



Sistema de Predicción de Riesgos Operacionales en Aviación

BY: ANDRE CORAHUA

AGENDA

1. 
CONTEXTO Y PROBLEMÁTICA
2. 
PROPIUESTA DE VALOR
3. 
MVP TÉCNICO:
PILOTO CON DATOS
PÚBLICOS
4. 
RESULTADOS DEL
MODELO DE
CLASIFICACIÓN
5. 
DETECCIÓN DE
ANOMALÍAS
6. 
DEMO DE
INFERENCIA
7. 
APLICACIÓN A
HELICÓPTEROS DE
CARGA EN PERÚ
8. 
ROADMAP Y
PRÓXIMOS PASOS
9. 
CONCLUSIONES

AeroVer



AeroOver

CONTEXTO Y PROBLEMÁTICA

EL PROBLEMA:

EN PERÚ Y LATINOAMÉRICA, LAS OPERACIONES DE HELICÓPTEROS CON CARGA EXTERNA SON CRÍTICAS PARA:

- 🚧 MINERÍA
- ⚡ ENERGÍA
- 🚧 CONSTRUCCIÓN
- 🚨 RESPUESTA A EMERGENCIAS

SITUACIÓN ACTUAL:

- ✗ MUCHOS OPERADORES NO UTILIZAN SISTEMÁTICAMENTE LOS DATOS DE VUELO QUE YA REGISTRAN
- 📋 DEPENDENCIA DE REGISTROS MANUALES, HOJAS DE CÁLCULO Y REPORTES AISLADOS
- 🔎 FALTA DE HERRAMIENTAS PARA IDENTIFICAR PATRONES DE RIESGO Y EXCEDENCIAS DE PARÁMETROS

CONSECUENCIAS:

- ⚠ MAYOR PROBABILIDAD DE INCIDENTES Y ACCIDENTES (SOBRECARGA, ALTURAS INADECUADAS, TORQUE EXCESIVO)
- 💰 ALTOSES COSTOS OPERATIVOS POR DAÑOS Y TIEMPOS FUERA DE SERVICIO
- 📈 DIFICULTAD PARA EVIDENCIAR GESTIÓN PROACTIVA ANTE REGULADORES (DGAC) Y CLIENTES

PROPUESTA DE VALOR - AEROCVER

¿QUÉ ES AEROCVER?

1.

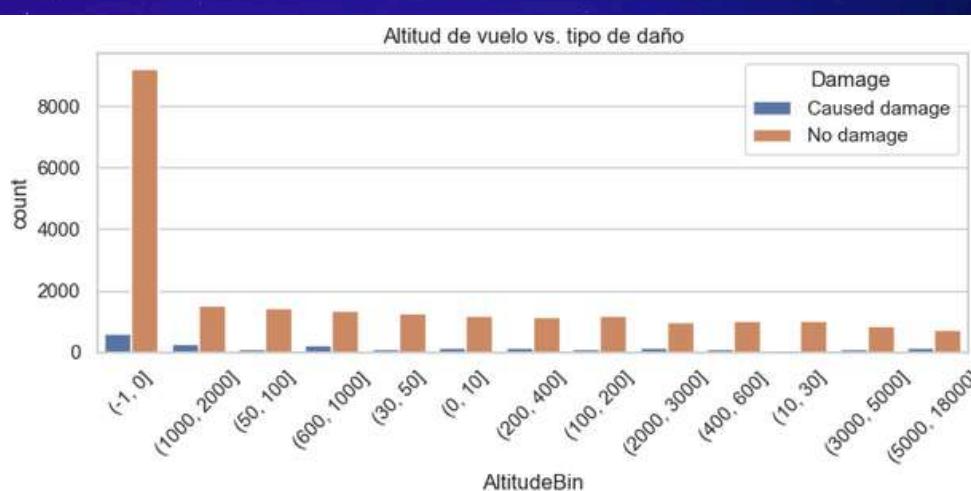
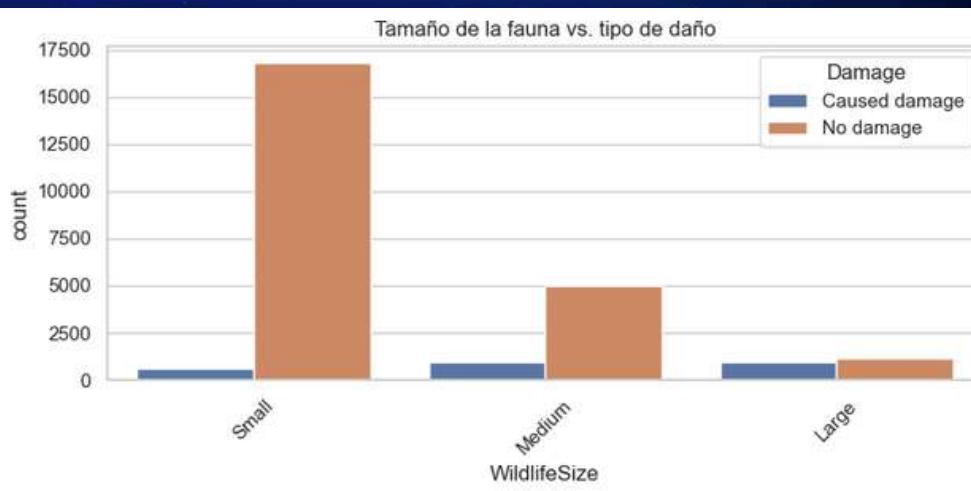
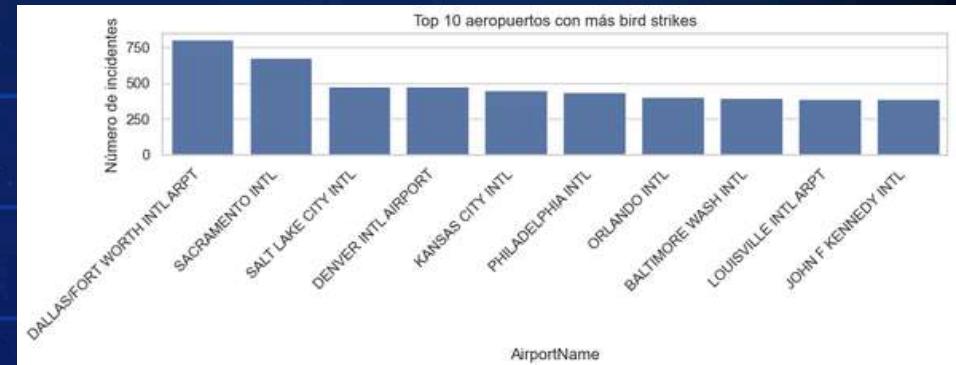
UNA PLATAFORMA DE ANÁLISIS DE DATOS DE VUELO Y GESTIÓN DE SEGURIDAD OPERACIONAL QUE:

1. INGESTA DATOS DE VUELO (ARCHIVOS EXPORTADOS DE FDR/QAR/FOQA)
2. ANALIZA PARÁMETROS CRÍTICOS (TORQUE, ALTURA, VELOCIDAD VERTICAL, TIEMPO EN HOVER CON CARGA, ETC)
3. DETECTA EXCEDENCIAS Y ANOMALÍAS USANDO ALGORITMOS DE MACHINE LEARNING
4. GENERA DASHBOARDS Y REPORTES COMPRENSIBLES PARA ÁREAS DE SEGURIDAD, MANTENIMIENTO Y GERENCIA

3

VALOR PARA EL CLIENTE:

- DETECCIÓN TEMPRANA DE MANIOBRAS FUERA DE PARÁMETROS
- PRIORIZACIÓN DE ENTRENAMIENTOS BASADA EN EVIDENCIA
- MATERIAL OBJETIVO PARA AUDITORÍAS Y REGULADORES (DGAC)
- REDUCCIÓN DE INCIDENTES Y COSTOS OPERATIVOS



MVP TÉCNICO- ESTRATEGIA DE VALIDACIÓN

OBJETIVO DEL MVP:

VALIDAR LA ARQUITECTURA TÉCNICA Y LOS MODELOS DE ANÁLISIS USANDO DATOS PÚBLICOS ANTES DE INTEGRAR DATOS REALES DE HELICÓPTEROS.

¿POR QUÉ BIRD STRIKES?

- DATASET PÚBLICO, BIEN DOCUMENTADO Y DE GRAN TAMAÑO
- PERMITE PROBAR EL PIPELINE COMPLETO: EDA → PREPROCESAMIENTO → MODELADO → INFERNERIA
- LA ARQUITECTURA ES REUTILIZABLE PARA CUALQUIER TIPO DE DATO DE VUELO

DATASET PILOTO:

-  FUENTE: BIRD STRIKES (FAA/NTSB) – IMPACTOS DE AVES EN AERONAVES
-  REGISTROS: ~25,429 INCIDENTES TRAS LIMPIEZA
-  VARIABLE OBJETIVO: ISSEVERE (1 = CAUSED DAMAGE, 0 = NO DAMAGE)
-  VARIABLES PREDICTORAS: TIPO DE AERONAVE, AEROPUERTO, RANGO DE ALTITUD, MODELO, TAMAÑO DE FAUNA, CONDICIONES DEL CIELO, FASE DE VUELO, MES, AÑO



ARQUITECTURA DEL PROYECTO

STACK TECNOLÓGICO:

- LENGUAJE: PYTHON 3.X
- LIBRERÍAS: PANDAS, NUMPY, SCIKIT-LEARN, SEABORN, MATPLOTLIB, JOBLIB
- MODELOS: RANDOMFORESTCLASSIFIER, ISOLATIONFOREST
- VERSIONADO: GIT + GITHUB

ESTRUCTURA DE NOTEBOOKS:

aerosafe-risk-predictor/

```
01_exploracion_datos_publicos.ipynb
    EDA: distribución de variables, análisis visual

02_preprocesamiento_base.ipynb
    Limpieza, creación de IsSevere, encoding

03_modelos_anomalias_pyod.ipynb
    Detección de anomalías con IsolationForest

04_modelos_clasificacion_basica.ipynb
    Random Forest, ajuste de umbral, demo de inferencia
```



andrecorahua	Update README.md	82ff68c · 11 hours
data	feat: modelo RF con ajuste de umbral y guardado en pkl	
notebooks	feat: análisis de anomalías con IsolationForest y conclusiones	
.gitignore	Initial commit	
LICENSE	Initial commit	
README.md	Update README.md	
requirements.txt	Update requirements.txt	

Name
..
01_exploracion_datos_publicos.ipynb
02_preprocesamiento_base.ipynb
03_modelos_anomalias_pyod.ipynb
04_modelos_clasificacion_basica.ipynb

MODELO DE CLASIFICACIÓN - ENFOQUE



AeroOver

OBJETIVO DEL
MODELO:

PREDECIR LA
PROBABILIDAD
DE QUE UN
INCIDENTE
CAUSE DAÑO
FÍSICO EN LA
AERONAVE.

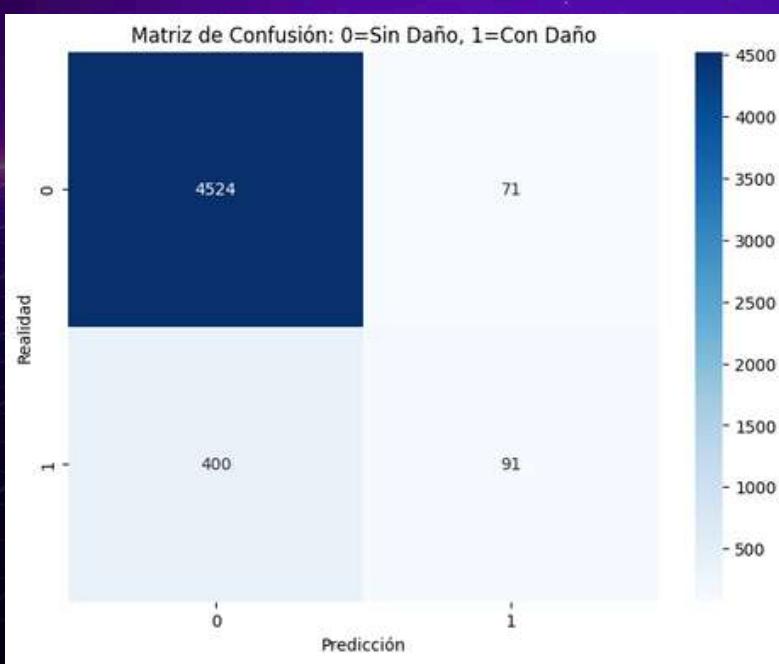
ALGORITMO:

- RANDOMFORESTCLASSIFIER (ENSEMBLE DE ÁRBOLES DE DECISIÓN)
VENTAJAS: ROBUSTO, MANEJA BIEN VARIABLES CATEGÓRICAS, INTERPRETABLE

PREPROCESAMIENTO: DIVISIÓN DE DATOS:

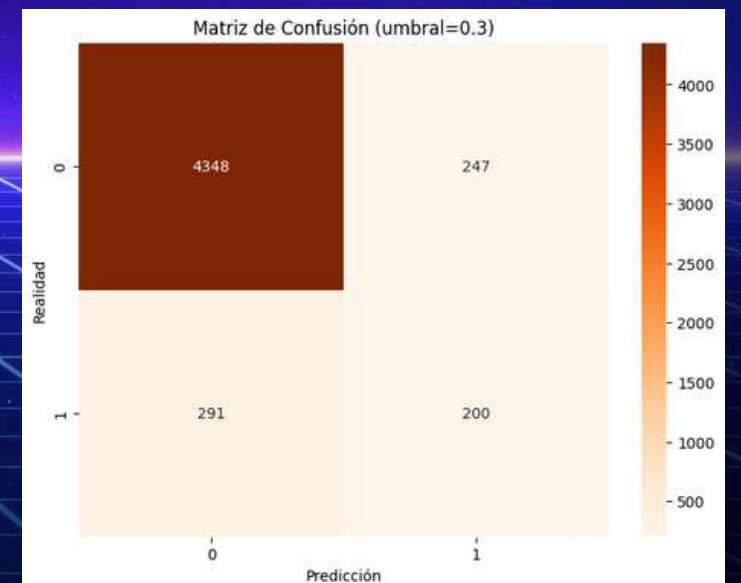
- ONE-HOT ENCODING PARA VARIABLES CATEGÓRICAS (AEROPUERTO, TIPO DE AERONAVE, ETC)
- VARIABLES NUMÉRICAS: MONTH, YEAR (SIN TRANSFORMACIÓN)

- 80% ENTRENAMIENTO, 20% PRUEBA
- ESTRATIFICACIÓN POR ISSEVERE PARA MANTENER PROPORCIONES



MANEJO DE
DESBALANCE:

CLASS_WEIGHT='BALANCED'
PARA DAR MÁS PESO A LA CLASE MINORITARIA (CON DAÑO)



RESULTADOS - UMBRAL ESTÁNDAR (0.50)

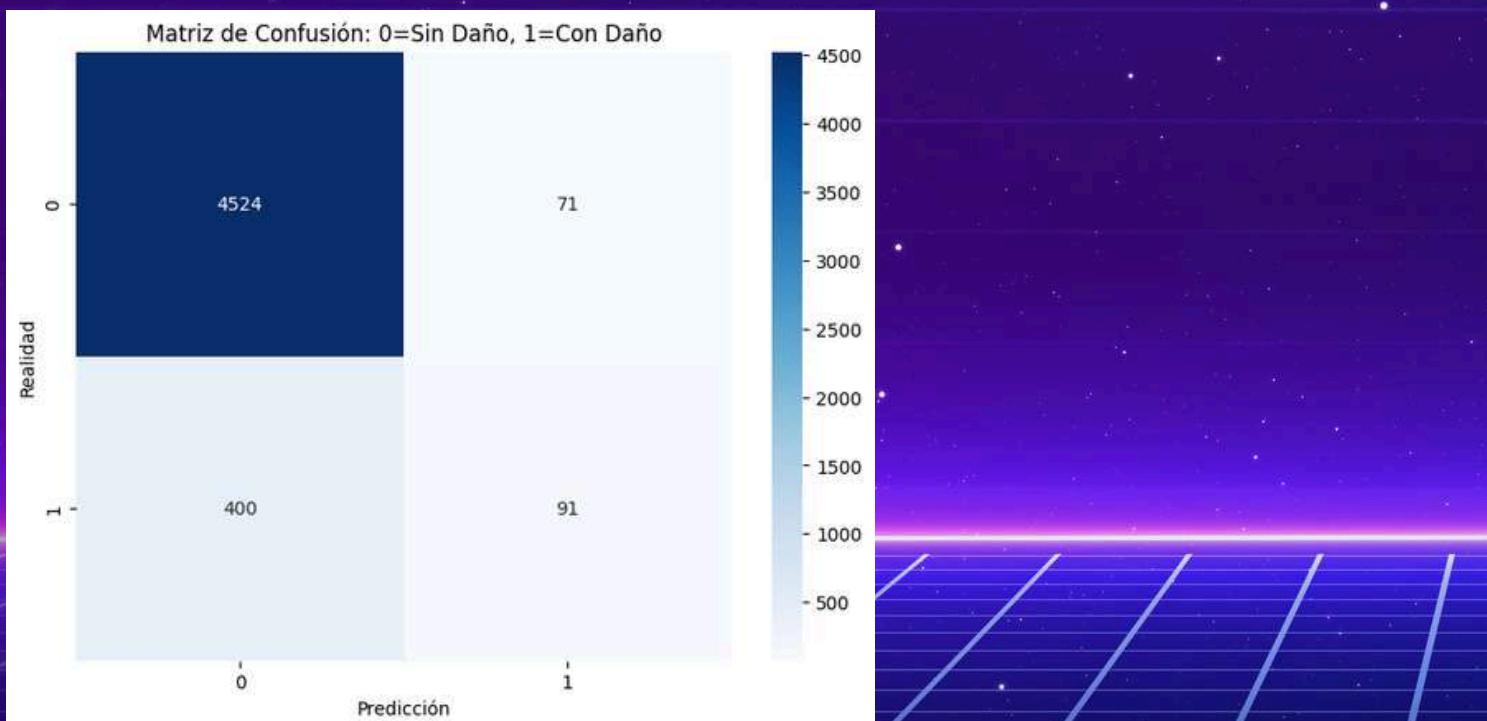


AeroVer

MÉTRICAS CON UMBRAL 0.50:

CLASE	DESCRIPCIÓN	PRECISION	RECALL	F1-SCORE
0	SIN DAÑO	0.92	0.98	0.95
1	CON DAÑO	0.56	0.19	0.28

ACCURACY GLOBAL: 0.91



INTERPRETACIÓN:

- ✅ EXCELENTE DESEMPEÑO PARA DETECTAR INCIDENTES SIN DAÑO
- ⚠️ RECALL BAJO (19%) PARA INCIDENTES CON DAÑO: SOLO DETECTA 91 DE 491 CASOS REALES
- ESTO ES TÍPICO EN DATASETS DESBALANCEADOS CON UMBRAL ESTÁNDAR

--- Reporte de Clasificación ---

	precision	recall	f1-score	support
0	0.92	0.98	0.95	4595
1	0.56	0.19	0.28	491
accuracy			0.91	5086
macro avg	0.74	0.58	0.61	5086
weighted avg	0.88	0.91	0.89	5086

AJUSTE DE UMBRAL - ENFOQUE DE SEGURIDAD

EVALUACIÓN DE UMBRALES:

UMbral	Precision (Clase 1)	Recall (Clase 1)
0.50	0.58	0.20
0.40	0.54	0.30
0.30	0.45	0.41
0.25	0.40	0.42
0.20	0.35	0.54

JUSTIFICACIÓN:

EN SEGURIDAD OPERACIONAL, ES PREFERIBLE REVISAR MÁS CASOS (AUNQUE ALGUNOS SEAN FALSOS POSITIVOS) QUE DEJAR ESCAPAR INCIDENTES CRÍTICOS.

--- Umbral: 0.50 ---

clase 0 - precision: 0.92, recall: 0.98
clase 1 - precision: 0.58, recall: 0.20

--- Umbral: 0.40 ---

clase 0 - precision: 0.93, recall: 0.97
clase 1 - precision: 0.54, recall: 0.30

--- Umbral: 0.30 ---

clase 0 - precision: 0.94, recall: 0.95
clase 1 - precision: 0.45, recall: 0.41

--- Umbral: 0.25 ---

clase 0 - precision: 0.94, recall: 0.92
clase 1 - precision: 0.40, recall: 0.47

--- Umbral: 0.20 ---

clase 0 - precision: 0.95, recall: 0.89
clase 1 - precision: 0.35, recall: 0.54

DECISIÓN: UMBRAL 0.30

- ✓ RECALL DE CLASE 1 SUBE DE 0.19 A 0.41 (MÁS DEL DOBLE)
- ✓ SE DETECTAN 200 CASOS CON DAÑO (VS. 91 CON UMBRAL 0.5)
- ⚠ A COSTA DE MÁS FALSAS ALARMAS (24P VS. 71)

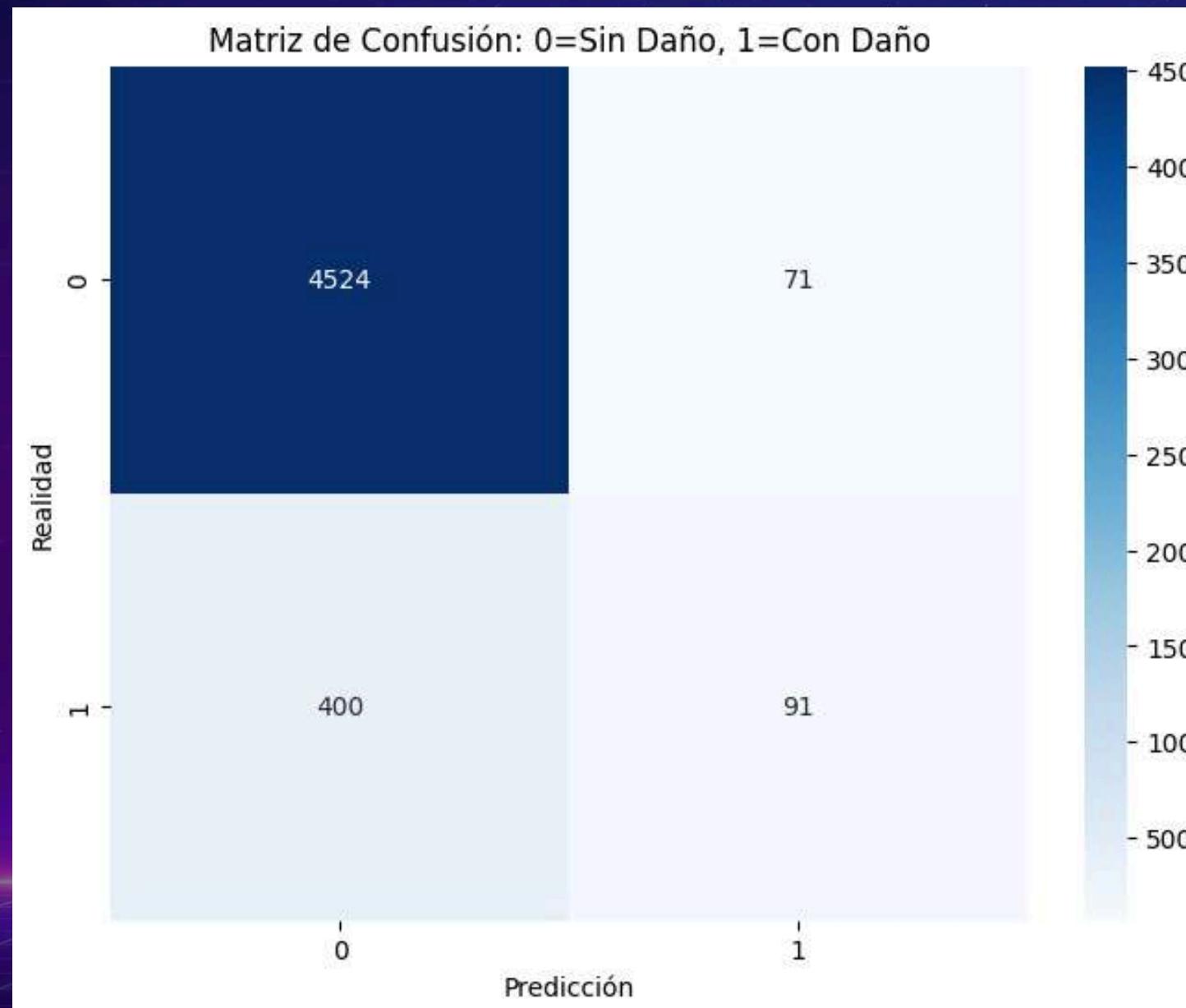


MATRIZ DE CONFUSIÓN - COMPARATIVA

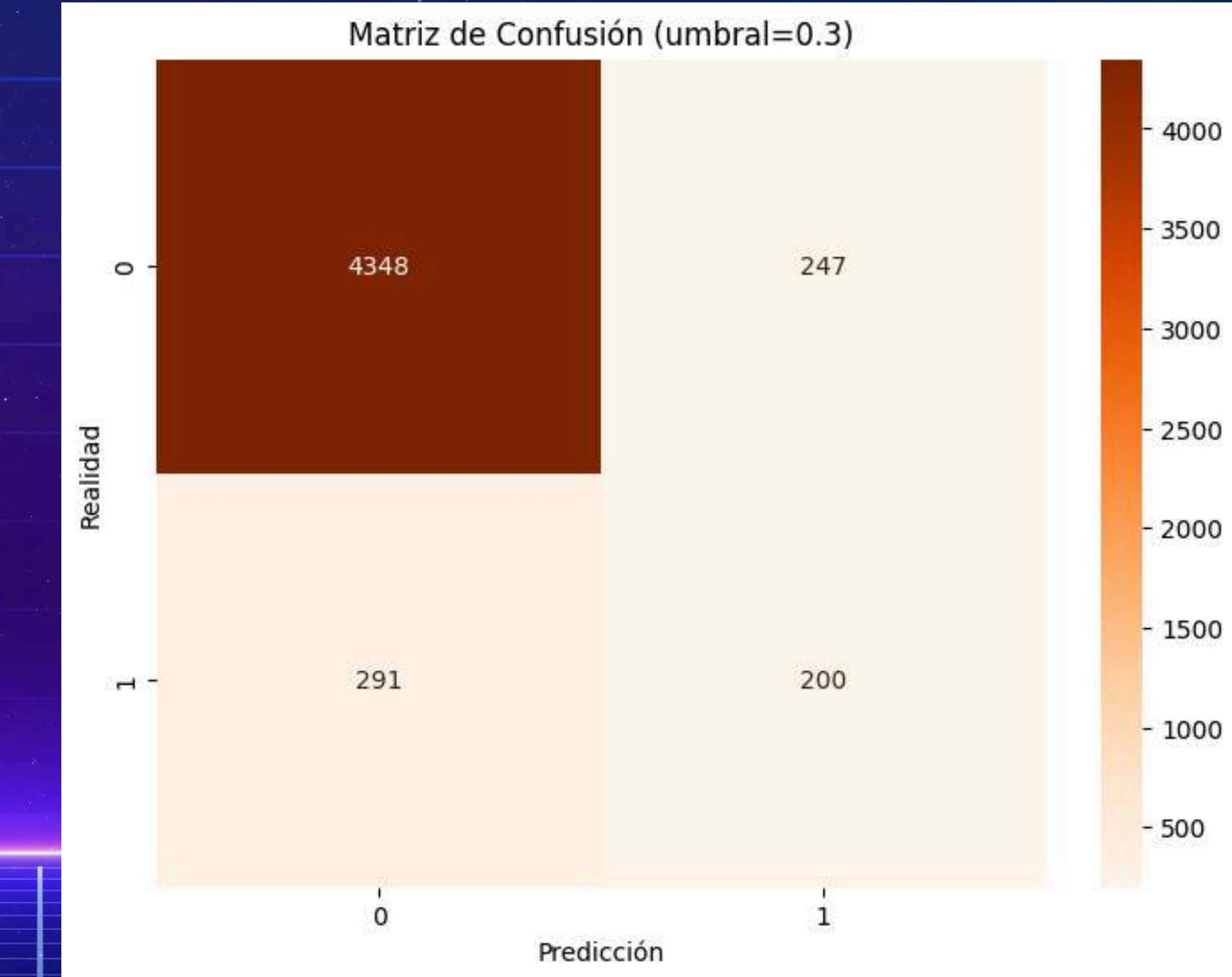


AeroVer

UMBRAL 0.50:



UMBRAL 0.30:



MEJORA CLAVE:

- ▼ FALSOS NEGATIVOS (CASOS CON DAÑO NO DETECTADOS): 400 → 291 (-27%)
- ▲ VERDADEROS POSITIVOS (CASOS CON DAÑO DETECTADOS): 91 → 200 (+120%)

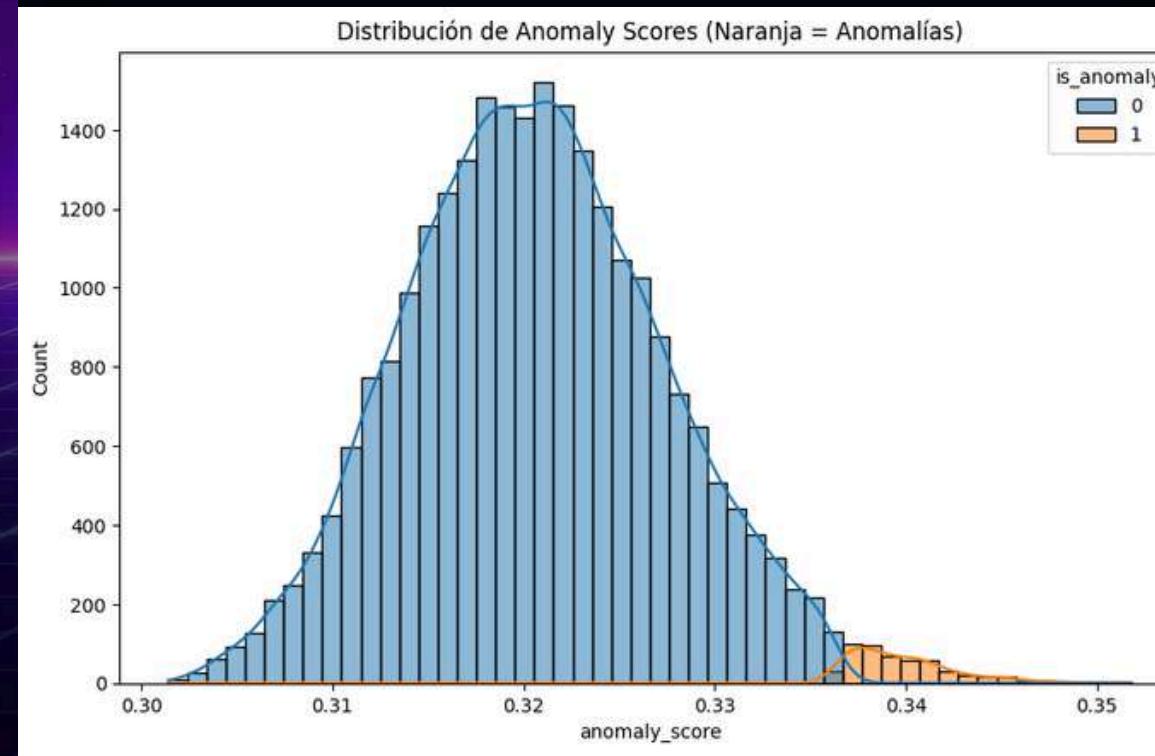
DETECCIÓN DE ANOMALÍAS - ENFOQUE



AeroVer

OBJETIVO:

IDENTIFICAR INCIDENTES "RAROS" O INUSUALES QUE PUEDAN REPRESENTAR SITUACIONES DE RIESGO ELEVADO, SIN NECESIDAD DE ETIQUETAS PREVIAS.



ALGORITMO:

- ISOLATION FOREST (SCIKIT-LEARN)
- PRINCIPIO: LOS PUNTOS ANÓMALOS SON MÁS FÁCILES DE AISLAR EN EL ESPACIO DE CARACTERÍSTICAS

CONFIGURACIÓN:

- CONTAMINACIÓN: 2% (ASUMIMOS ~2% DE INCIDENTES ANÓMALOS)
- VARIABLES: MISMAS QUE EL MODELO DE CLASIFICACIÓN (TRAS ONE-HOT ENCODING)

SALIDA:

- CADA INCIDENTE RECIBE UN ANOMALY_SCORE (MAYOR = MÁS RARO)
- CLASIFICACIÓN BINARIA: 0 = NORMAL, 1 = ANOMALÍA

RESULTADOS - DETECCIÓN DE ANOMALÍAS



AeroVer

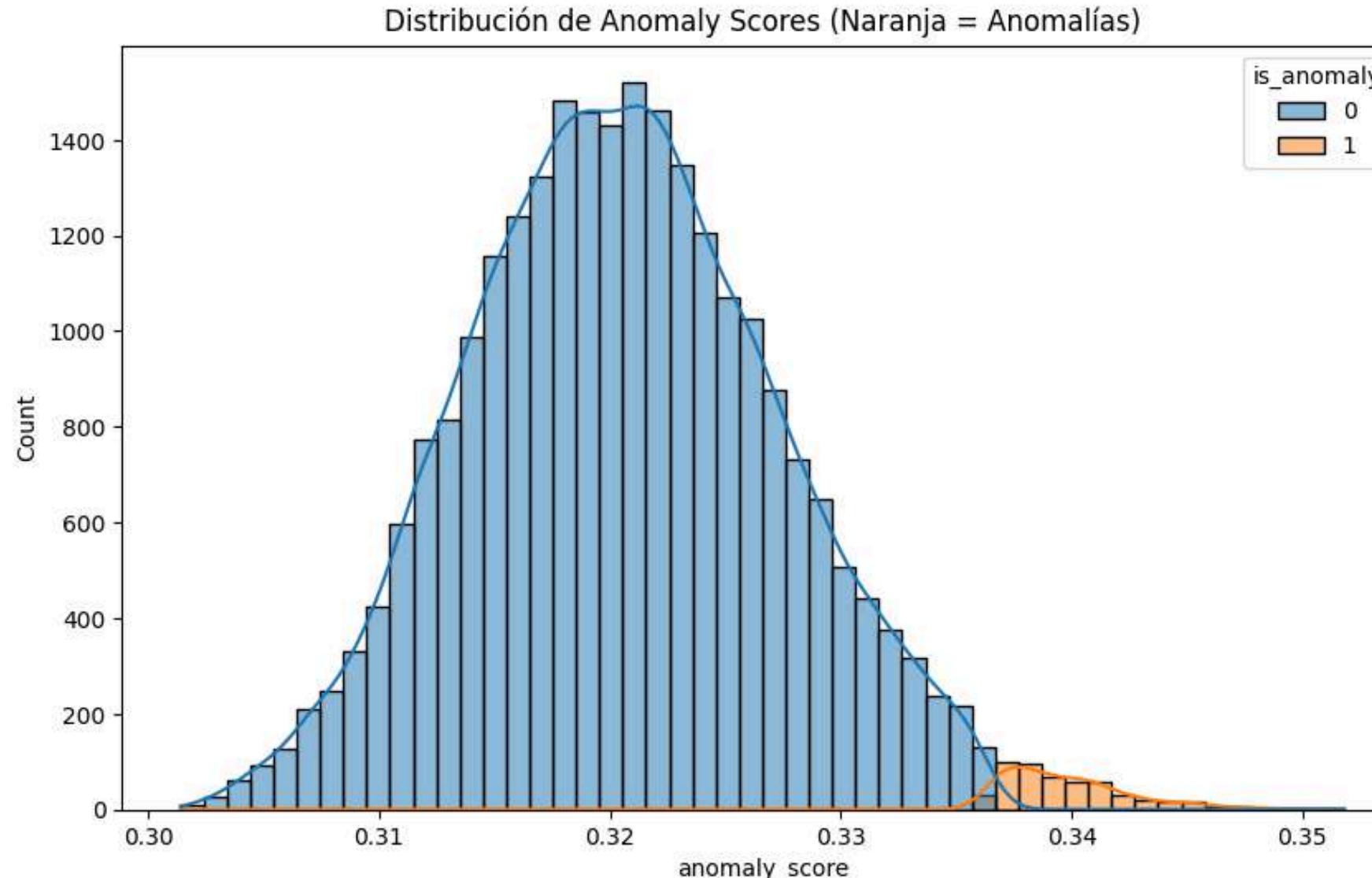
CANTIDAD DE ANOMALÍAS DETECTADAS:

- TOTAL DE REGISTROS: 25,429
- ANOMALÍAS DETECTADAS: 509 (~2%)

HALLAZGO CLAVE:

LOS INCIDENTES ANÓMALOS CONCENTRAN UN RIESGO DE DAÑO 5 VECES MAYOR QUE LOS NORMALES.

Distribución de Anomaly Scores (Naranja = Anomalías)



DISTRIBUCIÓN DE ANOMALY SCORES:

- GRÁFICO: HISTOGRAMA CON DOS COLORES (AZUL = NORMAL, NARANJA = ANOMALÍA)
- LAS ANOMALÍAS SE CONCENTRAN EN LA COLA DERECHA (SCORES MÁS ALTOS)

PROPORCIÓN DE DAÑO GRAVE (ISSEVERE = 1):

GRUPO	PROPORCIÓN CON DAÑO
NORMALES (0)	8.9%
ANOMALÍAS (1)	44.2%



AeroVer

PATRONES EN ANOMALÍAS

--- Top Aeropuertos en Anomalías ---

AirportName	
BALTIMORE WASH INTL	101
SACRAMENTO INTL	45
ORLANDO INTL	38
DALLAS/FORT WORTH INTL ARPT	35
KANSAS CITY INTL	32
PHILADELPHIA INTL	20
CHICAGO O'HARE INTL ARPT	20
GEORGE BUSH INTERCONTINENTAL	16
EPPLER AIRFIELD	15
LOS ANGELES INTL	11

Name: count, dtype: int64

--- Top Fases de Vuelo en Anomalías ---

FlightPhase	
Climb	216
Approach	161
Descent	117
Take-off run	14
Landing Roll	1

Name: count, dtype: int64

INTERPRETACIÓN:

- LAS ANOMALÍAS SE CONCENTRAN EN FASES CRÍTICAS (ASCENSO, APROXIMACIÓN, DESCENSO)
- ALGUNOS AEROPUERTOS PRESENTAN FRECUENCIAS INUSUALMENTE ALTAS, SUGIRIENDO FACTORES LOCALES (FAUNA, GEOGRAFÍA, TRÁFICO)

DEMO DE INFERENCIA

OBJETIVO:

DEMOSTRAR QUE EL MODELO PUEDE USARSE EN PRODUCCIÓN PARA PREDECIR EL RIESGO DE UN VUELO NUEVO.

PROCESO:

- EL MODELO ENTRENADO SE GUARDA COMO RF_MODEL_RISK.PKL (PIPELINE COMPLETO)
- SE CARGA CON JOBLIB.LOAD()
- SE INGRESA UN VUELO DE EJEMPLO CON SUS CARACTERÍSTICAS

```
... Vuelo de ejemplo:  
 AircraftType  AirportName  AltitudeBin  MakeModel WildlifeSize \  
 0  Airplane    LAGUARDIA NY  (1000, 2000]  B-737-400      Medium  
  
 ConditionsSky PilotWarned FlightPhase Month Year  
 0  No Cloud        N       Climb     11  2000  
  
Probabilidad estimada de daño: 0.74  
Clasificación final con umbral 0.3 (1=Con daño, 0=Sin daño): 1
```

SALIDA:

- PROBABILIDAD ESTIMADA DE DAÑO: 0.74 (74%)
- CLASIFICACIÓN CON UMBRAL 0.30: 1 (CON DAÑO / ALTO RIESGO)



```
racion_datos_publicos.ipynb  02_preprocesamiento_base.ipynb  04_modelos_clasificacion_basica.ipynb X  03_modelos_anomalias_pyod.ipynb  ...  
notebooks > 04_modelos_clasificacion_basica.ipynb > M4. Conclusiones del modelo de clasificación básica > M4 Optimización del Umbral de Riesgo > # CELDA 11: Definir un vuelo de ejemplo y predecir el riesgo  
Generate + Code + Markdown | Run All Restart Clear All Outputs Jupyter Variables Outline ... Python 3.14.2  
# CELDA 11: Definir un vuelo de ejemplo y predecir el riesgo  
# Debe tener las MISMAS columnas que X:  
# ['AircraftType', 'AirportName', 'AltitudeBin', 'MakeModel',  
# 'WildlifeSize', 'ConditionsSky', 'PilotWarned', 'FlightPhase',  
# 'Month', 'Year']  
  
ejemplo = {  
    "AircraftType": "Airplane",  
    "AirportName": "LAGUARDIA NY",  
    "AltitudeBin": "(1000, 2000]",  
    "MakeModel": "B-737-400",  
    "WildlifeSize": "Medium",  
    "Conditionssky": "No Cloud",  
    "PilotWarned": "N",  
    "FlightPhase": "Climb",  
    "Month": 11,  
    "Year": 2000,  
}  
  
df_ejemplo = pd.DataFrame([ejemplo])  
  
# Usamos el modelo cargado  
prob_dano = loaded_model.predict_proba(df_ejemplo)[:, 1][0]  
  
# Aplicar el mismo umbral que definimos antes (0.30)  
best_threshold = 0.30  
pred_clase = int(prob_dano >= best_threshold)  
  
print("Vuelo de ejemplo:")  
print(df_ejemplo)  
print("\nProbabilidad estimada de daño:", round(prob_dano, 3))  
print(f"Clasificación final con umbral {best_threshold} (1=Con daño, 0=Sin daño):", pred_clase)
```

APLICACIÓN A HELICÓPTEROS DE CARGA EN PERÚ

ADAPTACIÓN PARA HELICÓPTEROS: *AeroOver*

¿CÓMO SE TRASLADA ESTE ENFOQUE?

AUNQUE EL PILOTO USA DATOS DE BIRD STRIKES EN AVIACIÓN GENERAL, LA ARQUITECTURA ES AGNÓSTICA AL TIPO DE AERONAVE.

FUENTE DE DATOS:

- 📁 ARCHIVOS FOQA/HFDM EXPORTADOS DE HELICÓPTEROS
- 📋 REPORTES DE INCIDENTES INTERNOS DEL OPERADOR

DISTRIBUCIÓN DE ANOMALY SCORES:

- GRÁFICO: HISTOGRAMA CON DOS COLORES (AZUL = NORMAL, NARANJA = ANOMALÍA)
- LAS ANOMALÍAS SE CONCENTRAN EN LA COLA DERECHA (SCORES MÁS ALTOS)

BIRD STRIKES (PILOTO)	HELICÓPTEROS CON CARGA EXTERNA (AEROCVER)
TIPO DE AERONAVE	TIPO DE HELICÓPTERO (MI-171, MI-8AMT, MI-8MTV-1)
AEROPUERTO	BASE DE OPERACIONES / ZONA DE MISIÓN
RANGO DE ALTITUD	ALTURA SOBRE TERRENO (AGL)
FASE DE VUELO	FASE DE MISIÓN (HOVER, TRASLADO, DESCARGA)
TAMAÑO DE FAUNA	PESO DE CARGA EXTERNA
CONDICIONES DEL CIELO	CONDICIONES METEOROLÓGICAS
VARIABLE OBJETIVO: DAÑO	VARIABLE OBJETIVO: VUELO DE ALTO RIESGO (EXCEDENCIAS DE TORQUE, VELOCIDAD VERTICAL, ETC.)

ROADMAP Y PRÓXIMOS PASOS



FASE 1: MVP TÉCNICO (COMPLETADO)

- PILOTO CON DATOS PÚBLICOS DE BIRD STRIKES
- VALIDACIÓN DE ARQUITECTURA DE ANÁLISIS
- MODELOS DE CLASIFICACIÓN Y ANOMALÍAS FUNCIONALES

FASE 2: INTEGRACIÓN CON OPERADORES (MESES 1-4)

- IDENTIFICAR 1-2 OPERADORES DE HELICÓPTEROS PILOTO
- OBTENER DATOS FOQA/HFDM DE 10-50 VUELOS
- ADAPTAR VARIABLES Y REENTRENAR MODELOS
- GENERAR REPORTES Y DASHBOARDS PERSONALIZADOS

FASE 3: PRODUCTIZACIÓN (MESES 5-8)

- DESARROLLAR PLATAFORMA WEB (DASHBOARD INTERACTIVO)
- AUTOMATIZAR INGESTA DE DATOS Y GENERACIÓN DE REPORTES
- DEFINIR MODELO DE SUSCRIPCIÓN (USD 300-800/AERONAVE/MES)

FASE 4: ESCALAMIENTO REGIONAL (12+ MESES)

- EXPANDIR A OTROS OPERADORES EN PERÚ Y LATINOAMÉRICA
- INTEGRAR CON SISTEMAS DE MANTENIMIENTO (OMAS)
- EXPLORAR MODELOS AVANZADOS (LSTM, AUTOENCODERS PARA SERIES TEMPORALES)

IMPACTO ESPERADO



PARA OPERADORES:

- ✓ REDUCCIÓN DE INCIDENTES Y ACCIDENTES
- ✓ AHORRO EN COSTOS DE MANTENIMIENTO Y TIEMPOS FUERA DE SERVICIO
- ✓ MEJORA EN CULTURA DE SEGURIDAD OPERACIONAL
- ✓ EVIDENCIA OBJETIVA PARA AUDITORÍAS DGAC Y REGULADORES

PARA OMAS:

- ✓ DATOS OPERACIONALES QUE RESPALDAN MANTENIMIENTO PREVENTIVO
- ✓ IDENTIFICACIÓN DE PATRONES DE FALLAS RECURRENTES
- ✓ MEJOR PLANIFICACIÓN DE RECURSOS

PARA EL SECTOR:

- ✓ DEMOCRATIZACIÓN DEL ACCESO A HERRAMIENTAS DE FDM/FOQA
- ✓ ALINEAMIENTO CON ESTÁNDARES INTERNACIONALES (ICAO, IATA)
- ✓ CONTRIBUCIÓN A LA REDUCCIÓN DE LA TASA DE ACCIDENTES EN AVIACIÓN REGIONAL

CONCLUSIONES

PRINCIPALES HALLAZGOS:



EL PILOTO CON DATOS DE BIRD STRIKES DEMUESTRA QUE
ES POSIBLE CONSTRUIR MODELOS PREDICTIVOS DE
RIESGO CON ALTA CAPACIDAD DE DISCRIMINACIÓN



AL REDUCIR EL UMBRAL DE DECISIÓN A 0.30, EL RECALL DE LA CLASE CRÍTICA AUMENTA DE 19% A 41%, PRIORIZANDO LA DETECCIÓN DE EVENTOS DE ALTO RIESGO



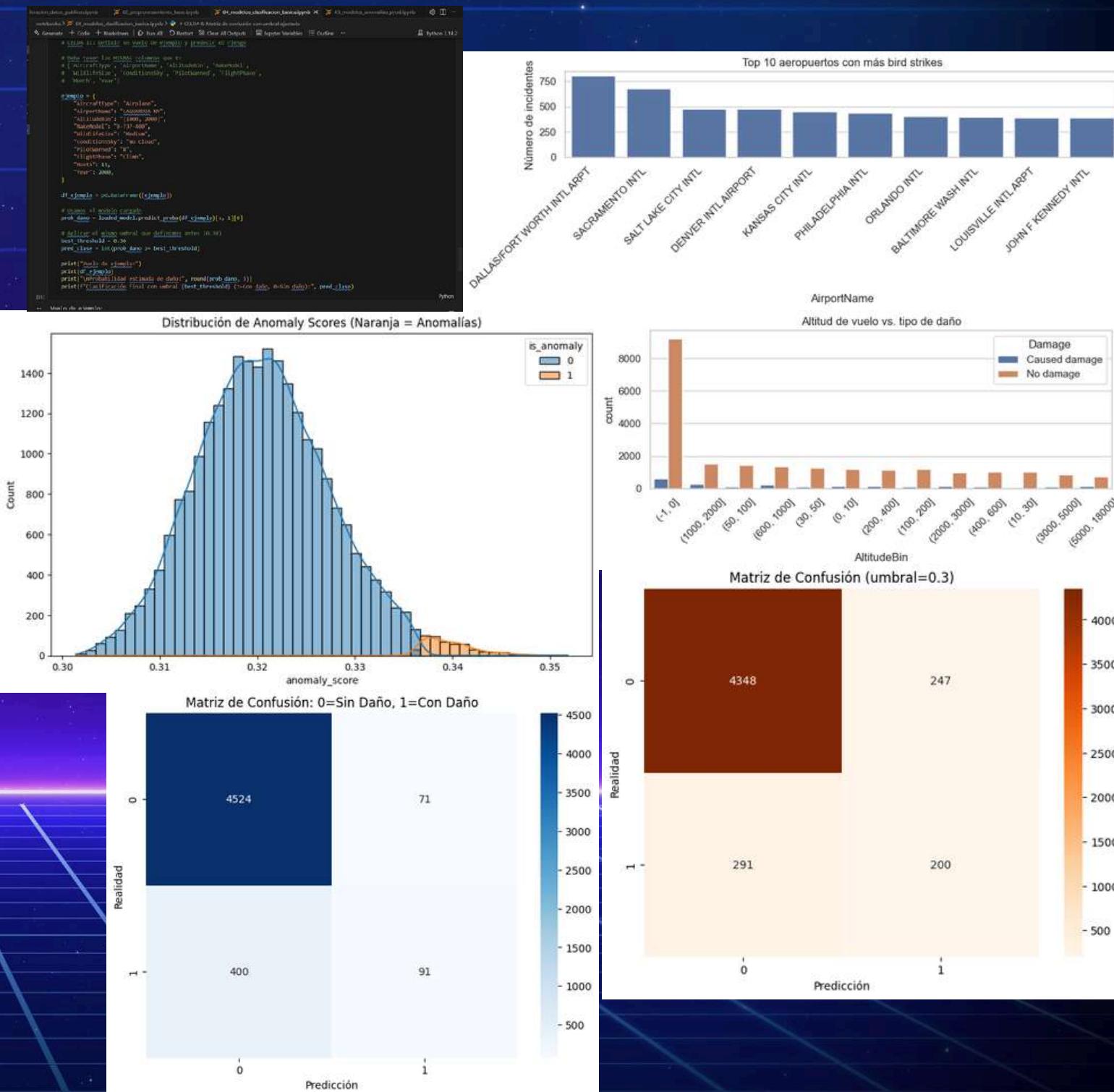
LOS INCIDENTES MARCADOS COMO ANÓMALOS CONCENTRAN UN RIESGO DE DAÑO 5 VECES MAYOR, VALIDANDO EL ENFOQUE DE RANKING DE VUELOS CRÍTICOS



EL PIPELINE DESARROLLADO ES DIRECTAMENTE APLICABLE A DATOS DE HELICÓPTEROS CON CARGA EXTERNA, SUSTITUYENDO VARIABLES Y REENTRENANDO MODELOS



AEROCVER CIERRA LA BRECHA ENTRE LAS RECOMENDACIONES INTERNACIONALES DE FDM Y LA REALIDAD DE OPERADORES MEDIANOS EN LATINOAMÉRICA





AeroOver

PREGUNTAS Y CONTACTO



PHONE
NUMBER

984 - 268 - 854



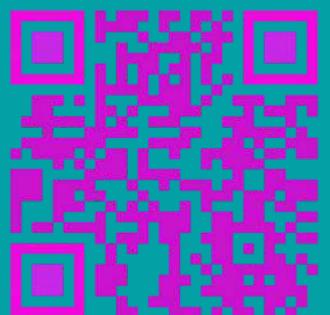
EMAIL
ADDRESS

andrecorahua@gmail.com



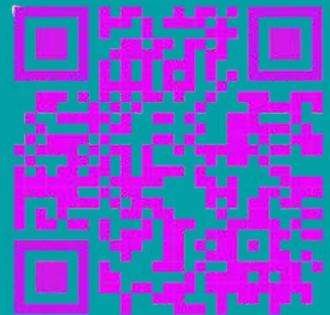
LINKEDIN

ANDRE CORAHUA Andre Corahua
andrecorahua



GITHUB

Andre Corahua
andrecorahua



RECURSOS:

- INFORME TÉCNICO AEROCVER (MVP)
- ANEXO TÉCNICO - RESULTADOS PILOTO BIRD STRIKES
- REPOSITORIO GITHUB: AEROSAFE-RISK-PREDICTOR (CÓDIGO COMPLETO Y REPRODUCIBLE)