-fine [28], HalvingRandomSearchCV [35] y BayesSearchCV [41]. La validación cruzada utilizada en esta etapa fue estratificada ($k = 5$) para preservar la proporción de clases en cada partición [40].

3.6. Estimación de desempeño y separación entre selección y evaluación

Para obtener una estimación más realista del desempeño, la evaluación se organizó en dos etapas separadas: primero se ajustaron los hiperparámetros de cada modelo y, después, se compararon los modelos usando esos hiperparámetros ya definidos. Esta separación ayuda a evitar resultados demasiado optimistas que pueden ocurrir cuando se usa el mismo proceso de validación cruzada tanto para “buscar la mejor configuración” como para “reportar el desempeño final” [45].

De forma práctica, el procedimiento sigue la lógica de una validación cruzada anidada: una etapa interna para seleccionar la mejor configuración del modelo, y una etapa externa para estimar su rendimiento de manera estable.

- Etapa 1 (selección): se probaron distintas configuraciones y se eligió la que mejor cumplía el objetivo del estudio (priorizando la clase positiva).
- Etapa 2 (evaluación): con la configuración elegida, se midió el desempeño del modelo en varias particiones para obtener un promedio y su variación.

Para cada algoritmo se reportó:

- La matriz de confusión por partición y su promedio ($\overline{TN}, \overline{FP}, \overline{FN}, \overline{TP}$).
- Las métricas de desempeño como media \pm desviación estándar.

3.7. Métricas de evaluación

Sea la matriz de confusión definida por (TP, FP, FN, TN) . Se priorizaron métricas consistentes con el costo operativo de FN.

Recall (sensibilidad) de la clase positiva

$$\text{Recall}_1 = \frac{TP}{TP + FN}. \quad (1)$$

Esta es la métrica crítica del estudio: valores altos implican menos FN.

Precisión de la clase positiva

$$\text{Precision}_1 = \frac{TP}{TP + FP}. \quad (2)$$

Controla la tasa de falsas alarmas.

F_β -score (con $\beta = 2$) Se utilizó F_2 para ponderar el recall por encima de la precision:

$$F_\beta = \frac{(1 + \beta^2) \cdot \text{Precision}_1 \cdot \text{Recall}_1}{\beta^2 \cdot \text{Precision}_1 + \text{Recall}_1}, \quad \beta = 2, \quad (3)$$

lo que penaliza con mayor fuerza escenarios con FN elevados [30].

PR-AUC/Average Precision (AP) Como resumen de la curva PR se empleó Average Precision:

$$\text{AP} = \sum_n (R_n - R_{n-1}) P_n, \quad (4)$$

donde P_n y R_n son precisión y recall evaluados en umbrales sucesivos [33]. Este resumen es especialmente recomendable en problemas desbalanceados [31, 1].

Conteos operativos Además de métricas normalizadas, se reportaron \overline{FN} y \overline{FP} (promedio por fold) como indicadores directos de impacto operativo.

3.8. Criterio de selección del modelo

El criterio de selección fue explícitamente operativo:

1. Minimizar \overline{FN} (prioridad principal).
2. Evitar soluciones con valores extremos de \overline{FP} mediante una restricción práctica: se consideraron únicamente los modelos cuyo \overline{FP} se ubicó en el percentil 75 o por debajo.
3. En caso de empate entre modelos con desempeño operativo similar, se aplicó un criterio de desempate basado en métricas complementarias. En primer lugar, se priorizó $\overline{\text{Recall}}_1$ para asegurar la mayor capacidad de detección de la clase positiva. Si el empate persistía, se utilizó \overline{F}_2 , dado que pondera el *recall* por encima de la *precision* y es coherente con el mayor costo de los falsos negativos. Posteriormente, se consideró $\overline{\text{AP}}$ (PR-AUC) como medida global del balance *precision-recall* a través de distintos umbrales de decisión. Finalmente, se empleó $\overline{\text{Precision}}_1$ para favorecer modelos que, manteniendo alta detección, generan menos falsas alarmas.

Este esquema previene seleccionar modelos con recall artificialmente alto a costa de una explosión de falsas alarmas.

3.9. Entrenamiento final

Una vez elegido el modelo con mejor desempeño, se entrenó una versión final utilizando el 100 % del conjunto de datos. Para mantener coherencia con las pruebas anteriores, se aplicaron los mismos pasos de preparación de datos y la configuración de parámetros que obtuvo los mejores resultados. El modelo final quedó listo para usarse en nuevos casos y generar predicciones en un contexto operativo.

Adicional a esto se reportaron coeficientes estandarizados en el modelo final para entender la contribución de variables. Se analizaron distribuciones por clase (densidades) y estadística descriptiva básica. Finalmente, se evaluó la correlación/multicolinealidad de las variables mediante una matriz de correlación.

En resumen, el proyecto se conforma por: Preparación de datos, búsqueda de hiperparámetros para encontrar la mejor configuración, validación cruzada con hiperparámetros fijos ($k = 5$) y finalmente entrenamiento del modelo elegido por medio de las métricas de evaluación para nuevos individuos.

3.10. Implementación

El desarrollo se realizó en Python, utilizando librerías estándar para manipulación de datos, visualización y aprendizaje automático. En particular:

- **Gestión y análisis de datos:** NumPy y pandas se emplearon para la manipulación, limpieza, filtrado, uniones y preparación del conjunto de datos.
- **Visualización:** matplotlib se utilizó para la generación de figuras y seaborn para gráficos estadísticos complementarios.
- **Preprocesamiento y flujo reproducible:** scikit-learn se utilizó para estructurar un flujo de trabajo reproducible mediante Pipeline y ColumnTransformer, incorporando estandarización (StandardScaler), codificación categórica (OneHotEncoder) y transformaciones auxiliares (FunctionTransformer).

- **Partición y validación:** la separación entrenamiento–prueba se realizó con `train_test_split`, y la validación cruzada estratificada con `StratifiedKFold`, preservando la proporción de clases en cada partición.
- **Modelado:** se evaluaron múltiples clasificadores disponibles en scikit-learn, incluyendo `SVC` (SVM), `KNeighborsClassifier` (KNN), `GaussianNB` (BAYES), `DecisionTreeClassifier` (ÁRBOL), `RandomForestClassifier` (BOSQUES), `GradientBoostingClassifier` (POTENCIA-CIÓN), `MLPClassifier` (REDES) y `LogisticRegression` (GLM/Elastic Net según configuración).
- **Optimización de hiperparámetros:** se implementaron estrategias de búsqueda con `GridSearchCV`, `RandomizedSearchCV` y `HalvingRandomSearchCV`. Adicionalmente, se incorporó `BayesSearchCV` de scikit-optimize, entre otras.
- **Métricas y evaluación:** se calcularon métricas enfocadas en la clase positiva con `precision_score`, `recall_score`, `fbeta_score` (F_2), `average_precision_score` (PR-AUC/AP), además de matrices de confusión (`confusion_matrix`) y curvas Precision–Recall (`precision_recall_curve`).
- **Ejecución y serialización:** `joblib` se utilizó para parallelizar ejecuciones (Parallel, delayed) y para guardar/cargar el modelo final.

3.11. Reproducibilidad

Debido a que las variables meteorológicas utilizadas pueden obtenerse desde distintas fuentes especializadas (y en algunos casos con formatos o resoluciones diferentes), el proceso específico de construcción del conjunto de datos base puede variar y se deja a criterio del lector. No obstante, para facilitar la reproducibilidad, el código que implementa los pasos descritos en las secciones previas se encuentra disponible en el siguiente enlace: https://github.com/JosePicadoM/Proyecto_Final_PROMiDAT.

Para ejecutar el notebook, únicamente es necesario actualizar la ruta del directorio donde se almacenan los archivos de entrada. Los datos empleados en este proyecto se incluyen en el mismo repositorio con el fin de permitir la replicación de los resultados.

Cabe señalar que algunos procedimientos incorporan aleatoriedad (por ejemplo, partición entrenamiento–prueba y procesos de búsqueda de hiperparámetros). Por esta razón, si no se fija una semilla, los resultados pueden presentar variaciones leves entre ejecuciones. Esta decisión busca mantener coherencia con el enfoque de validación cruzada y evitar depender de una única partición específica.

Finalmente, el código fue desarrollado y validado con el conjunto de datos descrito en este trabajo. Su aplicación a otros conjuntos de datos podría requerir ajustes menores, principalmente relacionados con nombres de variables, formatos de entrada y rutas de archivos.

4. Análisis de resultados

A continuación se muestran los principales resultados obtenidos en el proyecto.

Modelo	Recall	Precision	F_2	AP (PR-AUC)	FN	FP
BAYES	0.82 ± 0.03	0.53 ± 0.02	0.74 ± 0.03	0.64 ± 0.03	9.20 ± 1.72	35.80 ± 2.32
GLMNET	0.72 ± 0.06	0.62 ± 0.05	0.70 ± 0.05	0.71 ± 0.07	14.00 ± 2.83	22.00 ± 3.16
GLM	0.72 ± 0.06	0.62 ± 0.05	0.70 ± 0.05	0.71 ± 0.07	14.00 ± 2.83	22.00 ± 3.16
SVM	0.71 ± 0.05	0.62 ± 0.04	0.69 ± 0.04	0.71 ± 0.06	14.60 ± 2.24	21.80 ± 2.48
KNN	0.69 ± 0.04	0.61 ± 0.03	0.67 ± 0.04	0.62 ± 0.06	15.60 ± 2.06	22.40 ± 2.50
BOSQUES	0.68 ± 0.06	0.62 ± 0.05	0.67 ± 0.06	0.71 ± 0.04	15.80 ± 3.19	21.20 ± 3.82
ÁRBOL	0.66 ± 0.07	0.63 ± 0.04	0.66 ± 0.05	0.66 ± 0.04	16.80 ± 3.31	19.80 ± 3.25
RED	0.66 ± 0.06	0.64 ± 0.05	0.65 ± 0.05	0.65 ± 0.03	17.20 ± 2.79	18.80 ± 2.64
POTENCIACIÓN	0.61 ± 0.08	0.70 ± 0.05	0.62 ± 0.08	0.72 ± 0.04	19.60 ± 4.03	13.00 ± 1.41

Cuadro 3: Resultados por modelo (media \pm desviación estándar) en validación cruzada.



Figura 3: Mapa de calor del promedio de falsos negativos (\overline{FN}) y falsos positivos (\overline{FP}) por modelo..

Modelo	Recall	Precision	F_2	AP (PR-AUC)	FN	FP
GLMNET	0.72 ± 0.06	0.62 ± 0.04	0.70 ± 0.05	0.71 ± 0.07	14.0 ± 2.83	22.0 ± 3.16
GLM	0.72 ± 0.06	0.62 ± 0.04	0.70 ± 0.05	0.71 ± 0.07	14.0 ± 2.83	22.0 ± 3.16
SVM	0.71 ± 0.04	0.62 ± 0.04	0.69 ± 0.04	0.71 ± 0.06	14.6 ± 2.24	21.8 ± 2.48
BOSQUES	0.68 ± 0.06	0.62 ± 0.05	0.67 ± 0.05	0.71 ± 0.04	15.8 ± 3.19	21.2 ± 3.82
ÁRBOL	0.66 ± 0.07	0.63 ± 0.04	0.66 ± 0.05	0.66 ± 0.04	16.8 ± 3.31	19.8 ± 3.25
RED	0.66 ± 0.06	0.64 ± 0.05	0.65 ± 0.05	0.65 ± 0.02	17.2 ± 2.79	18.8 ± 2.64
POTENCIACIÓN	0.61 ± 0.08	0.70 ± 0.05	0.62 ± 0.08	0.72 ± 0.03	19.6 ± 4.03	13.0 ± 1.41

Cuadro 4: Resultados de validación cruzada (media \pm desviación estándar) para los modelos con \overline{FP} en el percentil 75 o por debajo.

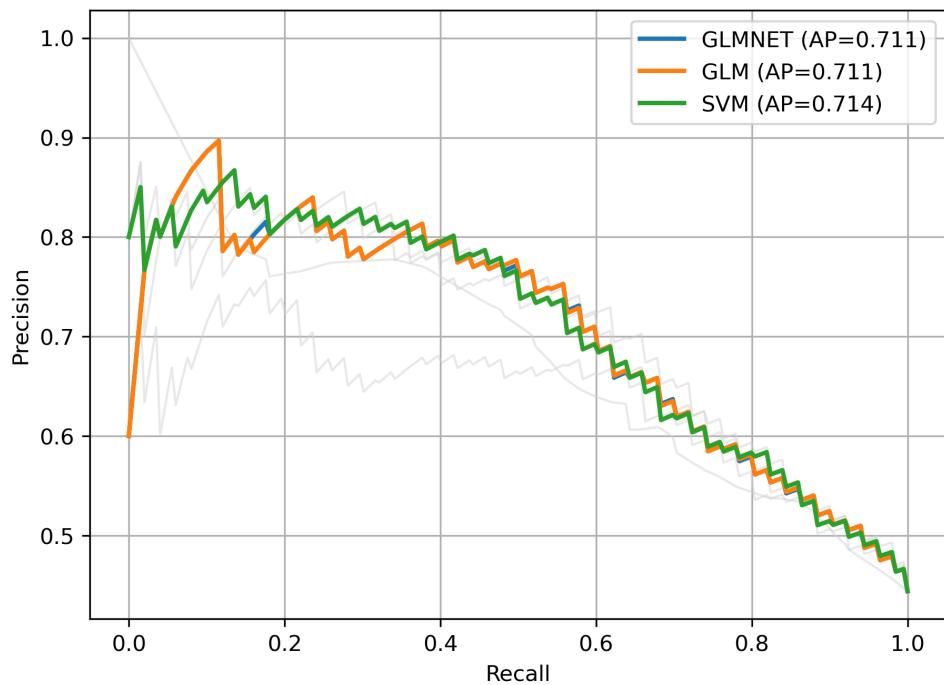


Figura 4: PR CURVE

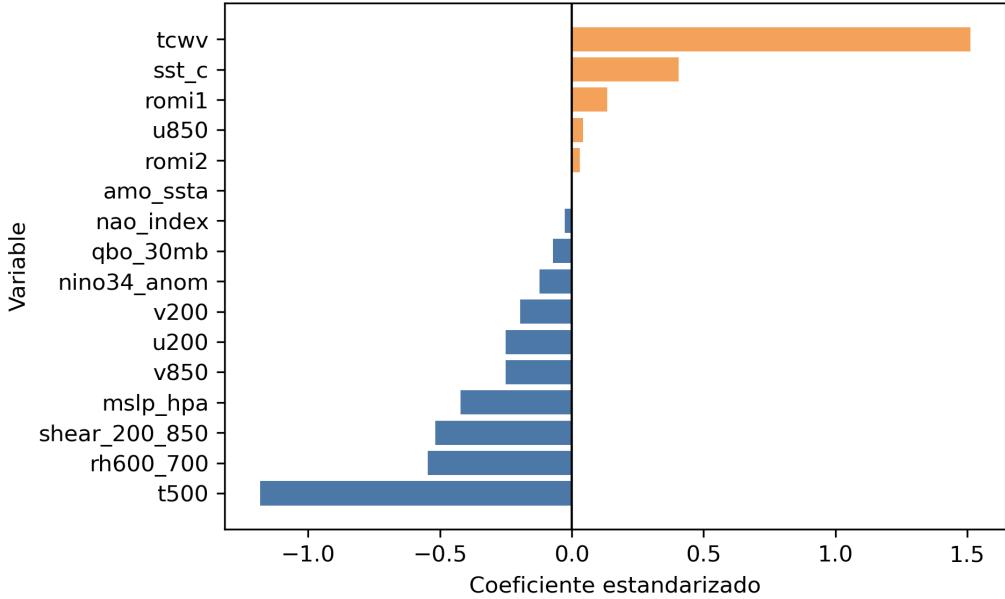


Figura 5: Coeficientes

5. Discusión

En primer lugar, el cuadro 3 evidencia que los modelos difieren de forma importante en el compromiso entre detectar huracanes (reducir falsos negativos, FN) y evitar falsas alarmas (reducir falsos positivos, FP). En particular, BAYES alcanzó el menor número promedio de falsos negativos ($\overline{FN} = 9.2$), lo cual es consistente con el objetivo de minimizar omisiones. No obstante, este beneficio se acompaña de un aumento marcado en falsas alarmas ($\overline{FP} = 35.8$). En un escenario operativo, un volumen elevado de FP puede traducirse en alertas innecesarias, mayor carga de análisis y, eventualmente, pérdida de confianza en el sistema, por lo que el desempeño no debe juzgarse únicamente por sensibilidad/recall [19, 32]. La mejor configuración de que cada modelo por medio de hiperparámetros se encuentra en el material complementario en el cuadro 5.

En la misma línea, la Figura 3 resume visualmente este compromiso al mostrar, para cada modelo, los promedios de \overline{FN} y \overline{FP} en forma de mapa de calor. A partir de esta síntesis se confirma que el mínimo \overline{FN} no coincide necesariamente con un nivel aceptable de \overline{FP} . Por esta razón, y con el fin de mantener una solución utilizable en la práctica, se incorporó la restricción adicional comentada en la metodología: que es considerar únicamente aquellos modelos cuyo \overline{FP} se ubica en el percentil 75 o por debajo. En otras palabras, antes de desempatar por métricas orientadas a la clase positiva, se fija el umbral con el fin de disminuir el impacto operativo que podría implicar un exceso de \overline{FP} como el caso del primer modelo del cuadro 3.

Bajo dicha restricción (Tabla 4), los modelos con mejor equilibrio se concentran en alternativas lineales (GLM, GLMNET y SVM), con valores muy cercanos en recall y en medidas resumen basadas en la curva Precision–Recall, particularmente AP (PR-AUC), la cual es más informativa cuando existe desbalance y el interés se centra en la clase positiva [32]. Esta consistencia también se observa en la Figura 4, donde las curvas Precision–Recall de los tres mejores modelos seleccionados bajo la restricción (líneas de color) se superponen ampliamente y presentan valores de AP similares, mientras que el resto de modelos (en gris) se ubican, en general, por debajo a lo largo de distintos niveles de recall. En particular, las tres curvas mantienen precisiones relativamente altas en rangos bajos a medios de recall, y convergen hacia precisiones menores al aproximarse a recall cercano a 1, lo cual refleja el incremento esperado de falsas alarmas cuando se fuerza la detección prácticamente todos los casos positivos. Es importante recalcar que el modelo con

mayor AP fue el POTENCIACIÓN, este es el que tiene la menor cantidad de falsos positivos, pero con un número mayor de falsos negativos. Si el objetivo fuese minimizar FP, este modelo podría ser de los mejores.

Con base a estos resultados, GLMNET se seleccionó como modelo final porque ofrece un balance favorable entre omisiones y falsas alarmas: mantiene \overline{FN} y \overline{FP} en niveles competitivos dentro del grupo con restricción, sin incurrir en un aumento desproporcionado de falsos positivos. Además, al incorporar regularización elastic net, el modelo tiende a ser más estable cuando existen predictores correlacionados o parcialmente redundantes, una condición frecuente en conjuntos de variables ambientales y meteorológicas [46].

En términos interpretativos, la Figura 5 (Cuadro 6 en el material complementario) presenta los coeficientes estandarizados del modelo GLMNET, lo que permite identificar qué variables separan con mayor fuerza los casos que alcanzan huracán de aquellos que no. En general, el patrón que emerge es coherente con el marco físico descrito por Gray, donde la intensificación depende de un entorno termodinámico favorable (océano cálido y alta disponibilidad de humedad) y de condiciones dinámicas que no inhibían la organización convectiva (por ejemplo, baja cizalladura) [12].

En particular, el mayor peso positivo corresponde a tcwv ($\beta_{std} = 1.51$), indicando que, manteniendo el resto constante, una columna atmosférica más húmeda se asocia con mayor probabilidad de intensificación. De forma consistente, sst_c también aporta positivamente ($\beta_{std} = 0.40$), reforzando la idea de que un océano más cálido incrementa el flujo de energía disponible para sostener convección profunda, elemento central en los mecanismos discutidos por Gray [14, 12].

Por el contrario, la cizalladura vertical shear_200_850 aparece con efecto negativo ($\beta_{std} = -0.52$), lo cual coincide con el rol inhibidor de la cizalladura sobre la estructura vertical y la persistencia del núcleo convectivo [11, 14]. Adicionalmente, t500 muestra el coeficiente negativo de mayor magnitud ($\beta_{std} = -1.18$), sugiriendo que, en este conjunto de datos, la señal asociada a la estructura térmica de niveles medios aporta información relevante para diferenciar entornos más o menos favorables. En el mismo bloque de variables locales, msdp_hpa también contribuye negativamente ($\beta_{std} = -0.42$), consistente con que presiones más bajas suelen acompañar ambientes más cercanos a sistemas organizados.

En cuanto a rh600_700 ($\beta_{std} = -0.55$), su signo puede ser engañoso, al ser un modelo multivariable regularizado, los coeficientes reflejan efectos condicionales (lo que aporta una variable dado que las otras ya están en el modelo). Dado que rh600_700 y tcwv capturan facetas relacionadas de la humedad, es esperable cierta redundancia y en presencia de colinealidad, el modelo puede “repartir” la señal entre predictores de manera no trivial sin que ello invalide la lectura física global [46].

Finalmente, los índices de gran escala (ENOS, MJO, QBO, NAO, AMO) aparecen con magnitudes menores (romi1 = 0.13; nino34_anom = -0.12; qbo_30mb = -0.07; amo_ssta ≈ 0), lo que sugiere que, bajo esta definición de observaciones y a esta escala temporal, la capacidad discriminante proviene principalmente del entorno local inmediato (humedad integrada, condiciones oceánicas y cizalladura), mientras que los índices operan como moduladores secundarios del “estado de fondo”, en congruencia con la interpretación clásica de Gray sobre controles ambientales y modulaciones climáticas [12, 14]. Con ello, GLMNET no solo cumple el criterio operativo de equilibrio entre omisiones y falsas alarmas, sino que también permite una lectura física directa y comparable con la literatura vía coeficientes estandarizados.

Desde una perspectiva operativa, el modelo propuesto podría incorporarse como una herramienta de apoyo a la decisión para priorizar vigilancia y emisión de alertas tempranas ante sistemas con alta probabilidad de intensificación. En particular, al estar optimizado para reducir falsos negativos, su principal valor práctico radica en disminuir la probabilidad de omitir eventos que posteriormente se convierten en huracanes, lo cual es crítico para la planificación de recursos, comunicación de riesgo y preparación anticipada.

Asimismo, el hecho de que el desempeño se evalúe con métricas centradas en la clase po-

sitiva (recall, F_2 y AP/PR-AUC) permite alinear la salida del modelo con necesidades reales de monitoreo, bajo el supuesto que, el costo de no activar protocolos de seguimiento ante un evento relevante suele ser mayor que el costo de analizar una alerta que no se materializa. Con este enfoque, el modelo puede utilizarse como un filtro de priorización, por ejemplo, identificar un subconjunto de casos que necesiten revisión experta inmediata o monitoreo reforzado en las siguientes 24 horas.

Adicionalmente, al tratarse de un modelo lineal, su uso facilita transparencia y trazabilidad: la dirección e intensidad de los coeficientes permiten explicar por qué un caso fue clasificado como de mayor riesgo (altos valores de TCWV y SST, o baja cizalladura), lo que resulta útil para su adopción institucional y para generar confianza en ambientes donde las decisiones deben justificarse. Finalmente, en un despliegue real, el umbral de decisión puede ajustarse según el nivel de tolerancia a falsas alarmas de la entidad usuaria, permitiendo adaptar el modelo a distintos escenarios (temporadas con mayor demanda operativa o recursos limitados), manteniendo el objetivo central de minimizar falsos negativos.

6. Conclusiones

Este trabajo desarrolló y evaluó un clasificador binario para anticipar la transición de CT a H en el Atlántico Norte, bajo un criterio operativo centrado en minimizar omisiones (FN) sin comprometer la utilidad práctica del sistema por un exceso de falsas alarmas (FP). Los resultados confirman que, en un problema desbalanceado, la selección del modelo no puede basarse únicamente en sensibilidad/recall, pues el mejor desempeño en FN puede venir acompañado de un incremento sustancial en FP [19, 32]. En particular, BAYES logró el menor \bar{FN} , pero a costa del mayor \bar{FP} , evidenciando el compromiso inherente entre detección y falsas alarmas (Cuadro 3 y Figura 3).

Para asegurar una solución aplicable en contextos reales, se incorporó una restricción explícita sobre falsas alarmas, reteniendo únicamente los modelos con \bar{FP} en el percentil 75 o por debajo antes de proceder al desempate por métricas enfocadas en la clase positiva. Bajo este filtro, los modelos con mejor rendimiento se concentraron principalmente en alternativas lineales (GLM, GLMNET y SVM), que mostraron desempeños muy cercanos en recall, F_2 y medidas resumen de la curva Precision–Recall (AP/PR-AUC), métrica particularmente informativa ante desbalance [1, 32]. La superposición de las curvas Precision–Recall de los tres mejores modelos seleccionados respalda esta conclusión y refleja el intercambio esperado entre precisión y recall al incrementar la sensibilidad hacia valores cercanos a 1 (Figura 4). Aunque POTENCIACIÓN alcanzó la mayor AP y el menor \bar{FP} , su mayor \bar{FN} lo hace menos adecuado cuando la prioridad es reducir omisiones; no obstante, podría ser preferible si el objetivo operativo se orientara a minimizar falsas alarmas.

Con base en el criterio operativo planteado, GLMNET se seleccionó como modelo final al ofrecer un balance favorable entre \bar{FN} y \bar{FP} dentro del conjunto restringido, sin incurrir en un aumento desproporcionado de falsas alarmas. Adicionalmente, su regularización elastic net aporta estabilidad ante predictores correlacionados o parcialmente redundantes, condición frecuente en variables meteorológicas [46]. Esta elección combina desempeño y robustez, y además facilita interpretabilidad, un atributo deseable para su adopción en entornos institucionales.

El análisis de coeficientes estandarizados del GLMNET mostró que la señal predictiva dominante proviene de variables ambientales locales, especialmente asociadas con la humedad integrada y condiciones oceánicas, así como con factores dinámicos que modulan la organización convectiva. Destaca el aporte positivo de tcwv y sst_c, y el efecto negativo de la cizalladura shear_200_850, consistente con el marco físico clásico de Gray sobre condiciones favorables para intensificación [12, 14, 11]. Por su parte, los índices climáticos de gran escala (ENOS, MJO, QBO, NAO, AMO) presentaron contribuciones relativamente menores, lo que sugiere que, para esta definición de observación y escala temporal, estos índices actúan como moduladores del

estado de fondo más que como separadores principales entre clases.

Desde una perspectiva aplicada, el modelo final puede utilizarse como herramienta de apoyo a la decisión para priorizar vigilancia y alertas tempranas, orientando recursos hacia casos con mayor probabilidad de intensificación en horizontes cortos (24 horas). Su naturaleza lineal facilita transparencia y trazabilidad, permitiendo comunicar de forma clara por qué un caso fue clasificado como de mayor riesgo. A su vez, el umbral de decisión puede ajustarse según la tolerancia operativa a FP, lo que habilita adaptar el sistema a distintos escenarios de monitoreo sin perder el objetivo central de reducir FN.

Como trabajos a futuros, se recomienda validar el enfoque bajo esquemas que respeten la dependencia temporal y ampliar el tamaño muestral para fortalecer la generalización, especialmente considerando la restricción temporal impuesta por la disponibilidad del índice ROMI. Asimismo, resulta pertinente evaluar conjuntos alternativos de predictores, nuevas combinaciones o transformaciones físicas de variables (por ejemplo, gradientes, anomalías estandarizadas o promedios espaciales con distintas ventanas), e incorporar variables adicionales potencialmente informativas. De igual forma, conviene explorar distintos lapsos para definir las características previas a la intensificación, comparando la ventana utilizada en este estudio de 24 h con horizontes alternativos como 12 h, 36 h o 48 h antes de la transición a H, con el fin de cuantificar cómo cambia el poder predictivo y la utilidad operativa del modelo conforme se amplía o reduce el tiempo de anticipación. Finalmente, se puede complementar el modelo con una salida más “accionable” para la operación. En lugar de entregar solo una clase (H/no H), el sistema podría entregar un nivel de riesgo o una recomendación de prioridad (baja, media, alta) y definir desde el inicio en qué punto se activa una alerta según la capacidad real de respuesta del equipo. Esto permitiría adaptar el uso del modelo a diferentes escenarios (temporadas muy activas o turnos con menos personal), manteniendo el objetivo de reducir omisiones, pero evitando generar alertas innecesarias que saturen el monitoreo.

7. Referencias

- [1] Jesse Davis y Mark Goadrich. «The Relationship Between Precision-Recall and ROC Curves». En: *Proceedings of the 23rd International Conference on Machine Learning (ICML)*. 2006.
- [2] Mark DeMaria. «The effect of vertical shear on tropical cyclone intensity change». En: *Journal of Atmospheric Sciences* 53.14 (1996), págs. 2076-2088.
- [3] Mark DeMaria et al. «Further improvements to the statistical hurricane intensity prediction scheme (SHIPS)». En: *Weather and Forecasting* 20.4 (2005), págs. 531-543.
- [4] Kerry Emanuel. «Tropical cyclones». En: *Annual review of earth and planetary sciences* 31.1 (2003), págs. 75-104.
- [5] Kerry Emanuel et al. «Environmental control of tropical cyclone intensity». En: *Journal of the atmospheric sciences* 61.7 (2004), págs. 843-858.
- [6] Kerry A Emanuel. «An air-sea interaction theory for tropical cyclones. Part I: Steady-state maintenance». En: *Journal of Atmospheric Sciences* 43.6 (1986), págs. 585-605.
- [7] Kerry A Emanuel. «The maximum intensity of hurricanes». En: *J. Atmos. Sci* 45.7 (1988), págs. 1143-1155.
- [8] William M Frank y Elizabeth A Ritchie. «Effects of vertical wind shear on the intensity and structure of numerically simulated hurricanes». En: *Monthly weather review* 129.9 (2001), págs. 2249-2269.
- [9] Stanley B Goldenberg y Lloyd J Shapiro. «Physical mechanisms for the association of El Niño and West African rainfall with Atlantic major hurricane activity». En: *Journal of Climate* 9.6 (1996), págs. 1169-1187.
- [10] Stanley B Goldenberg et al. «The recent increase in Atlantic hurricane activity: Causes and implications». En: *Science* 293.5529 (2001), págs. 474-479.
- [11] William M Gray. «Global view of the origin of tropical disturbances and storms». En: *Monthly Weather Review* 96.10 (1968), págs. 669-700.
- [12] William M Gray. «Hurricanes: Their formation, structure and likely role in the tropical circulation. Meteorology over the tropical oceans». En: *Roy. Meteor. Soc.* (1979), págs. 155-218.
- [13] William M Gray et al. «Tropical cyclone genesis». En: (1975).
- [14] William M. Gray. «The Formation of Tropical Cyclones». En: *Meteorology and Atmospheric Physics* (1998).
- [15] Hans Hersbach et al. «The ERA5 global reanalysis». En: *Quarterly Journal of the Royal Meteorological Society* 146.730 (2020), págs. 1999-2049.
- [16] James W Hurrell. «Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation». En: *Science* 269.5224 (1995), págs. 676-679.
- [17] John Kaplan y Mark DeMaria. «Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin». En: *Weather and forecasting* 18.6 (2003), págs. 1093-1108.
- [18] Christopher W. Landsea. *The Revised Atlantic Hurricane Database (HURDAT2)*. Inf. téc. HURDAT2 format documentation for the Atlantic basin (1851–2021). NOAA National Hurricane Center, abr. de 2022. URL: <https://www.nhc.noaa.gov/data/hurdat/hurdat2-format-atl-1851-2021.pdf>.
- [19] Charles X. Ling y Victor S. Sheng. *Cost-Sensitive Learning and the Class Imbalance Problem*. Inf. téc. Technical report / working paper. The University of Western Ontario, 2004. URL: https://www.csd.uwo.ca/~xling/papers/cost_sensitive.pdf.

- [20] Eric D Maloney y Dennis L Hartmann. «Modulation of eastern North Pacific hurricanes by the Madden–Julian oscillation». En: *Journal of climate* 13.9 (2000), págs. 1451-1460.
- [21] NASA Goddard Space Flight Center. *Quasi-Biennial Oscillation (QBO)*. Atmospheric Chemistry and Dynamics Branch. n.d. URL: https://acd-ext.gsfc.nasa.gov/Data_services/met/qbo/qbo.html (visitado 12-02-2026).
- [22] National Center for Atmospheric Research. *Atlantic Multi-decadal Oscillation (AMO)*. Climate Data Guide, UCAR. 2026. URL: <https://climatedataguide.ucar.edu/climate-data/atlantic-multi-decadal-oscillation-amo> (visitado 12-02-2026).
- [23] NOAA Climate Prediction Center. *North Atlantic Oscillation (NAO) Index: Monthly Normalized Values (b5001)*. Monthly mean NAO index time series (normalized). Accessed from CPC Teleconnections (NAO) data downloads. NOAA/NWS/NCEP Climate Prediction Center (CPC), 2026. URL: <https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/norm.nao.monthly.b5001.current.ascii> (visitado 10-02-2026).
- [24] NOAA Climate Prediction Center. *Sea Surface Temperature (SST) Indices (Niño 1+2, Niño 3, Niño 3.4, Niño 4) from OI SST (sstoi.indices)*. Monthly time series of SST indices used to monitor ENSO; distributed via CPC Indices directory (see README). NOAA/NWS/NCEP Climate Prediction Center, 2026. URL: <https://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices> (visitado 10-02-2026).
- [25] NOAA Physical Sciences Laboratory. *Madden-Julian Oscillation (MJO)*. National Oceanic and Atmospheric Administration. 2024. URL: <https://psl.noaa.gov/mjo/> (visitado 12-02-2026).
- [26] NOAA Physical Sciences Laboratory. *Quasi-biennial Oscillation (QBO) Index, monthly time series*. Monthly QBO index time series (standard PSL format). Dataset landing page: <https://psl.noaa.gov/data/timeseries/month/DS/QBO/>. NOAA Physical Sciences Laboratory (PSL), 2026. URL: <https://psl.noaa.gov/data/correlation/qbo.data> (visitado 10-02-2026).
- [27] NOAA Physical Sciences Laboratory. *ROMI: Real-time OLR MJO Index (ROMI values time series)*. Daily ROMI principal components and amplitude time series (ASCII). NOAA Physical Sciences Laboratory (PSL), 2026. URL: <https://psl.noaa.gov/mjo/mjoindex/romi.cpcolr.1x.txt> (visitado 10-02-2026).
- [28] Louis Owen. *The Art of Hyperparameter Tuning in Python: Grid, Random, Coarse to Fine, Bayesian, Manual Search, and Genetic Algorithm*. Medium (Data Science). Sep. de 2021. URL: <https://medium.com/data-science/the-art-of-hyperparameter-tuning-in-python-c581a129d4e4> (visitado 11-02-2026).
- [29] Germán Poveda y Óscar J Mesa. «Las fases extremas del fenómeno ENSO (El Niño y La Niña) y su influencia sobre la hidrología de Colombia». En: *Tecnología y ciencias del agua* 11.1 (1996), págs. 21-37.
- [30] C. J. van Rijsbergen. *Information Retrieval*. Butterworths, 1979.
- [31] Takaya Saito y Marc Rehmsmeier. «The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets». En: *PLOS ONE* 10.3 (2015), e0118432. DOI: [10.1371/journal.pone.0118432](https://doi.org/10.1371/journal.pone.0118432).
- [32] Takaya Saito y Marc Rehmsmeier. «The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets». En: *PLOS ONE* 10.3 (2015), e0118432. DOI: [10.1371/journal.pone.0118432](https://doi.org/10.1371/journal.pone.0118432).
- [33] scikit-learn developers. *average_precision_score — scikit-learn documentation*. Online documentation. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.average_precision_score.html (visitado 10-02-2026).

- [34] scikit-learn developers. *GridSearchCV* — scikit-learn documentation. Online documentation. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html (visitado 11-02-2026).
- [35] scikit-learn developers. *HalvingRandomSearchCV* — scikit-learn documentation. Online documentation. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.HalvingRandomSearchCV.html (visitado 11-02-2026).
- [36] scikit-learn developers. *OneHotEncoder* — scikit-learn documentation. Online documentation. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html> (visitado 10-02-2026).
- [37] scikit-learn developers. *Pipeline* — scikit-learn documentation. Online documentation. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html> (visitado 10-02-2026).
- [38] scikit-learn developers. *RandomizedSearchCV* — scikit-learn documentation. Online documentation. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html (visitado 11-02-2026).
- [39] scikit-learn developers. *StandardScaler* — scikit-learn documentation. Online documentation. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html> (visitado 10-02-2026).
- [40] scikit-learn developers. *StratifiedKFold* — scikit-learn documentation. Online documentation. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html (visitado 11-02-2026).
- [41] scikit-optimize developers. *BayesSearchCV* — scikit-optimize documentation. Online documentation. URL: <https://scikit-optimize.github.io/stable/modules/generated/skopt.BayesSearchCV.html> (visitado 11-02-2026).
- [42] Lynn K. Shay, Gustavo J. Goni y Peter G. Black. «Effects of a warm oceanic feature on Hurricane Opal». En: *Monthly Weather Review* 128.5 (2000), págs. 1366-1383. DOI: [10.1175/1520-0493\(2000\)128<1366:EOAWOF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<1366:EOAWOF>2.0.CO;2).
- [43] Brian Tang y Kerry Emanuel. «A ventilation index for tropical cyclones». En: *Bulletin of the American Meteorological Society* 93.12 (2012), págs. 1901-1912.
- [44] Brian H Tang y JD Neelin. «ENSO influence on Atlantic hurricanes via tropospheric warming». En: *Geophysical Research Letters* 31.24 (2004).
- [45] Sudhir Varma y Richard Simon. «Bias in error estimation when using cross-validation for model selection». En: *BMC Bioinformatics* 7 (2006), pág. 91. DOI: [10.1186/1471-2105-7-91](https://doi.org/10.1186/1471-2105-7-91).
- [46] Hui Zou y Trevor Hastie. «Regularization and Variable Selection via the Elastic Net». En: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2 (2005), págs. 301-320. DOI: [10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x).

Material suplementario

A continuación se presenta el material complementario.

7.1. Cuadros

Modelo	Mejores hiperparámetros (valores redondeados)
BAYES	Sin ajuste (<i>baseline</i>).
GLM	$C = 10.00$, <code>class_weight=balanced</code> .
GLMNET	$C = 9.09$, <code>class_weight=balanced</code> , <code>l1_ratio=0.51</code> .
SVM	$C = 100.00$, <code>kernel=linear</code> , <code>gamma=scale</code> , <code>class_weight=balanced</code> .
KNN	<code>n_neighbors=21</code> , $p = 1$, <code>weights=uniform</code> .
ÁRBOL	<code>max_depth=7</code> , <code>min_samples_split=45</code> , <code>min_samples_leaf=5</code> , <code>class_weight=balanced</code> .
BOSQUES	<code>n_estimators=463</code> , <code>max_depth=2</code> , <code>min_samples_split=2</code> , <code>min_samples_leaf=1</code> , <code>max_features=log2</code> , <code>bootstrap=True</code> , <code>criterion=entropy</code> , <code>class_weight=balanced_subsample</code> .
POTENCIACIÓN	<code>n_estimators=109</code> , <code>max_depth=2</code> , <code>learning_rate=0.04</code> , <code>subsample=0.70</code> .
RED	<code>activation=tanh</code> , <code>alpha=0.00051</code> <code>learning_rate_init=0.00123</code> .

Cuadro 5: Configuración final de hiperparámetros por modelo.

Variable	coef_std	abs_coef_std
tcwv	1.512212	1.512212
t500	-1.181923	1.181923
rh600_700	-0.546243	0.546243
shear_200_850	-0.516669	0.516669
mslp_hpa	-0.422357	0.422357
sst_c	0.404557	0.404557
v850	-0.251690	0.251690
u200	-0.250956	0.250956
v200	-0.195319	0.195319
romi1	0.134251	0.134251
nino34_anom	-0.123214	0.123214
qbo_30mb	-0.071565	0.071565
u850	0.042918	0.042918
romi2	0.030758	0.030758
nao_index	-0.026029	0.026029
amo_ssta	0.002548	0.002548

Cuadro 6: Coeficientes estandarizados y su valor absoluto por variable.

7.2. Visualizaciones

A continuación se presenta el material suplementario relevante.

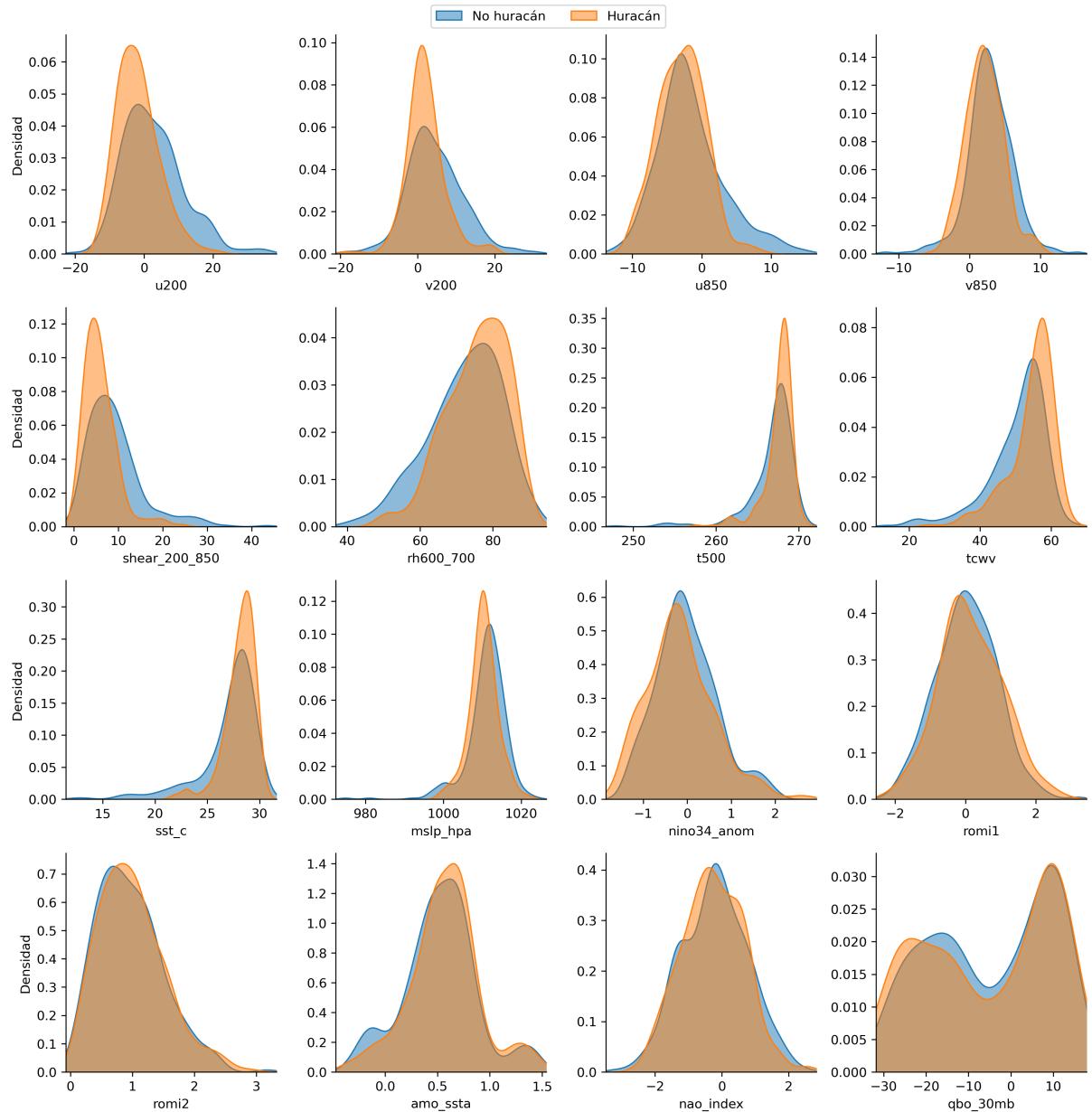


Figura 6: Densidades por clase `label`.