



UNIVERSIDAD
SAN SEBASTIAN
Ilumina el futuro

**FACULTAD DE INGENIERÍA
SEDE SANTIAGO**

**MODELO PREDICTIVO MEDIANTE *MACHINE LEARNING* PARA EL
ENVÍO DE MENSAJES SAE ANTE LA OCURRENCIA DE INCENDIOS
FORESTALES EN EL TERRITORIO NACIONAL**

Tesis para optar al Grado de Magíster en Data Science

Profesor Tutor
Estudiante

: Ph.D. Cesar Ravello Serrano
: **Francisco Javier Páez Ramírez**



UNIVERSIDAD
SAN SEBASTIAN
Ilumina el futuro

© **Francisco Javier Páez Ramírez**

Se autoriza la reproducción parcial o total de esta obra con fines académicos, por cualquier forma, medio o procedimiento, siempre y cuando se incluya la cita bibliográfica del documento.

Santiago de Chile
2026



UNIVERSIDAD
SAN SEBASTIAN
Ilumina el futuro

HOJA DE CALIFICACIÓN

En Santiago, el 2026, los abajo firmantes dejan constancia que el estudiante Francisco Javier Páez Ramírez del Magíster en Data Science han aprobado el Documento y la Defensa del Proyecto para optar al Grado de Magíster Data Science con una nota de....

(Nombre y firma profesor evaluador)

(Nombre y firma profesor evaluador)

(Nombre y firma profesor evaluador)

DEDICATORIA

Dedico esta investigación a aquellas familias que han perdido a sus seres queridos producto de los grandes incendios forestales que han afectado a Chile en los últimos años. Que este trabajo represente, aunque sea de manera modesta, un aporte orientado a fortalecer la prevención y la protección de la vida ante emergencias futuras.

Asimismo, la dedico a los funcionarios y funcionarias de carrera de la Administración Pública con vocación de servicio que, desde sus respectivos espacios institucionales, buscan permanentemente mejorar los procesos internos del Estado, muchas veces enfrentando las complejidades propias de la gestión pública y las limitaciones estructurales. Que este esfuerzo sea un reconocimiento a quienes creen que el servicio público puede y debe evolucionar para proteger de mejor manera a la ciudadanía.



UNIVERSIDAD
SAN SEBASTIAN
Ilumina el futuro

AGRADECIMIENTOS

Agradezco en primer lugar, a mi familia, por su apoyo incondicional, comprensión y paciencia durante el desarrollo de esta investigación. Su respaldo ha sido fundamental para culminar esta etapa.

Extiendo mi reconocimiento a mi profesor guía por su acompañamiento constante y confianza en esta investigación, aportando rigurosidad y claridad en cada una de sus observaciones.

Asimismo, agradezco a los funcionarios de CONAF y a mis compañeros y compañeras de trabajo de SENAPRED, quienes desde su experiencia y compromiso con la gestión del riesgo de desastres han contribuido directa e indirectamente en el desarrollo de este trabajo.



RESUMEN

Esta investigación desarrolla un modelo predictivo basado en **Machine Learning** para optimizar la activación del **Sistema de Alerta de Emergencia (SAE)** ante incendios forestales en Chile, superando las limitaciones de las evaluaciones manuales actuales que retrasan la obtención de información de activar la alerta SAE.

Metodológicamente, se consolidó un *dataset* con datos históricos de CONAF (meteorología, topografía y combustible), registros de SENAPRED y cartografía de Riesgo Territorial. Se integró el **Déficit de Presión de Vapor (VPD)** como indicador clave y se aplicó un filtrado condicional para corregir el desbalance de clases. Se comparó el desempeño de los algoritmos **Random Forest** y **XGBoost** bajo un enfoque de gestión de riesgos prioritario en la seguridad pública.

Los resultados demostraron que el modelo **XGBoost** resultó superior, identificando al **Riesgo Territorial** como la variable predictora dominante (39.4%), seguida por la humedad y el viento. Mediante un ajuste operativo del umbral de decisión, se estableció un **umbral de decisión de 0.30**, logrando una **sensibilidad (Recall) del 89.9%**. Este ajuste permitió detectar un 7% adicional de alertas reales en comparación con el umbral estándar, priorizando la reducción de falsos negativos. El estudio concluye que el modelo se alinea con el principio de prevención de SENAPRED, consolidándose como una herramienta estratégica de apoyo a la toma de decisiones para la protección civil.

Palabras clave: Riesgo Territorial - *Machine Learning* - Modelos Predictivos - Mensajería SAE - Gestión del Riesgo.

ABSTRACT

This research develops a machine learning–based predictive model to optimize the activation of the Emergency Alert System (SAE) in response to wildfires in Chile, overcoming the limitations of current manual assessments that delay the timely activation of SAE alerts.

A consolidated dataset was built using historical data from CONAF (meteorological, topographic, and fuel variables), records from SENAPRED, and Territorial Risk mapping. Vapor Pressure Deficit (VPD) was incorporated as a key indicator, and conditional filtering was applied to address class imbalance. The performance of Random Forest and XGBoost algorithms was compared under a risk management approach that prioritizes public safety.

The results showed that the XGBoost model outperformed the alternative model, identifying Territorial Risk as the dominant predictive variable (39.4%), followed by humidity and wind. Through operational calibration, a decision threshold of 0.30 was established, achieving a sensitivity (Recall) of 89.9%. This adjustment enabled the detection of an additional 7% of true alerts compared to the standard threshold, prioritizing the reduction of false negatives. The study finds that the model aligns with SENAPRED's prevention principle, consolidating it as a strategic decision-support tool for civil protection.

Keywords: Territorial Risk - Machine Learning - Predictive Models - SAE Messaging - Risk Management



TABLA DE CONTENIDOS

HOJA DE CALIFICACIÓN	ii
DEDICATORIA	iii
AGRADECIMIENTOS	iv
RESUMEN.....	v
ABSTRACT	vi
TABLA DE CONTENIDOS	vii
ÍNDICE DE TABLAS.....	xii
ÍNDICE DE FIGURAS	xiii
INTRODUCCIÓN.....	1
CAPITULO I: ANTECEDENTES DEL PROBLEMA.....	4
1.1 Antecedentes de la Organización.....	4
1.1.1 <i>Evolución Histórica</i>	4
1.1.2 <i>Presencia y Propósito</i>	4
1.2 Formulación del Problema	6
1.3 Alcance	6
1.4 Delimitaciones	6
1.4.1 <i>Delimitación Geográfica</i>	6
1.4.2 <i>Delimitación Poblacional</i>	7
1.4.3 <i>Delimitación de Tiempo</i>	7
1.5 Limitaciones	7
1.6 Objetivos	8



1.6.1	Objetivo General	8
1.6.2	Objetivos Específicos	8
CAPÍTULO II: MARCO TEÓRICO		9
2.1	Sistema de Alerta de Emergencia (SAE)	9
2.1.1	Implementación del Sistema de Alerta de Emergencia (SAE) en Chile	10
2.1.1.1	Marco Institucional y Tecnológico del SAE	10
2.1.1.2	Base tecnológica: Cell Broadcast System (CBS)	11
2.1.1.3	Funcionamiento Operacional del SAE para Incendios Forestales	11
2.1.1.4	Operación del SAE en Chile	13
2.1.1.5	Relevancia del SAE para Incendios Forestales	14
2.1.1.6	El SAE como Soporte a la Toma de Decisiones Basados en Datos	14
2.2	Modelos de Aprendizaje Automático Random Forest y XGBoost	15
2.2.1	Random Forest (Bosque Aleatorio)	16
2.2.1.1	Fundamento Matemático: Impureza de Gini	16
2.2.1.2	Estrategia ante Desbalance: <code>class_weight</code>	17
2.2.1.3	Hiperparámetros Estructurales	18
2.2.2	XGBoost (Extreme Gradient Boosting)	18
2.2.2.1	Fundamento Matemático: Función Objetivo Regularizada	18
2.2.2.2	Estrategia ante Desbalance: <code>scale_pos_weight</code>	19
2.2.2.3	Tasa de Aprendizaje: <code>learning_rate</code>	19
2.2.2.4	Comparación Operativa: Sesgo y Varianza	20
2.2.2.5	Interpretabilidad y Jerarquía de Variables	20
2.3	Métricas de Evaluación de Desempeño	21



2.3.1	<i>Matriz de Confusión</i>	21
2.3.2	<i>Métricas Derivadas</i>	22
2.3.3	<i>Validación cruzada (Cross-Validation)</i>	22
2.3.4	<i>Análisis de Curvas de Rendimiento</i>	22
2.3.4.1	Curva ROC: Capacidad de Discriminación Global	23
2.3.4.2	Curva de Precision-Recall: Rendimiento en la Clase 1 (SAE)	23
2.3.4.3	Métrica Resumen: Average Precision (AP)	24
CAPÍTULO III: METODOLOGÍA		26
3.1	Diseño y Enfoque de la Investigación	26
3.2	Materiales y Herramientas Tecnológicas	26
3.2.1	<i>Lenguaje de Programación y Entorno</i>	27
3.2.2	<i>Librerías y Bibliotecas Especializadas</i>	27
3.2.3	<i>Herramientas de Sistemas de Información Geográficas (SIG)</i>	27
3.3	Población	28
3.4	Fase 1: Recolección e Integración de Datos (respuesta al OE1)	28
3.4.1	<i>Fuentes de Información</i>	28
3.4.2	<i>Procesamiento y Limpieza</i>	28
3.5	Análisis Exploratorio de los Datos: Características y Análisis Espacial (Respuesta al OE2).....	29
3.5.1	<i>Cruce Espacial (Spatial Join)</i>	29
3.5.2	<i>Categorización del Riesgo</i>	30
3.5.3	<i>Definición de la Variable Objetivo SAE</i>	30
3.5.4	<i>Cálculo del Déficit de Presión de Vapor (VPD)</i>	30



3.6	Estrategia de Reducir el Dataset Base.....	31
3.6.1	<i>Verificación de Consistencia Estadística del Subconjunto Reducido</i>	32
3.7	Codificación de Variables (Encoding).....	34
3.8	Modelo Predictivo: Random Forest (Respuesta al OE3).....	35
3.8.1	<i>Estrategia de Muestreo (Train/Test Split)</i>	35
3.8.2	<i>Manejo del Desbalance en Random Forest</i>	35
3.9	Modelo Predictivo: XGBoost y Análisis Comparativo (respuesta al OE4)	36
3.9.1	<i>Manejo del Desbalance en XGBoost</i>	36
3.10	<i>Evaluación y Validación</i>	37
3.10.1	<i>Métricas de Desempeño Orientadas a la Gestión del Riesgo de Desastres</i> .	37
3.10.2	<i>Análisis de Variables Determinantes</i>	37
3.10.3	<i>Validación de Estabilidad</i>	38
3.10.4	<i>Estrategia de Ajuste Operativo del Umbral de Decisión (Threshold Tuning)</i> .	38
CAPÍTULO IV: ANÁLISIS DE RESULTADOS		40
4.1	Análisis Exploratorio de Datos (EDA).....	40
4.1.1	Caracterización Estadística y Temporal.	40
4.1.2	Análisis Exploratorio Espacial de Incendios Forestales	44
4.1.3	Comportamiento Meteorológico	45
4.1.4	Coherencia del Mapa de Riesgo con la Activación de SAE	47
4.2	Evaluación Comparativa del Desempeño Predictivo	50
4.2.1	<i>Análisis de la Capacidad de Detección (Sensibilidad)</i>	51
4.2.2	<i>Análisis de Falsas Alarmas (Precisión)</i>	52
4.2.3	<i>Selección del Modelo Bajo Criterio de Riesgo Operacional</i>	53

4.2.4	<i>Análisis de Curvas de Rendimientos (ROC y Precisión-Recall)</i>	53
4.3	Validación de Estabilidad (Cross-Validation).....	55
4.4	Análisis de Importancia de Variables	56
4.5	Ajuste Operativo del Umbral de Decisión	60
4.5.1	<i>Análisis Visual del Intercambio de Métricas</i>	62
CAPÍTULO V: CONCLUSIONES.....		65
5.1	Verificación de Hipótesis y Conclusión General	65
5.2	Síntesis de Hallazgos	66
5.3	Conclusiones Generales	67
CAPÍTULO VI: RECOMENDACIONES		69
6.1	Implementación del Modelo Predictivo	69
6.1.1	Despliegue del modelo XGBoost.....	69
6.1.2	Configuración del Umbral Operativo	69
6.2	Gestión del Dato y Mejora Continua.....	70
6.2.1	Actualización de la Variable Riesgo	70
6.2.2	Retroalimentación Continua del Sistema	70
6.2.3	Profundización del Análisis Exploratorio Temporal	70
BIBLIOGRAFÍA.....		72
ANEXOS.....		77
Anexo A. Oficio N° 273/2026 – Servicio Nacional de Prevención y Respuesta ante Desastres.		77
Anexo B. Carta Oficial N°80/2026 – Corporación Nacional Forestal		79

ÍNDICE DE TABLAS

Tabla 1 Matriz de Responsabilidades: Protocolo SAE	13
Tabla 2 Tipos de Mensajes de Emergencia Emitidos por SENAPRED, según Idioma y Nivel de Riesgo	14
Tabla 3 Estadística Descriptiva de Variables Numéricas	42
Tabla 4 Reporte de Clasificación por Clase (Random Forest y XGBoost)	50
Tabla 5 Comportamiento de las Métricas.....	61

ÍNDICE DE FIGURAS

Figura 1 Protocolo de Activación: Alerta de Emergencia SAE.....	12
Figura 2 Fórmula Gini	17
Figura 3 Función Objetivo Regularizada.....	18
Figura 4 Fórmula Tasa de Falsos Positivos.....	23
Figura 5 Relación Precisión y Umbral de Decisión	24
Figura 6 Fórmula Déficit de Presión de Vapor	31
Figura 7 Gráfico Comparativo entre Dataset Original y Dataset Filtrado Condicional.....	32
Figura 8 Matriz de Correlación: Incendios Relevantes (>2.6 ha)	33
Figura 9 Representación Calibración del Umbral de Decisión.....	39
Figura 10 Gráfico de Total de Incendios vs. Alertas SAE (Escala Logarítmica).....	42
Figura 11 Gráfico de Tasa de Activación SAE (% de incendios que requirieron Alerta).....	43
Figura 12 Mapa de Incendios con Activación SAE en el Territorio Nacional	45
Figura 13 Gráfico de Distribución de Temperatura, Humedad y Velocidad del Viento.....	46
Figura 14 Gráfico de Distribución de Incendios v/s Activación SAE por Nivel de Riesgo	48
Figura 15 Validación: Probabilidad de SAE según Nivel de Riesgo CONAF	48
Figura 16 Matrices de Confusión Resultantes	51
Figura 17 Gráfico Comparativo: Curva ROC (capacidad de discriminación) y la Curva Precision-Recall (rendimiento en clase desbalanceada).....	54
Figura 18 XGBoost: Ranking de Variables Determinantes para Alerta SAE.....	57
Figura 19 Gráfico Análisis de Sensibilidad.....	63

INTRODUCCIÓN

La activación de la mensajería del Sistema de Alerta de Emergencias (SAE) frente a incendios forestales es responsabilidad de **SENAPRED**, basándose principalmente en información técnica proporcionada por **CONAF y Bomberos** mediante evaluaciones manuales y en terreno. Sin embargo, en escenarios críticos, esta información no siempre está disponible con la rapidez necesaria, lo que genera retrasos en la emisión de alertas y en la evacuación preventiva de la población. Actualmente, la gestión de emergencias carece de modelos predictivos que estandaricen y agilicen este proceso, manteniendo una dependencia exclusiva del análisis humano. Esta brecha técnica contraviene el espíritu de la **Ley 21.180 sobre Transformación Digital del Estado**, que insta a las instituciones a incorporar soluciones semi-automatizadas para una gestión pública más eficiente y oportuna.

Considerando este contexto, el potencial del *Machine Learning* para procesar grandes volúmenes de datos y fundamentar decisiones complejas constituye el motor de esta investigación. El estudio se centra en el entrenamiento de dos algoritmos basados en árboles de decisión, **Random Forest** y **XGBoost**, diseñados para proporcionar información pertinente que apoye a los equipos técnicos y profesionales en la determinación crítica de activar alertas ciudadanas ante eventos naturales.

Debido a que la activación de una alerta exige rapidez y fiabilidad, la implementación de modelos predictivos capaces de responder con agilidad se presenta como la solución óptima para las necesidades de **SENAPRED**. Esta herramienta busca fortalecer el soporte técnico en la gestión de la **Mensajería SAE**, especialmente ante incendios forestales, los cuales representan una amenaza creciente en Chile.

Al examinar el registro histórico de **CONAF** de los incendios forestales entre los periodos 2019 y 2025, se observa una dinámica de cambio preocupante que ejemplifica la magnitud de esta problemática. Se destaca que el periodo 2022-2023 fue el de mayor impacto, ya que la superficie devastada superó las 350.000 ha. con un total de 5.326

incendios. Esto evidencia, además, que la agresividad de los siniestros no siempre es proporcional a su cantidad. Por otra parte, en el periodo más reciente 2024-2025, se registra el número máximo de incendios de 5.442, acompañado de un incremento sostenido en la superficie afectada, correspondiente a 83.057,2 ha. respecto al año anterior.

Se observa también, un incremento en el uso del sistema de Alerta de Emergencia (SAE) como mecanismo de comunicación hacia la población para alertar sobre el riesgo que la puede afectar. De acuerdo con el informe estadístico del Servicio Nacional de Prevención y Respuesta ante Desastres (SENAPRED), entre los años 2016 y 2024 se han emitido más de 500 mensajes SAE asociados a incendios forestales, concentrándose un número significativo de estos en las temporadas más recientes. Antecedentes preliminares indican que esta tendencia se mantiene en los años posteriores, registrándose durante el año 2025 un total de 277 mensajes SAE por incendios forestales.

Este escenario evidencia una creciente demanda operativa sobre los protocolos de evacuación y la toma de decisiones para la activación del **Sistema de Alerta de Emergencias (SAE)**. Lo anterior refuerza la necesidad de implementar herramientas analíticas que respalden dicho proceso de manera objetiva, consistente y oportuna. Bajo esta premisa, la presente investigación aborda esta problemática mediante el desarrollo de una solución que optimice la gestión de alertas.

Para ello, se detalla el diseño, entrenamiento y validación de los algoritmos **Random Forest** y **XGBoost**, utilizando como insumo una base de datos histórica normalizada de incendios forestales. Esta base integra factores topográficos, niveles de riesgo y, fundamentalmente, variables meteorológicas críticas como temperatura, humedad y viento. La relevancia de estas últimas se sustenta en la literatura especializada a través de la regla empírica '30-30-30' (temperaturas > 30°C, humedad < 30% y vientos > 30 km/h), condiciones que incrementan significativamente la probabilidad de un comportamiento extremo del fuego y que influyen directamente en la evaluación técnica previa a la activación de una alerta. Al contrastar estas variables con el registro histórico,

el estudio permite determinar qué modelo ofrece un desempeño superior en términos de sensibilidad operativa, pertinencia y completitud, con el fin de optimizar la determinación final de activación de la mensajería SAE por parte de los equipos técnicos pertinentes.

CAPITULO I: ANTECEDENTES DEL PROBLEMA

1.1 Antecedentes de la Organización

La presente investigación se realizó desde el **Servicio Nacional de Prevención y Respuesta ante Desastres (SENAPRED)**. Este organismo técnico del Estado es el encargado de coordinar el Sistema Nacional de Prevención y Respuesta ante Desastres (**SINAPRED**). Como institución pública descentralizada, opera bajo un enfoque integral que busca proteger la vida de las personas y reducir los impactos de las emergencias en el país.

1.1.1 Evolución Histórica

La trayectoria de este organismo se remonta a **1974** con la creación de la Oficina Nacional de Emergencia (**ONEMI**), pionera en la coordinación de recursos frente a catástrofes. Sin embargo, ante la alta exposición de Chile a amenazas naturales y antrópicas, la **Ley 21.364** marcó un hito en 2021, transformando a la ONEMI en SENAPRED. Este salto institucional consolidó el paso de un modelo reactivo a uno basado en la **gestión integral del riesgo**.

1.1.2 Presencia y Propósito

Actualmente, SENAPRED cuenta con **16 Direcciones Regionales** distribuidas en todo el territorio nacional, trabajando de manera articulada con:

- Ministerios y servicios públicos.
- Gobiernos regionales y municipios.
- Comunidades y organismos privados.

El objetivo fundamental es promover una cultura preventiva y garantizar una gestión eficiente de los recursos para fortalecer la preparación y respuesta, minimizando los riesgos que enfrenta nuestro país. Según lo declarado en el sitio web institucional la Misión y la Visión de SENAPRED es la siguiente:



Misión

“Generar conciencia y acciones que reduzcan el riesgo de desastres en el país a través de planificar, coordinar, organizar, supervisar y asesorar al Sistema Nacional de Prevención y Respuesta ante Desastres (SINAPRED), mediante los instrumentos de coordinación y gestión vigentes, incorporando el enfoque de género en las fases de mitigación, preparación, respuesta y en la etapa de rehabilitación.”

Visión

“Ser una institución reconocida, robusta y confiable que lidere y promueva territorialmente el enfoque integral de gestión del riesgo de desastres para hacer de Chile un país más resiliente y sostenible.”

1.2 Formulación del Problema

La activación de la mensajería SAE (Sistema de Alerta de Emergencias) frente a incendios forestales es responsabilidad de SENAPRED, basándose principalmente en la información entregada por el organismo técnico a cargo de la amenaza, CONAF y Bomberos, mediante evaluaciones manuales y en terreno. Sin embargo, en diversos escenarios críticos, dicha información no siempre está disponible con la rapidez requerida, lo que genera retrasos en la emisión del mensaje SAE y, en consecuencia, en el inicio oportuno de la evacuación preventiva de la población expuesta.

Actualmente no se dispone de un modelo predictivo que complemente y apoye este proceso de decisión. La ausencia de herramientas semi-automatizadas dificulta la estandarización y agilización en la determinación de cuándo activar un SAE, aumentando la dependencia exclusiva de los análisis manuales.

En este contexto, la Ley N° 21.180 sobre Transformación Digital del Estado establece el deber de avanzar hacia sistemas de toma de decisiones más eficientes, incorporando soluciones semi-automatizadas que fortalezcan la gestión pública.

1.3 Alcance

El proyecto permitirá contar con una herramienta de apoyo a la decisión de enviar un mensaje SAE a la población en el proceso de evacuación de SENAPRED, optimizando la seguridad de la población y la confianza institucional. Técnicamente el alcance contempla modelos supervisados de clasificación binaria (se envía o no mensaje SAE), no se contempla implementación en tiempo real.

1.4 Delimitaciones

1.4.1 Delimitación Geográfica

El área de estudio comprende la totalidad del territorio nacional identificado bajo categorías de riesgo o vulnerabilidad ante la ocurrencia de incendios forestales.

1.4.2 Delimitación Poblacional

El universo de estudio comprende a los habitantes del territorio nacional ubicados en zonas de riesgo o áreas de influencia de incendios forestales.

1.4.3 Delimitación de Tiempo

El estudio contempla una ventana de observación desde el año 2019 hasta el 2025, enfocada en la recopilación y el procesamiento de datos históricos oficiales de SENAPRED y CONAF, correspondiente a pequeños focos hasta grandes incendios forestales ocurridos en el territorio nacional.

1.5 Limitaciones

Las limitaciones que contempla la investigación son las siguientes:

- No se desarrollará un sistema productivo operativo.
- Sólo se trabajará con bases de datos históricas y públicas (CONAF y SENAPRED).
- El foco estará en el prototipo analítico, no en la integración con plataformas de mensajería SAE.
- La validación del modelo se realizó dentro del mismo bloque temporal 2019-2025, lo que no permite evaluar el comportamiento del modelo ante temporadas completamente nuevas fuera del periodo de entrenamiento.
- Dado que los incendios con activación SAE presentan una concentración espacial significativa en la zona centro-sur del país, como se evidencia en la Figura 12, la división aleatoria del dataset podría introducir casos geográficamente similares en ambos conjuntos (entrenamiento y prueba), produciendo una estimación levemente optimista del desempeño. Una validación geográfica por macrozona o región habría sido metodológicamente más exigente. Esta condición se declara como una restricción del estudio.

1.6 Objetivos

1.6.1 *Objetivo General*

El objetivo general de la presente investigación consiste en: “Desarrollar y evaluar un modelo predictivo que utilice *Machine Learning* para estimar la probabilidad de emitir un mensaje de evacuación mediante el Sistema de Alerta de Emergencia (SAE) ante incendios forestales, integrando variables meteorológicas, topográficas, espaciales y categóricas, con el fin de apoyar la gestión de emergencias de SENAPRED”.

1.6.2 *Objetivos Específicos*

Con el fin de dar cumplimiento al objetivo general, la investigación contempla el desarrollo de los siguientes objetivos específicos:

OE1. Recolectar y limpiar datos históricos de incendios forestales y variables asociadas (mapa de riesgo de incendios forestales, meteorológicas, topográficas, espaciales).

OE2. Realizar un análisis exploratorio y generar nuevas características, a partir del mapa de riesgo de incendios forestales según las categorías: bajo, medio, alto y muy alto.

OE3. Entrenar y evaluar un modelo de Random Forest para la predicción de activación del SAE, analizando su desempeño mediante métricas de clasificación y matrices de confusión.

OE4. Entrenar y evaluar un modelo de XGBoost para la predicción de activación del SAE, comparando su efectividad técnica frente al modelo de Random Forest.

CAPÍTULO II: MARCO TEÓRICO

2.1 Sistema de Alerta de Emergencia (SAE)

El Sistema de Alerta de Emergencia (SAE) es un sistema de mensajería de emergencia, implementado por el Estado de Chile, cuyo objetivo es advertir e instruir a la población localizada en zonas de riesgo ante situaciones de emergencia, mediante el envío de mensajes directamente a los teléfonos móviles ubicados en el área afectada.

El SAE permite emitir alertas asociadas a múltiples amenazas, entre ellas:

- Incendios forestales
- Tsunami
- Remociones en masa
- Inundaciones
- Actividad volcánica

Lo anterior se alinea con la definición de los sistemas de alerta temprana multiamenaza promovidos por la Organización de las Naciones Unidas para la reducción del Riesgo de Desastres.

El sistema está diseñado para comunicar mensajes preventivos y de evacuación, orientados a la autoprotección de la población en función de la evolución de cada amenaza.

Este tipo de sistemas corresponde en la literatura internacional a lo que se denomina Public Warning Systems (PWS), es decir, infraestructura tecnológica diseñada para la difusión masiva y localizada de alertas tempranas, avisos y mensajes de emergencia dirigidos a la población.

2.1.1 Implementación del Sistema de Alerta de Emergencia (SAE) en Chile

La implementación del SAE en Chile surge como una respuesta directa al terremoto y tsunami del 27 de febrero de 2010, evento que evidenció fallas estructurales en los mecanismos de coordinación, generación de alarmas y comunicación hacia la población.

Según el documento oficial: “ Que producto del terremoto y posterior tsunami, acaecido el 27 de febrero de 2010, se pudo constatar la ausencia de coordinaciones adecuadas, entre los organismos de emergencia con los encargados de generar las alarmas de catástrofes, así como la falta de herramientas de comunicaciones robustas entre estos organismos y la inexistencia de un sistema de comunicación masivo para alertar a la población” (Decreto N° 60 del Ministerio de Transporte y Telecomunicaciones, Abril de 2012).

Este fenómeno ha sido ampliamente documentado en la literatura como un ejemplo de fallas en los sistemas de alerta temprana, donde la ausencia de comunicación oportuna incrementa el impacto de los desastres.

Como resultado, el Estado de Chile decidió implementar un sistema de alerta pública capaz de emitir mensajes masivos y georreferenciados, que no dependiera de registros previos ni de la congestión de las redes tradicionales.

2.1.1.1 Marco Institucional y Tecnológico del SAE

En 2011, la entonces Oficina Nacional de Emergencias (ONEMI) en conjunto con el Ministerio del Interior y Seguridad Pública y la Subsecretaría de Telecomunicaciones (SUBTEL), lideró la implementación del SAE mediante la creación de una plataforma de distribución de alerta desarrollada por una empresa israelí.

Esta plataforma se integró con las redes de todas las compañías de telecomunicaciones móviles del país permitiendo la difusión coordinada de las alertas a nivel nacional.

Los objetivos operaciones definidos para el SAE fueron:

- Difusión masiva y localizada de alertas
- Operación continua 24/7

- Capacidad de difusión a múltiples compañías de telefonía móvil

La integración del SAE con todas las compañías móviles del país responde al principio de interoperabilidad, considerado un requisito crítico en los Public Warning Systems modernos.

2.1.1.2 Base tecnológica: Cell Broadcast System (CBS)

El SAE utiliza la tecnología Cell Broadcast System (CBS), que corresponde a un servicio de mensajería unidireccional diseñada para enviar mensajes simultáneamente a todos los dispositivos móviles conectados a una red celular dentro de un área geográfica específica.

A diferencia de los SMS, el CBS:

- No se ve afectado por congestión de red
- No requiere números telefónicos
- No necesita registro previo
- No consume saldo ni datos del usuario
- Es gratuito para la población

Esto permite que un solo mensaje pueda ser recibido por millones de usuarios al mismo tiempo, lo que lo convierte en una infraestructura crítica para la gestión del riesgo de desastres.

La literatura técnica establece que el CBS es el mecanismo más robusto para alertamiento público porque no depende de congestión de red ni de bases de datos de usuarios, por ello el CBS es tecnología base de sistemas como WEA (EE. UU), EU-Alert y J-Alert.

2.1.1.3 Funcionamiento Operacional del SAE para Incendios Forestales

El proceso de activación del SAE sigue una cadena de decisiones institucionales, las que se detallan en la siguiente figura:

Figura 1

Protocolo de Activación: Alerta de Emergencia SAE



Nota. Figura elaborada con la herramienta NotebookLM para una mejor representación de la información obtenida del Manual del Sistema de Alerta de Emergencias: Monitoreo, alertamiento y producción de información operacional (Código: MO-MAC-02; Versión 07). Ministerio del Interior y Seguridad Pública, Gobierno de Chile.

Este flujo, representado en la Figura 1, confirma que el SAE es el eslabón final del Sistema Nacional de Alerta Temprana, materializando la decisión de alertar en una comunicación directa a la población.

En la Tabla 1, se puede apreciar el detalle de las responsabilidades a nivel de rol e institución que contempla el protocolo de SAE, evidenciando los actores multiinstitucionales involucrados y la función que cumple cada uno de ellos.

Tabla 1

Matriz de Responsabilidades: Protocolo SAE

Actor / Institución	Nivel	Responsabilidad Principal	Función Clave
CONAF / Bomberos	Técnico / Terreno	Detección	Identificar la amenaza crítica y solicitar la evacuación.
Director/a Regional SENAPRED	Regional	Decisión Estratégica	Evaluar si la magnitud de la emergencia amerita el uso del SAE.
Profesional SAE	Operativo	Configuración	Redactar el mensaje y delimitar el polígono de cobertura (Georreferenciación).
Jefe UAT Nacional	Nacional	Validación	Control de calidad final y autorización del envío masivo.
Compañías de Telecomunicaciones	Externo	Difusión	Transmitir la señal de alerta a las antenas del sector afectado.

Nota. Tabla elaborada con la información obtenida del *Manual del Sistema de Alerta de Emergencias: Monitoreo, alertamiento y producción de información operacional* (Código: MO-MAC-02; Versión 07). Ministerio del Interior y Seguridad Pública, Gobierno de Chile.

2.1.1.4 Operación del SAE en Chile

El SAE está operativo desde marzo de 2017 y es utilizado por todas las compañías móviles del país. La infraestructura y conectividad del sistema es supervisada por SUBTEL, que garantiza la interoperabilidad entre las redes móviles y la plataforma SAE.

El sistema cuenta con dos tipos de mensajes para incendios forestales:

- Mensajes preventivos: se utilizan para informar con anticipación que se prevé que una amenaza pueda afectar a una comunidad.
- Mensajes de amenaza inminente: se utiliza para complementar y reforzar el proceso de evacuación.

Tabla 2

Tipos de Mensajes de Emergencia Emitidos por SENAPRED, según Idioma y Nivel de Riesgo

Idioma	Tipo	Mensaje
Español	Mensaje preventivo	SENAPRED: Incendio forestal en [Nombre de la comuna]. Mira a tu alrededor e informate.
Español	Amenaza inminente	SENAPRED: EVACUE [nombre sector] por incendio forestal [nombre comuna].
Inglés	Mensaje preventivo	SENAPRED: Wildfire in your area. Be aware of your location and be informed.
Inglés	Amenaza inminente	SENAPRED: EVACUATE [nombre sector] due to wildfire.

Nota. Tabla extraída del Manual del Sistema de Alerta de Emergencias: Monitoreo, alertamiento y producción de información operacional (Código: MO-MAC-02; Versión 07). Ministerio del Interior y Seguridad Pública, Gobierno de Chile.

La diferencia entre mensajes preventivos y de amenaza inminente coincide con la teoría de escalamiento de alertas o niveles, que busca reducir falsas alarmas y mejorar la reacción de la población, respecto a la entrega de información graduada según la amenaza.

2.1.1.5 Relevancia del SAE para Incendios Forestales

En el contexto de los incendios forestales, el SAE constituye la principal herramienta de alerta temprana a la ciudadanía, realizando evacuaciones ante la amenaza de propagación del fuego hacia zonas pobladas.

La activación del mensaje SAE por amenaza inminente dependerá de la evaluación de la amenaza que realice el mando en terreno. Este mando es presidido por la Corporación Nacional Forestal (CONAF) o por Bomberos de Chile, en caso de ausencia de CONAF.

2.1.1.6 El SAE como Soporte a la Toma de Decisiones Basados en Datos

Dado que la activación del SAE constituye una decisión crítica, su uso implica evaluar múltiples variables técnicas y operacionales. En este contexto, la predicción de la mensajería SAE mediante modelos de *Machine Learning* no busca automatizar ni sustituir la decisión del organismo técnico, sino apoyarla mediante una segunda opinión técnica

basada en datos históricos. El modelo aprende de patrones observados en eventos pasados, como características del incendio, condiciones meteorológicas, ubicación y mapa de riesgo de CONAF, para estimar la probabilidad de que una situación similar haya requerido el envío de un SAE en el pasado. Desde la perspectiva de los sistemas de apoyo a la decisión, los modelos de *Machine Learning* permiten aprender patrones históricos para apoyar decisiones complejas.

De este modo, el modelo funciona como un sistema de apoyo a la decisión entregando a las autoridades de SENAPRED un antecedente cuantitativo adicional que complementa el juicio experto, los protocolos y la información en tiempo real. Esta aproximación permite:

- Reducir la variabilidad subjetiva en la evaluación del riesgo
- Fortalecer la trazabilidad de las decisiones
- Mejorar la oportunidad de las evacuaciones

Así, la predicción de la mensajería SAE se inserta como una herramienta de apoyo dentro del proceso institucional de evaluación y decisión, contribuyendo a una gestión del riesgo más informada, transparente y basada en evidencia. En este marco, el modelo entrega una probabilidad de que un incendio forestal requiere la activación de un SAE, actuando como segunda opinión técnica que complementa el juicio experto de la autoridad. Este enfoque se alinea con el principio de “IA para apoyo a la decisión” promovido por la Política Nacional de Inteligencia Artificial de Chile (2024), que establece que los sistemas algorítmicos deben fortalecer y no reemplazar la toma de decisiones.

2.2 Modelos de Aprendizaje Automático Random Forest y XGBoost

Este trabajo se fundamenta en métodos de aprendizaje por conjuntos, específicamente en arquitecturas basadas en árboles de decisión. Según Géron (2019), la premisa de estos métodos es que “un grupo de predictores a menudo obtendrá mejores resultados que el mejor predictor individual” (p.189), reduciendo el riesgo de sesgo y varianza inherente a los modelos simples.

2.2.1 Random Forest (Bosque Aleatorio)

El algoritmo Random Forest se basa en la técnica de *bagging*, un método de aprendizaje por conjuntos cuyo objetivo es reducir la varianza de modelos inestables mediante la combinación de múltiples predictores entrenados sobre distintas muestras del conjunto de datos.

El *bagging* consiste en generar múltiples subconjuntos aleatorios de entrenamiento a partir de muestreo con reemplazo. Cada subconjunto es utilizado para entrenar un modelo base independiente, y posteriormente sus predicciones se combinan mediante votación mayoritaria (en clasificación) o promedio (en regresión). Este procedimiento permite estabilizar modelos que, por lo general, son sensibles a pequeñas variaciones en los datos. En el caso de Random Forest el modelo base corresponde a un Árbol de Decisión. (Géron, 2019, p.192).

Un Árbol de Decisión es un modelo predictivo estructurado en forma jerárquica, compuesto por nodos (nodo raíz, nodo terminal y nodo hijo). En cada nodo, el algoritmo selecciona una variable y un punto de corte que maximiza la separación entre clases, con el objetivo de reducir la impureza del conjunto de datos. Este proceso se repite de manera recursiva hasta cumplir un criterio de detención.

Los Árboles de Decisión suelen presentar alta varianza, pequeñas modificaciones en los datos pueden generar estructuras completamente distintas. Random Forest mitiga esta inestabilidad construyendo múltiples árboles en paralelo y combinando sus predicciones.

2.2.1.1 Fundamento Matemático: Impureza de Gini

La construcción de cada árbol se rige por la minimización de la impureza en los nodos. En este trabajo se utiliza el índice de Impureza de Gini, que mide la probabilidad de clasificar incorrectamente un elemento aleatorio si se etiquetara según la distribución de las clases en el nodo (Fernández-Avilés & Montero, 2024).



Figura 2

Fórmula Gini

$$Gini(\mathbf{X}) = 1 - \sum_{i=1}^{\kappa} p_i^2,$$

Nota. La impureza de Gini para X se define como: Siendo X un conjunto de datos con k

clases y sea p_i la probabilidad de que una observación pertenezca a la clase $i = 1, \dots, k$. De “Árboles de clasificación y regresión” por Carrasco, R. A., Bueno, I., y Montero, J.-M., en G. Fernández-Avilés y J.-M. Montero (Eds.), 2024, *Fundamentos de ciencia de datos con R*, 2024, McGraw-Hill. CC BY-NC-ND.

Donde en este caso al ser la respuesta binaria, $K = 2$. El algoritmo selecciona iterativamente la división que maximiza la reducción de esta impureza, buscando que las hojas finales del árbol sean lo más homogéneas posible respecto a la clase objetivo (SAE).

En resumen, la impureza de Gini mide el grado de desorden o mezcla en un grupo de datos. Con $Gini = 0$ (pureza total) el nodo es puro, es decir, sólo contiene elementos pertenecientes a un único grupo. Por ejemplo, si en una hoja del árbol quedaron 50 incendios y los 50 son SAE, no hay mezcla. Mientras que un $Gini = 0.5$ (máxima impureza) el nodo es un desorden, ejemplo en el nodo hay 50 incendios; 25 son SAE y 25 no son SAE.

2.2.1.2 Estrategia ante Desbalance: `class_weight`

Dado el severo desbalance de clases inherente a la predicción de incendios forestales (donde los eventos con SAE son minoritarios), el entrenamiento estándar tiende a sesgarse hacia la clase mayoritaria. Para mitigar esto, *Random Forest* implementa el

hiperparámetro `class_weight = balanced`. Este mecanismo ajusta los pesos de cada clase de forma inversamente proporcional a su frecuencia en los datos. Matemáticamente, el algoritmo penaliza el error de clasificación en la clase minoritaria con un costo mayor, forzando a los árboles a priorizar la correcta identificación de las alertas SAE durante la fase de división de nodos.

2.2.1.3 Hiperparámetros Estructurales

- **n_estimators:** Define el número de árboles en el bosque. Un mayor número aumenta la estabilidad de la votación final. Sin embargo, un valor excesivamente alto incrementa el costo computacional, sin necesariamente mejorar el desempeño predictivo.
- **max_depth:** Controla la profundidad máxima de los árboles actuando como un mecanismo de regularización para evitar que el modelo memorice el ruido de los datos (sobreajuste o overfitting).

2.2.2 XGBoost (*Extreme Gradient Boosting*)

XGBoost es una implementación optimizada y escalable del algoritmo *Gradient Boosting*. A diferencia del enfoque paralelo de *Random Forest*, XGBoost opera de manera secuencial y cada nuevo modelo se entrena para corregir los errores residuales (la diferencia entre lo predicho y lo real) de los modelos anteriores. Géron (2019).

2.2.2.1 Fundamento Matemático: Función Objetivo Regularizada

La superioridad de XGBoost en competencias de ciencia de datos radica en su función objetivo, la cual combina la función de pérdida con un término de regularización para controlar la complejidad expresado por árbol:

Figura 3

Función Objetivo Regularizada



$$obj(\theta) = L(\theta) + \Omega(\theta)$$

Nota. La función objetivo de XGBoost combina una función de pérdida $L(\theta)$ que mide el error de predicción del modelo, con un término de regularización $\Omega(\theta)$ que penaliza la complejidad del árbol para evitar sobre ajuste. De "XGBoost: A scalable tree boosting system" por Chen, T., & Guestrin, C., 2016, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794 (<https://doi.org/10.1145/2939672.2939785>). Derechos reservados 2016 para Association for Computing Machinery

Donde $L(\theta)$ es la función de pérdida (LogLoss para clasificación binaria) y $\Omega(\theta)$ penaliza la complejidad del modelo (basada en el número de hojas y sus pesos). Como señala Chen y Guestrin (2016), esta regularización es fundamental para prevenir los sobreajustes permitiendo que el modelo generalice mejor en escenarios con ruido.

2.2.2.2 Estrategia ante Desbalance: `scale_pos_weight`

Para maximizar la detección de eventos críticos, XGBoost utiliza el hiperparámetro `scale_pos_weight`. Este parámetro modifica la función de pérdida asignando un peso multiplicativo al error de la clase = 1 (SAE). Al configurar este valor (típicamente como la razón entre casos negativos y positivos) se logra que un Falso Negativo (no se alerta un incendio que si debería tener alerta) tenga un impacto cuantitativo mucho mayor en la función de costo que un Falso Positivo. Esto orienta el descenso por el gradiente hacia la optimización de la Sensibilidad (Recall), alineando el modelo con el objetivo de seguridad pública de SENAPRED. (XGBoost developers, 2025).

2.2.2.3 Tasa de Aprendizaje: `learning_rate`

Conocido también como η (eta), este hiperparámetro escala la contribución de cada nuevo árbol al modelo global. Utilizar una tasa de aprendizaje baja (ejemplo < 0.05) en combinación con un mayor número de árboles permite un proceso de optimización más robusto, reduciendo el riesgo de caer en mínimos locales. (XGBoost developers, 2025).

2.2.2.4 Comparación Operativa: Sesgo y Varianza

Desde una perspectiva teórica, la evaluación conjunta de ambos modelos se justifica por sus enfoques complementarios:

- **Random Forest** se enfoca en la reducción de la varianza: crea estabilidad promediando múltiples modelos independientes propensos al sobreajuste.
- **XGBoost** se enfoca en la reducción del sesgo: refina iterativamente modelos simples para capturar patrones complejos.

Entonces, el término sesgo hace referencia a cuánto se alejan en promedio las predicciones de un modelo respecto a los valores reales, y el término varianza hace referencia a cuánto cambia el modelo dependiendo de los datos utilizados en su entrenamiento. (Amat Rodrigo, 2020).

2.2.2.5 Interpretabilidad y Jerarquía de Variables

1. Random Forest: Capacidad de Segregación

En Random Forest la importancia de una variable se define por su capacidad para ordenar los datos. Scikit-Learn implementa esto mediante la Disminución Media de la Impureza (*Mean Decrease Impurity*).

El algoritmo mide qué tan bien una variable (ejemplo velocidad del viento) logra separar los incendios en grupos homogéneos o “puros” Clase SAE versus Clase no SAE.

Si al usar una variable para dividir un nodo, la mezcla de clases disminuye drásticamente, esa variable recibe una puntuación alta. La importancia final es el promedio de cuánto redujo la impureza esa variable en todos los árboles del bosque (Géron, 2019, p. 198). En resumen, premia a las variables que mejor **clasifican estructuralmente** los datos.



2. XGBoost: Eficiencia Predictiva

XGBoost utiliza una métrica distinta llamada **Ganancia** (*Gain*), la cual está alineada con su naturaleza de corrección de errores. La Ganancia mide la mejora relativa en la precisión del modelo al utilizar una variable específica. No solo busca separar grupos, sino minimizar la función de pérdida global del modelo.

Una variable tiene alta importancia en XGBoost si su inclusión en el árbol resulta en una reducción significativa del error de predicción. Es una métrica de **eficiencia**, el modelo privilegia aquellas variables que resuelven el problema de forma más directa, descartando aquellas que aportan información redundante (Chen & Guestrin, 2016).

2.3 Métricas de Evaluación de Desempeño

En el contexto de la gestión del riesgo de desastres, la exactitud global (Accuracy) es una métrica engañosa debido al desbalance de las clases. Por ello, la evaluación se centra en métricas específicas de la matriz de confusión.

2.3.1 Matriz de Confusión

Herramienta fundamental que desglosa el rendimiento de clasificador en cuatro categorías:

- **Verdaderos Positivos (VP):** incendios con SAE correctamente predichos.
- **Falsos Negativos (FN):** incendios con SAE omitidos por el modelo (error crítico).
- **Falsos Positivos (FP):** incendios que no correspondía clasificarlos como SAE.
- **Verdaderos Negativos (VN):** incendios sin SAE correctamente clasificadas.

2.3.2 Métricas Derivadas

- Sensibilidad (Recall): $VP/VP+FN$ Mide la capacidad del sistema para detectar la totalidad de los incendios con SAE. Es la métrica prioritaria en nuestro estudio.
- Precisión: $VP/VP+FP$ Mide la fiabilidad de los SAE.
- F1-Score: Media armónica entre Sensibilidad y Precisión, útil para comparar modelos de forma balanceada.

2.3.3 Validación cruzada (Cross-Validation)

Para asegurar la robustez estadística de los resultados, se aplica la validación cruzada k-fold. Esta técnica divide los datos en K subconjuntos, entrenando y validando el modelo iterativamente en diferentes particiones. Esto garantiza que las métricas de rendimiento reportadas no sean producto de un sesgo de selección de los datos de prueba (Géron, 2019, p.73).

La validación cruzada no solo proporciona una estimación del desempeño promedio del modelo, sino que ofrece una métrica de la precisión de dicha estimación a través de su desviación estándar. Una desviación elevada es indicativa de que el modelo es sensible al ruido de los datos de entrenamiento (sobreajuste) reduciendo su confiabilidad para inferencias futuras.

Basado en este principio se estableció como criterio de aceptación una desviación estándar inferior al 5% ($SD < 0.05$). Este umbral asegura que el margen de incertidumbre del modelo se mantenga acotado, garantizando que la capacidad predictiva de la mensajería SAE sea consistente independiente del subconjunto de datos históricos utilizados.

2.3.4 Análisis de Curvas de Rendimiento

Más allá de las métricas puntuales como un solo valor *Recall* o Precisión, es fundamental evaluar cómo se comporta el modelo entre diferentes umbrales de decisión. Para ello se utilizan dos herramientas gráficas complementarias.

2.3.4.1 Curva ROC: Capacidad de Discriminación Global

La Curva ROC (*Receiver Operating Characteristic*) muestra la capacidad de diagnóstico de un sistema clasificador binario al variar su umbral de discriminación. (Géron, 2019, p.97).

Ejes: En el eje Y Gráfica la tasa de Verdaderos Positivos (Sensibilidad) y en eje X la tasa de Falsos Positivos (1-Especificidad).

Interpretación de (AUC): La métrica derivada es el área bajo la curva (AUC). Un AUC de 0.5 indica un desempeño aleatorio, mientras que un AUC de 1 representa una discriminación perfecta. En el contexto de este trabajo, el AUC permite medir qué tan bien el modelo distingue, en términos globales, entre un incendio que requiere SAE y otro que no. (Géron, 2019 página 93).

2.3.4.2 Curva de Precision-Recall: Rendimiento en la Clase 1 (SAE)

Si bien la curva ROC es estándar, puede ser engañosa cuando existe un fuerte desbalance de clases. Esto ocurre porque el eje X de la curva ROC (Tasa de Falsos Positivos = TFP) incorpora en su cálculo los Verdaderos Negativos, siendo:

Figura 4

Fórmula Tasa de Falsos Positivos

$$TFP = \frac{FP}{FP+VN}$$

Nota. La tasa de falsos positivos (TFP) corresponde a la proporción de casos negativos clasificados incorrectamente como positivos, calculada como $FP/(FP+VN)$. De *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*, por Géron, A. 2019. Derechos reservados 2019 para Kiwisoft S.A.S.

En escenarios como el de esta investigación, donde la clase “incendios sin SAE” es mayoritaria (1.226 con SAE y sin SAE 5.117), la gran cantidad de Verdaderos Negativos tiende a diluir el impacto de los Falsos Positivos, manteniendo la curva ROC considerablemente alta.

Por esta razón se prioriza el análisis de la Curva Precision-Recall (PR), ya que las fórmulas de Precisión y Recall no incluyen a los Verdaderos Negativos en sus ecuaciones. Según Géron (2019), la curva PR es la herramienta idónea cuando la clase positiva es menor. Al excluir la clase mayoritaria de la ecuación, esta curva revela el desempeño real del modelo al identificar eventos críticos (SAE).

2.3.4.3 Métrica Resumen: Average Precision (AP)

La evaluación de la curva Precision-Recall puede ser subjetiva en ocasiones. Para cuantificar la calidad global del modelo en un solo indicador numérico, se utiliza el *Average Precision (AP)*.

Comparar clasificadores basándose únicamente en curvas visuales puede ser impreciso, por lo que utilizar el área bajo la curva proporciona una métrica objetiva para determinar la calidad de un modelo sobre otro.

En esta investigación, se utiliza la implementación de la librería Scikit-Learn, la cual aproxima esta área mediante una suma ponderada de precisiones en cada umbral de decisión.

Figura 5

Relación Precisión y Umbral de Decisión

$$AP = \sum_n (R_n - R_{n-1}) P_n$$

Nota. El Average Precision (AP) estima el área bajo la curva Precision-Recall. De *sklearn.metrics.average_precision_score*, por Scikit-learn developers, s.f., Scikit-learn (<https://scikit-learn.org>)

learn.org/stable/modules/generated/sklearn.metrics.average_precision_score.html). CC BY 3.0.

Donde P_n y R_n son la precisión y el recall en el n-ésimo umbral de decisión

El AP estima el área bajo la curva Precision-Recall. Un valor de 1.0 indica un modelo perfecto. Dado que el objetivo es maximizar la detección de incendios (Recall) manteniendo una precisión operativa aceptable, el AP ofrece una medida de la estabilidad del clasificador. Un AP alto confirma qué modelo es más consistente y fiable para tomar decisiones.

CAPÍTULO III: METODOLOGÍA

3.1 Diseño y Enfoque de la Investigación

En concordancia con el objetivo general de “desarrollar y evaluar un modelo predictivo” para el apoyo a la gestión de emergencias, la presente investigación adopta un enfoque cuantitativo, no experimental y transversal.

- **Enfoque cuantitativo:** se fundamenta en el procesamiento estadístico y matemático de datos históricos de los incendios forestales (periodo 2019-2025) para establecer patrones de probabilidad objetiva, utilizando algoritmos de *Machine Learning* como instrumentos de medición y predicción.
- **Alcance correlacional predictivo:** La investigación busca determinar la relación funcional entre un conjunto de variables independientes (meteorológicas, topográficas, mapa de riesgo y variables propias de la caracterización del incendio forestal) y una variable dependiente binaria (la emisión de un mensaje SAE), para luego construir un modelo capaz de apoyar en la toma de decisión de enviar un SAE.
- **Diseño No experimental:** Se analizan los fenómenos en su contexto natural sin manipular las condiciones físicas de los incendios forestales, observando los datos en un corte temporal retrospectivo para entrenar y evaluar los modelos *Random Forest* y *XGBoost*.

La unidad de análisis corresponde a eventos individuales de incendios forestales registrados, caracterizados por su contexto espacial, meteorológico y operacional al momento de ocurrencia.

3.2 Materiales y Herramientas Tecnológicas

Para el desarrollo de la solución, procesamiento de datos y entrenamiento de modelos se seleccionaron las siguientes herramientas de software y hardware, justificadas por su idoneidad técnica:

3.2.1 Lenguaje de Programación y Entorno

- **Python:** Seleccionado como lenguaje principal debido a su importancia en el ecosistema de Ciencia de Datos. Su amplia colección de librerías permite integrar procesos de limpieza, análisis espacial y modelado predictivo en un solo flujo de trabajo.
- **Google Colab:** Utilizado como entorno de desarrollo integrado (IDE) en la nube. Se justifica por su capacidad de cómputo escalable (acceso a GPU/TPU) necesaria para el entrenamiento de algoritmos de *Machine Learning*.

3.2.2 Librerías y Bibliotecas Especializadas

- **Pandas y NumPy:** Para la manipulación eficiente de estructuras de datos tabulares y cálculos. Se utilizó específicamente Pandas para la transformación de variables categóricas mediante Codificación Dummy (*Dummy Encoding*).
- **Scikit-Learn:** Utilizada para la implementación de algoritmos *Random Forest*, preprocesamiento de datos y métricas de evaluación.
- **XGBoost (Extreme Gradient Boosting):** Librería específica seleccionada por su eficiencia en la implementación de árboles de decisión optimizados por gradiente, superior en velocidad y rendimiento para datos desbalanceados.
- **Seaborn y Matplotlib:** Herramientas de visualización utilizadas para el análisis exploratorio y la generación de gráficos.

3.2.3 Herramientas de Sistemas de Información Geográficas (SIG)

- **ArcGIS Pro:** Software utilizado para el manejo de datos vectoriales y la realización de cruce espacial (*Spatial Join*) entre las coordenadas de los incendios forestales y las capas del Mapa de Riesgo de CONAF.

3.3 Población

La población objetivo corresponde a la totalidad de los eventos de incendios forestales registrados en la base de datos histórica de CONAF y SENAPRED durante el periodo 2019-2025. Esta población abarca eventos de todas las magnitudes, desde pequeños focos hasta grandes incendios forestales.

3.4 Fase 1: Recolección e Integración de Datos (respuesta al OE1)

Dando cumplimiento al objetivo específico 1 (Recolectar y limpiar datos...), esta fase se centró en la consolidación de un dataset histórico (periodo 2019-2025) a partir de fuentes heterogéneas.

3.4.1 Fuentes de Información

Se integraron tres dimensiones de datos fundamentales:

1. **Registro histórico de incendios forestales:** Base de datos de ocurrencia de incendios forestales de CONAF que incluye, el nombre del incendio, localización, combustible inicial, además de condiciones meteorológicas (temperatura, humedad relativa, nubosidad, velocidad y dirección del viento) y características topográficas (pendiente, exposición, topografía), obtenidos mediante solicitud oficial a CONAF.
2. **Variable SAE:** Datos históricos sobre el envío de mensajes de evacuación asociados a incendios forestales.
3. **Cartografía de Mapa de Riesgo:** Capas vectoriales (polígonos) del “Mapa de Riesgo” de CONAF, considerado los niveles de amenaza territorial.

3.4.2 Procesamiento y Limpieza

Se aplicaron técnicas de depuración para garantizar la calidad de los inputs de entrada:

- **Integridad de Datos:** Tras el análisis exploratorio, no se detectaron valores nulos críticos, por lo que se mantuvo la integridad de los registros del dataset base.
- **Normalización Temporal:** Se estandarizaron los formatos de fecha generando variables derivadas como mes de ocurrencia para capturar la estacionalidad de los incendios

3.5 Análisis Exploratorio de los Datos: Características y Análisis Espacial (Respuesta al OE2)

Como etapa previa al entrenamiento de los modelos predictivos, se realizó un Análisis Exploratorio de Datos (EDA) con el propósito de examinar la estructura general del conjunto de datos y evaluar la consistencia interna, estadística y física de las variables consideradas.

Este proceso comprendió la revisión de distribuciones, la identificación de posibles valores atípicos, el análisis preliminar de relaciones entre variables, la caracterización estadística y temporal de los registros, así como un análisis exploratorio espacial de los incendios forestales y del comportamiento de las variables meteorológicas.

En paralelo al análisis exploratorio de los datos, se procedió a la generación de nuevas características, y en particular se implementaron técnicas de análisis geoespacial para enriquecer la capacidad predictiva del modelo.

3.5.1 Cruce Espacial (*Spatial Join*)

Dado que el dataset histórico de incendios forestales no contiene el nivel de riesgo del terreno, se realizó un procedimiento de interacción espacial. Utilizando las coordenadas geográficas (latitud/longitud) de cada incendio forestal, se extrajo la información subyacente de las capas poligonales del mapa de riesgo, a través del software ArcGIS Pro.

3.5.2 Categorización del Riesgo

A partir del cruce espacial, se generó una nueva variable numérica que clasifica cada incendio según la amenaza del territorio en cuatro niveles: 1 = Bajo, 2 = Medio, 3 = Alto y 4 = Muy Alto. Esta variable sintetiza la probabilidad de daños sociales, ambientales y económicos que provoca un incendio en el territorio.

3.5.3 Definición de la Variable Objetivo SAE

A partir del registro de SENAPRED, se etiquetó cada incendio forestal del dataset histórico mediante un cruce manual y validando incendio por incendio.

Para resolver casos con potencial ambigüedad en el emparejamiento, se aplicaron los siguientes criterios: (1) coincidencia temporal entre el registro del incendio de CONAF y la emisión del mensaje SAE en SENAPRED, (2) coincidencia geográfica a nivel de comuna y (3) coherencia entre el nombre del incendio registrado por CONAF y la descripción del sector en el mensaje SAE.

- **Variable Dependiente:** Activación del Sistema de Alerta de Emergencia (SAE).
- **Variable Binaria:** 0 = no se envió SAE y 1 = sí se envió SAE.

3.5.4 Cálculo del Déficit de Presión de Vapor (VPD)

Se calculó una nueva variable denominada Déficit de Presión de Vapor (VPD) a partir de la temperatura y la humedad relativa del dataset base. El VPD mide la “sed” atmosférica o la capacidad del aire para secar el combustible vegetal, siendo un indicador físico más robusto que la humedad relativa por sí sola.

El cálculo se realizó respecto de la metodología estándar propuesta por la *Food and Agriculture Organization of the United Nations* – FAO.

Figura 6

Fórmula Déficit de Presión de Vapor

$$VPD = e_s \cdot \left(1 - \frac{RH}{100}\right)$$

Donde:

- e_s : Presión de vapor de saturación, calculada como $0.6108 \cdot \exp\left(\frac{17.27 \cdot T}{T+237.3}\right)$.
- RH : Humedad Relativa (%).

Nota. El VPD representa la diferencia entre presión de vapor de saturación y la presión real del vapor de agua en la atmósfera. Adaptado de *Crop evapotranspiration: Guidelines for computing crop water requirements*, por R. G. Allen, L. S. Pereira, D. Raes y M. Smith, 1998. Derechos reservados 1998 para FAO.

3.6 Estrategia de Reducir el Dataset Base

Debido al severo desbalance de clases original (3.99% registros con SAE) y a la presencia masiva de micro incendios que no constituyen una amenaza operativa significativa, se implementó una estrategia de filtrado condicional.

A diferencia de un corte arbitrario, se aplicó la siguiente regla lógica de inclusión para conformar el dataset final (D_{final}):

$$x \in D_{final} \Leftrightarrow (Superficie \geq 2.66ha) \vee (SAE = 1)$$

El umbral de 2.66 hectáreas se estableció estadísticamente basado en el percentil 25 histórico de los incendios que tuvieron SAE. Los resultados de esta estrategia fueron:

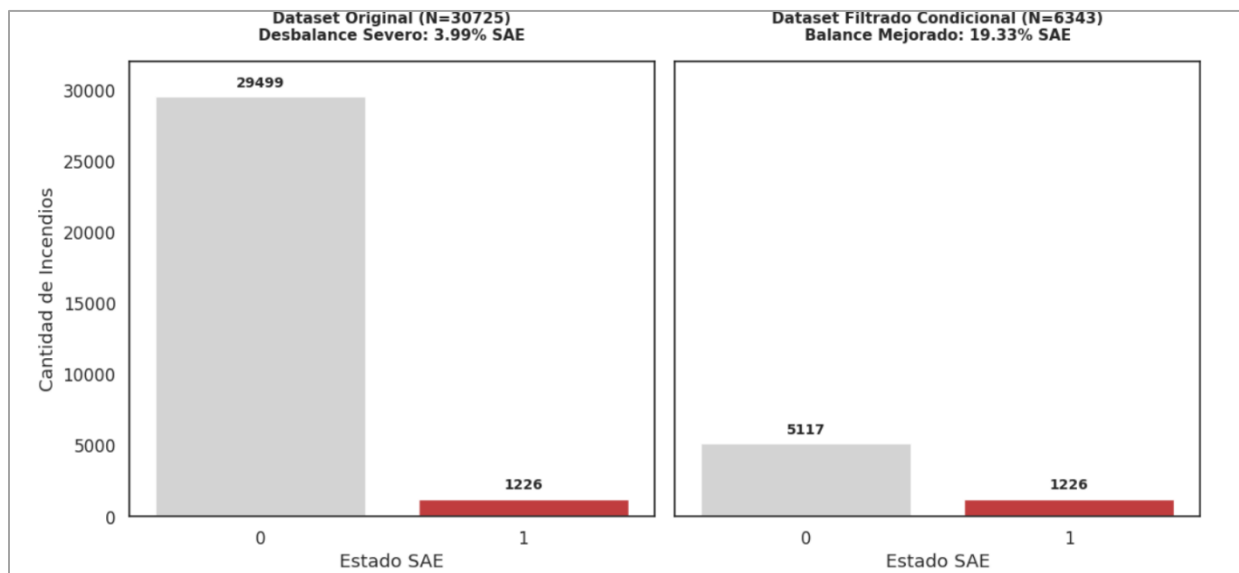
- Se mantuvo el 100% de los incendios con SAE (N=1.226), garantizando que el modelo aprenda de todos los casos históricos de alerta, independiente de la superficie quemada del incendio.
- Se eliminaron 24.382 eventos de la clase mayoritaria (Clase 0) que correspondientes a superficie quemada menor.

- Se mejoró el balance entre clases, la prevalencia de la clase objetivo aumentó del 3.99% al 19.33% creando un espacio de búsqueda mucho más denso para los algoritmos (*Random Forest* y *XGBoost*), como se muestra en la siguiente imagen.

Es importante precisar que todas las métricas reportadas en este estudio son válidas para el dataset de filtrado 6.343 registros y no deben interpretarse como métricas del universo completo de 30.725 incendios forestales registrados en periodo 2019-2025. El filtrado condicional modifica la composición del conjunto de análisis, por lo que los resultados corresponden al subconjunto de incendios operativamente relevantes para la decisión SAE.

Figura 7

Gráfico Comparativo entre Dataset Original y Dataset Filtrado Condicional



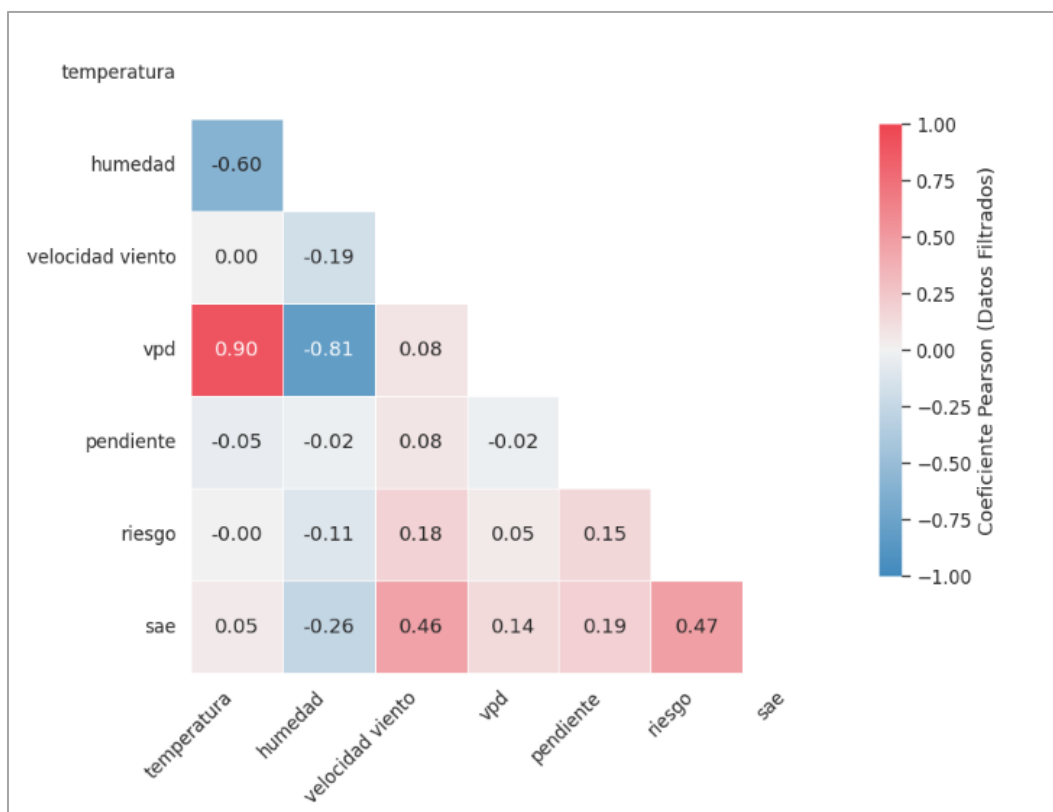
3.6.1 Verificación de Consistencia Estadística del Subconjunto Reducido

Dado que la aplicación de criterios de reducción del *dataset* modifica la distribución original de los datos y la proporción entre incendios con y sin activación SAE, se realizó un análisis exploratorio adicional sobre el conjunto el subconjunto final utilizado para el entrenamiento de los modelos. Este análisis tuvo como objetivo verificar que las

correlaciones entre variables predictoras se mantuvieran coherentes desde el punto de vista físico y operativo, y que el proceso de filtrado no introdujera distorsiones estructurales en los datos.

Figura 8

Matriz de Correlación: Incendios Relevantes (>2.6 ha)



La Figura 8, presenta la matriz de correlación de Pearson calculada sobre el conjunto de incendios relevantes (superficie afectada mayor a 2.6 ha). Se observa que las relaciones esperadas entre variables meteorológicas se mantienen, destacando la fuerte correlación positiva entre la temperatura y el déficit de presión de vapor (VPD), así como la relación negativa entre la humedad relativa y estas variables, lo cual confirma la consistencia interna del subconjunto reducido.

Asimismo, se identifican correlaciones moderadas entre la variable de riesgo territorial y la activación de mensajes SAE, y entre la velocidad de viento y la ocurrencia de SAE, lo que indica que el proceso de reducción del dataset preserva la señal relevante asociada a la decisión de alerta. En contraste, variables topográficas como la pendiente continúa mostrando asociaciones débiles con la variable objetivo.

Si bien se evidencia la presencia de colinealidad entre algunas variables meteorológicas, esta condición no representa una limitación para los algoritmos de aprendizaje automático basados en árboles de decisión empleados en esta investigación, los cuales son robustos frente a este fenómeno. En este sentido, la matriz de correlación cumple un rol de validación metodológica, asegurando que el dataset final utilizado para el modelamiento conserva las propiedades estructurales necesarias para una predicción fiable del envío de mensajes SAE.

3.7 Codificación de Variables (Encoding)

Dado que los algoritmos de *Machine Learning* basados en árboles requieren entradas numéricas, se transformaron las variables categóricas; combustible_inicial, topografía, dirección viento, exposición y mes.

Para ello se implementó la técnica de codificación Dummy (Dummy Encoding). A diferencia del One-Hot Encoding estándar, se configuró la eliminación de la categoría base (drop_first = True) para convertir una variable K categorías en K-1 columnas binarias, optimizando la dimensionalidad.

Variables Finales del Modelo: El *dataset* resultante incluyó:

- **Numéricas:** Riesgo (codificado), temperatura, humedad, velocidad del viento, pendiente, VPD.
- **Categóricas (Codificadas):** Combustible inicial (*Pastizal, Matorral, Plantación de eucaliptus, Plantación pino, Desechos, Basural, Cultivo agrícola, Arbolado natural, otros*), topografía (*Suave, Irregular, Abrupta*), dirección del viento (*SE,*

W, S, Calmo, N, SW, NW, NE, E), exposición (*N, W, Plano, SW, E, NE, S, SE, NW*), mes (*mes_01, mes_02, mes_03 etc*)

3.8 Modelo Predictivo: Random Forest (Respuesta al OE3)

Para cumplir con el objetivo específico 3 se entrenó y evaluó un modelo de clasificación basado en el algoritmo Random Forest. Este modelo actúa como la línea base de la investigación, permitiendo establecer un estándar de desempeño inicial para la predicción de activación del SAE.

La variable objetiva corresponde a la activación del Sistema de Alerta de Emergencia (SAE), modelada como una variable binaria (0 = no activación, 1 = activación)

3.8.1 Estrategia de Muestreo (Train/Test Split)

Se dividió el conjunto de datos en Entrenamiento (70%) y Prueba (30%). Debido a la baja prevalencia de alertas SAE (desbalance de clases) se aplicó un muestreo estratificado (*stratify = y*) asegurando que la proporción de eventos con SAE (aproximadamente 19%) se mantuviera representativa en ambos subconjuntos.

Cabe señalar que, dado que los incendios con activación SAE presentan una concentración espacial significativa en la zona centro sur del país (véase Figura 12), la división aleatoria del dataset podría introducir casos geográficamente similares en ambos conjuntos. Esta condición, conocida en la literatura como fuga espacial o *spatial leakage*, puede producir una estimación levemente optimista del desempeño en comparación con una validación geográfica por macrozona o región.

3.8.2 Manejo del Desbalance en Random Forest

Dado que los incendios con SAE constituyen la clase minoritaria, se configuró el hiperparámetro `class_weight='balanced'`. Esta técnica penaliza de forma automática los errores de clasificación en la clase positiva, forzando al algoritmo a aprender los patrones de los eventos críticos sin necesidad de alterar artificialmente el volumen de datos mediante técnicas de sobre muestreo. En lugar de generar datos

sinéuticos, se optó por penalizar de forma algorítmica los errores en la clase minoritaria (incendios con SAE).

3.9 Modelo Predictivo: XGBoost y Análisis Comparativo (respuesta al OE4)

Finalmente, para dar cumplimiento al objetivo específico 4, se entrenó y evaluó el algoritmo XGBoost (*Extreme Gradient Boosting*), manteniendo la misma estrategia de muestreo (70/30). Se midió el desempeño de los modelos desde dos dimensiones complementarias: el rendimiento estadístico y la coherencia de las variables que determinan el envío del SAE.

3.9.1 Manejo del Desbalance en XGBoost

A diferencia del modelo anterior, en XGBoost se ajustó el hiperparámetro `scale_pos_weight`, calculado como la razón entre el número de casos negativos y positivos presentes en el conjunto de entrenamiento. Este parámetro permite compensar el desbalance de clases al asignar un mayor peso al error asociado a la clase minoritaria ($SAE = 1$), favoreciendo así la capacidad del modelo para detectar eventos que requieren la activación del SAE.

De esta manera, el modelo orienta el proceso de optimización hacia el incremento de la **Sensibilidad (Recall)**, reduciendo el riesgo de omitir incendios forestales que potencialmente podrían derivar en una alerta a la población.

Con excepción de los hiperparámetros `scale_pos_weight` en XGBoost y `class_weight = 'balanced'` en Random Forest, ambos modelos fueron entrenados utilizando los valores por defecto provistos por las librerías **XGBoost** y **Scikit-Learn**, respectivamente. Esta decisión se fundamenta en que dichos valores han sido diseñados para ofrecer un desempeño robusto en problemas de clasificación binaria con desbalance de clases, constituyendo una línea base metodológicamente aceptada en la literatura de *Machine Learning* aplicado.

3.10 Evaluación y Validación

Para validar ambos modelos y determinar el algoritmo más idóneo para SENAPRED, se definió un enfoque de evaluación orientado al riesgo, priorizando métricas que minimicen consecuencias críticas para la población por sobre métricas tradicionales basadas exclusivamente en la exactitud global.

3.10.1 Métricas de Desempeño Orientadas a la Gestión del Riesgo de Desastres

La selección del mejor modelo no se basó en la exactitud global, sino en métricas jerarquizadas por su impacto en la gestión del riesgo de desastres “Salvar vidas”:

1. **Sensibilidad (*Recall*):** Métrica primaria. Asegurar la detección de incendios que requieren SAE (minimizando los Falsos Negativos) aquellos incendios que el modelo omite.
2. **Área Bajo la Curva (AUC-ROC):** Para medir la capacidad de discriminación global del modelo.
3. **Matriz de confusión:** Herramienta utilizada para cuantificar operacionalmente los Falsos Positivos y Falsos Negativos.
4. **Curva Precision-Recall:** Herramienta idónea cuando la clase positiva es menor al excluir de la clase mayoritaria de la ecuación, enfocándose en el desempeño sobre las alertas SAE reales.

3.10.2 Análisis de Variables Determinantes

Más allá de las métricas, se evaluó la interpretabilidad del modelo para asegurar que sus decisiones fuesen consistentes con el comportamiento del fuego y la gestión del riesgo de desastres que es salvar vidas y reducir los impactos ante un evento.

- **Ranking de importancia:** Se extrajo el peso relativo de cada predictor utilizando los atributos nativos de los algoritmos (Feature Importance):
 - En Random Forest, se analizó la disminución media de la impureza (MDI)

- En XGBoost, se analizó la Ganancia de información (Gain) identificando qué variables dominan la estructura de los árboles de decisión.

Con esto, el ranking de importancia obtenido fue contrastado con el conocimiento experto y con la literatura especializada en incendios forestales, considerando antecedentes que describen la influencia conjunta de variables meteorológicas críticas (como la temperatura, humedad relativa y velocidad del viento) en el comportamiento del fuego, resumidas operativamente en reglas empíricas como el denominado factor 30-30-30.

3.10.3 Validación de Estabilidad

Finalmente, se verificó la **estabilidad estadística** del modelo con el mejor Recall, asegurando que su capacidad predictiva no fuera producto del azar o de una partición de datos favorable.

Procedimiento de Validación Cruzada (Estabilidad):

Se ejecutó una validación cruzada estratificada de **10 pliegues (StratifiedKFold, k = 10)**.

- Al dividir los datos en 10 subconjuntos distintos, se garantiza que el modelo sea probado en la totalidad de los datos históricos.
- Se evaluó la desviación estándar de la métrica **Recall**. Una desviación baja (<5%) se estableció como indicador estructural del modelo predictivo.

3.10.4 Estrategia de Ajuste Operativo del Umbral de Decisión (Threshold Tuning)

Los algoritmos de clasificación como Random Forest y XGBoost no predicen directamente una categoría (0 o 1), sino que estiman una probabilidad de pertenencia a la clase positiva. Por convención, las librerías establecen un umbral de corte estándar de 0.5, donde:

Figura 9

Representación Calibración del Umbral de Decisión

$$\hat{y} = \begin{cases} 1 & \text{si } P(y = 1) \geq 0.5 \\ 0 & \text{si } P(y = 1) < 0.5 \end{cases}$$

Nota. Regla de clasificación basada en un umbral de probabilidad de 0.5, donde la observación se asigna a la clase positiva si $P(y = 1) \geq 0.5$. Adaptado de *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*, por Géron, A. 2019. Derechos reservados 2019 para Kiwisoft S.A.S.

Sin embargo, en el contexto de la gestión del riesgo de desastres en la fase de respuesta, asumir un costo simétrico de errores (donde un Falso Negativo pesa lo mismo que un Falso Positivo) es metodológicamente incorrecto. Un umbral de 0.5 podría resultar excesivamente conservador, omitiendo alertas SAE reales.

Por tanto, se diseñó un procedimiento de umbral post-entrenamiento:

1. Se evaluó el desempeño del modelo en un rango continuo de umbrales de 0.10 a 0.90.
2. Se seleccionó un nuevo umbral que priorice la maximización del Recall, aceptando una degradación controlada de la precisión con el objeto de minimizar el riesgo de omisión de alertas SAE reales.

CAPÍTULO IV: ANÁLISIS DE RESULTADOS

4.1 Análisis Exploratorio de Datos (EDA)

Previo a la etapa de modelamiento predictivo, se llevó a cabo un Análisis Exploratorio de Datos (EDA) con el propósito de caracterizar la estructura del *dataset*, evaluar la coherencia física y estadística de las variables, e identificar patrones, relaciones y posibles inconsistencias que pudieran influir en el desempeño de los modelos de *Machine Learning*. Este análisis permitió fundamentar las decisiones metodológicas adoptadas en las etapas posteriores del proceso de modelación.

4.1.1 Caracterización Estadística y Temporal.

En términos generales, el *dataset* se compone de **30.725 registros**, correspondiente a eventos de incendios forestales, sin presencia de valores nulos en las variables utilizadas, lo que permitió preservar la totalidad de los registros para las etapas posteriores de análisis.

Las **variables meteorológicas** presentan comportamientos consistentes con el contexto físico del fenómeno. **La temperatura** muestra un valor promedio de 23.3 °C con un rango entre 5°C y 40,3 °C, reflejando condiciones típicas de la temporada estival, periodo en el cual se concentra la mayor ocurrencia de incendios forestales. **La humedad relativa** presenta un promedio cercano al 47% con una mediana de 45% y un rango amplio entre 0% y 100% lo que evidencia una alta variabilidad ambiental.

Respecto a la **velocidad del viento**, se observa un valor medio de 8.6 km/h con una mediana de 6.5 km/h y valores máximos extremos que alcanzan los 177 km/h. Estos valores elevados corresponden a eventos atípicos, probablemente a condiciones excepcionales, los cuales fueron mantenidos en el análisis al no constituir valores nulos o inconsistentes desde el punto de vista estructural del *dataset*.

Las **variables topográficas** presentan una distribución concentrada en valores bajos de pendiente, con una mediana igual a 0% y un percentil 75 de 16% lo que sugiere que una proporción importante de los incendios analizados ocurren en terrenos de topografía relativamente suave. Este patrón es consistente con la localización de incendios forestales en zonas de interfaz urbano-rural.

En cuanto a la **superficie total quemada**, la variable muestra una marcada asimetría positiva, con una mediana de 0,4 ha y un valor máximo cercano a 69.800 ha, reflejando la coexistencia de una gran cantidad de incendios de pequeñas magnitudes junto con un número reducido de incendios forestales de gran escala (como los incendios forestales de Viña del Mar el año 2024).

La **variable categórica riesgo**, derivada del cruce espacial con el mapa de riesgo de CONAF, presenta una distribución centrada en niveles alto y muy alto, lo que indica que una proporción relevante de los incendios forestales analizados ocurre en áreas con condiciones de amenaza significativa a la población.

Finalmente, la **variable objetivo SAE** presenta una prevalencia del 3,99 % confirmado un fuerte desbalance de clases.



Tabla 3

Estadística Descriptiva de Variables Numéricas

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
alerta	30725	3	Verde	29649	NaN	NaN	NaN	NaN	NaN	NaN	NaN
combustible_inicial	30725	9	Pastizal	16183	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Superficie_total	30725.0	NaN	NaN	NaN	21.432184	488.710268	0.001	0.1	0.4	1.76	69080.51
temperatura	30725.0	NaN	NaN	NaN	23.324872	6.071062	5.0	19.0	23.5	28.0	40.3
nubosidad	30725.0	NaN	NaN	NaN	1.191733	2.276384	0.0	0.0	0.0	1.0	8.0
humedad	30725.0	NaN	NaN	NaN	46.983717	17.382189	0.0	35.0	45.0	58.0	100.0
exposicion	30725	9	Plano	16781	NaN	NaN	NaN	NaN	NaN	NaN	NaN
direccion viento	30725	9	S	10560	NaN	NaN	NaN	NaN	NaN	NaN	NaN
velocidad viento	30725.0	NaN	NaN	NaN	8.560342	7.422416	0.0	3.5	6.5	11.6	177.0
topografia	30725	3	Suave	19400	NaN	NaN	NaN	NaN	NaN	NaN	NaN
pendiente	30725.0	NaN	NaN	NaN	12.323749	20.602494	0.0	0.0	0.0	16.0	100.0
latitud	30725.0	NaN	NaN	NaN	-36.401044	2.964595	-53.946944	-38.108333	-36.681389	-34.410182	-17.721111
longitud	30725.0	NaN	NaN	NaN	-72.383992	3.282375	-109.440556	-72.789167	-72.229444	-71.441389	-68.010556
riesgo	30725.0	NaN	NaN	NaN	2.670203	1.013565	1.0	2.0	3.0	3.0	4.0
mes	30725	12	mes_01	6751	NaN	NaN	NaN	NaN	NaN	NaN	NaN
sae	30725.0	NaN	NaN	NaN	0.039902	0.195733	0.0	0.0	0.0	0.0	1.0

Figura 10

Gráfico de Total de Incendios vs. Alertas SAE (Escala Logarítmica)

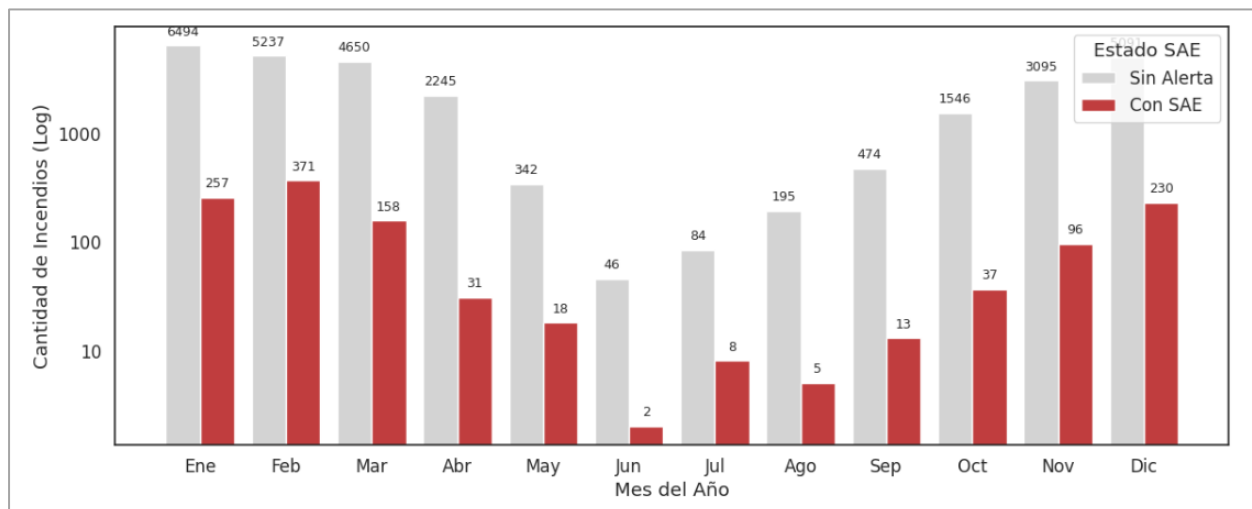
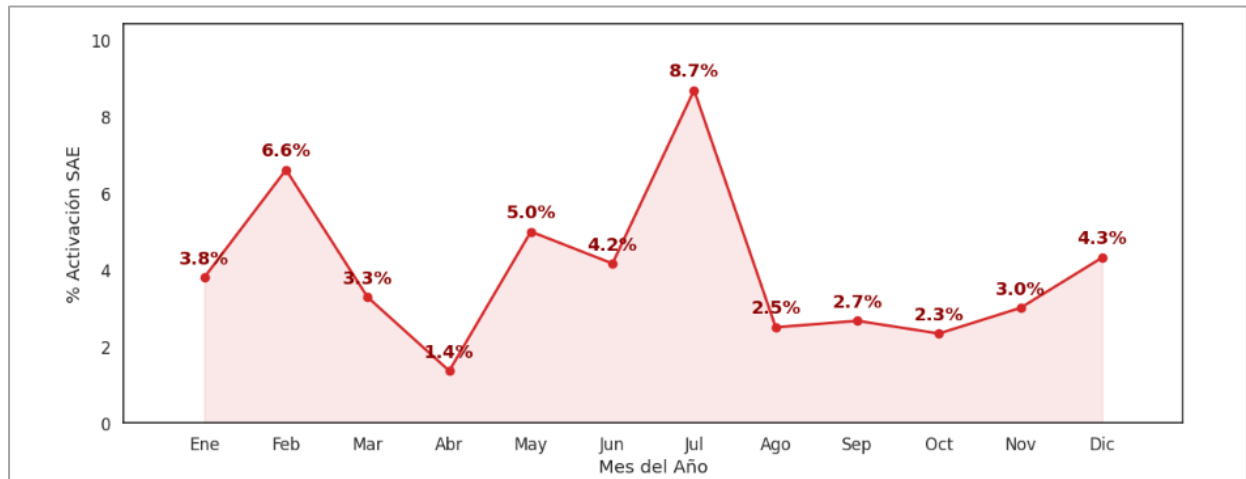


Figura 11

Gráfico de Tasa de Activación SAE (% de incendios que requirieron Alerta)



En Figura 10, se observa una fuerte concentración de eventos en el periodo estival (diciembre, enero y febrero), donde la frecuencia mensual supera los 6.000 incendios aproximadamente. Se evidencia un desbalance de clases importante, incluso en los meses de mayor actividad, la proporción de alertas SAE es significativamente menor a los incendios sin alerta. Gracias a la Transformación logarítmica se detecta que, aunque la cantidad descende drásticamente en invierno (junio - agosto) la clase (SAE = 1) no desaparece. Esto sugiere que las condiciones para una alerta SAE no son exclusivas del verano, sino que responden a factores multivariables que pueden converger en cualquier época del año.

Respecto al segundo gráfico, Figura 11 “Tasa de Activación SAE”, revela que los *insights* operativos quedan ocultos al observar solo la cantidad en bruto. Contra todo pronóstico intuitivo, el mes de Julio presenta la tasa de activación más alta del periodo 2019-2025 (8.7%), superando incluso el mes de febrero. Esto se podría entender cómo, si bien julio registra un número bajo de incendios totales, aquellos que logran desarrollarse en pleno invierno suelen estar asociados a condiciones anómalas severas (ejemplo, sequía extrema o viento puelche en la zona centro-sur).

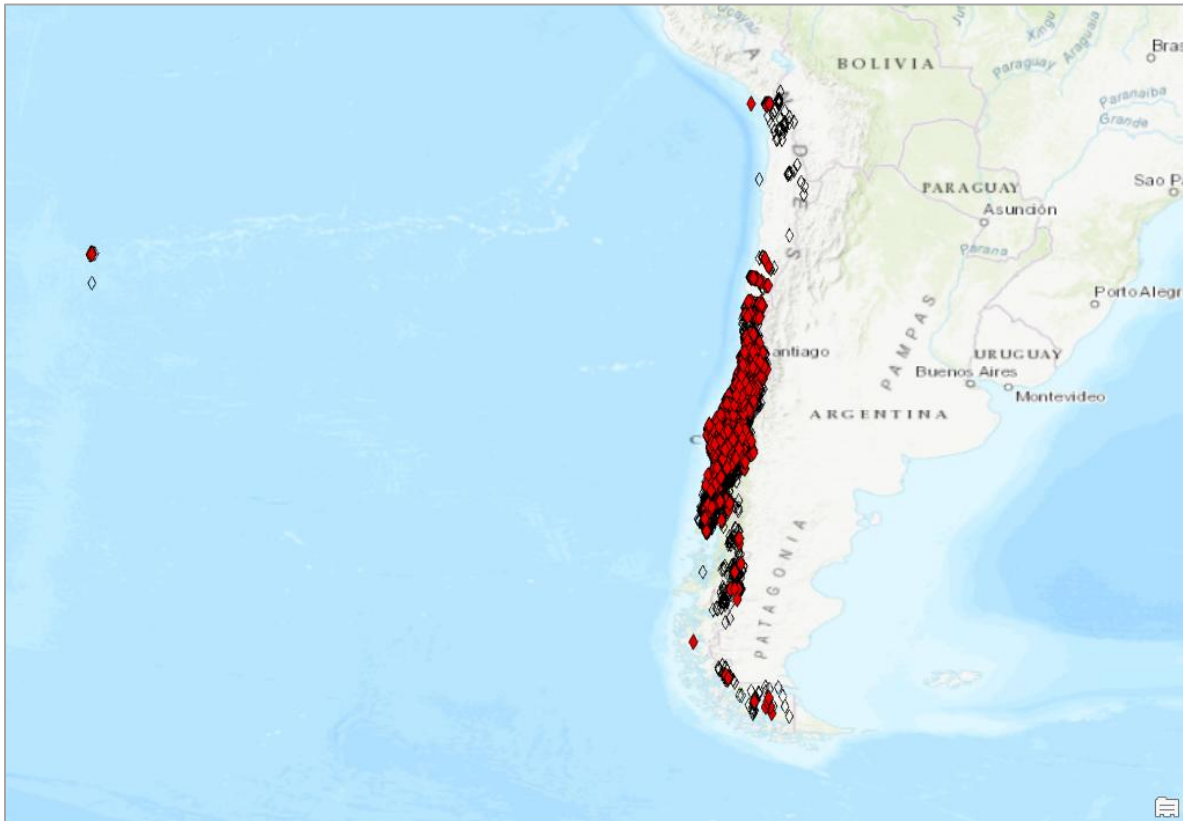
4.1.2 Análisis Exploratorio Espacial de Incendios Forestales

Como parte del análisis exploratorio de datos se examinó la distribución espacial de incendios forestales registrados entre los años 2019 - 2025, diferenciando aquellos eventos que derivaron en la activación de un mensaje SAE de aquellos que no lo hicieron.

Este análisis tuvo como objetivo identificar patrones espaciales preliminares y evaluar si la ocurrencia de incendios con SAE presenta concentración territorial asociadas a zonas de mayor exposición o riesgo, descartando un comportamiento aleatorio del fenómeno.

Figura 12

Mapa de Incendios con Activación SAE en el Territorio Nacional



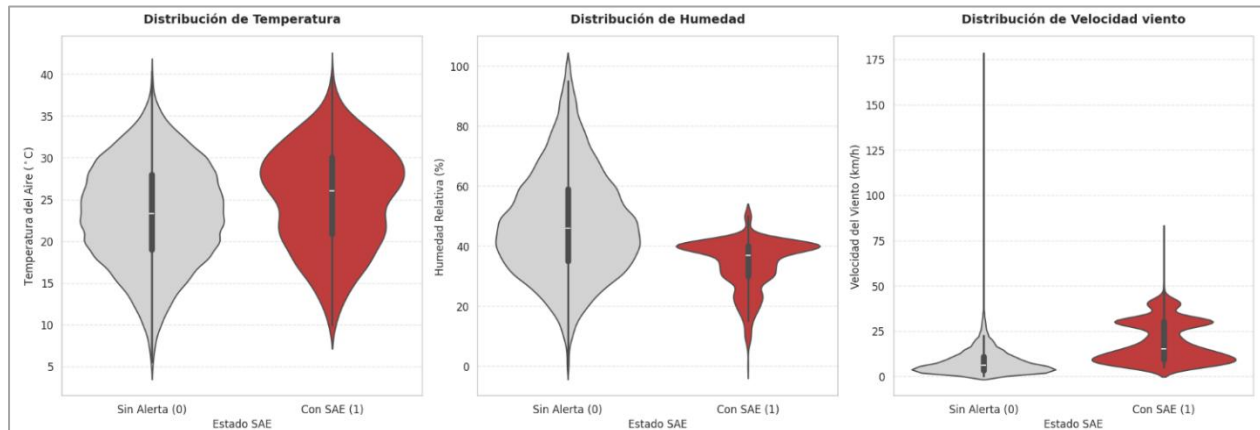
La Figura 12, muestra que los incendios con activación SAE (diamantes de color rojo) se concentran principalmente en la zona centro-sur del país, particularmente en áreas con alta interfaz urbano-rural, mientras que los incendios sin SAE (diamantes transparentes) presentan una distribución más dispersa. Esta evidencia respalda la incorporación de variables espaciales y de riesgo territorial dentro del modelo predictivo.

4.1.3 Comportamiento Meteorológico

Para confirmar la coherencia física en el medio, se contrastaron las distribuciones de variables climáticas entre incendios con y sin SAE. Los gráficos de violín evidencian que los eventos SAE (rojo) están asociados a mayor velocidad del viento y menor humedad.

Figura 13

Gráfico de Distribución de Temperatura, Humedad y Velocidad del Viento



A. Temperatura:

- Se observa un desplazamiento vertical claro de la masa de datos. Mientras la clase 0 (Gris/sin Alerta) tiene una distribución casi normal centrada en los 23°C, la clase 1 (Rojo/Con Alerta) muestra un achatamiento en la base y concentra su mayor densidad (la parte más ancha del violín) en el rango 26°C-32°C.
- La temperatura actúa como un catalizador basal. Nótese que el violín rojo es más "ancho" arriba, indicando que el calor es una condición necesaria pero no suficiente. Existe un solapamiento importante entre los 20°C y 25°C lo que explica por qué la temperatura por sí sola tiene una correlación, por tanto, el calor predispone el combustible, pero requiere de otros factores para gatillar la alerta SAE.

B. Humedad Relativa:

- Esta variable es la con mayor poder discriminante, la clase 0 (gris) distribución dispersa y alargada, llegando hasta el 100% de humedad, la clase 1 (roja) distribución extremadamente compacta y comprimida hacia abajo. La joroba principalmente se ubica bajo el 30-35%.



- El gráfico confirma que la humedad funciona como un umbral de corte, prácticamente no existen alertas SAE con humedades superiores al 50%. Esto valida la regla empírica del factor "30-30-30". Estadísticamente la baja varianza en la clase roja indica que la sequedad atmosférica es un requisito importante para la propagación de los incendios.

C. Velocidad del viento:

- Aquí se revela un fenómeno interesante, la clase 0 presenta *outliers* extremos (la línea negra vertical que llega a 175 km/h). Esto corresponde a eventos temporales en el extremo sur de Chile o en zonas cordilleranas.
- La clase (rojo), aunque no alcanza esos extremos, su base es más ancha entre los 15 y 30 km/h.
- Lo que define al SAE no es una ráfaga, sino un viento sostenido y constante en rangos medios altos (15-30 km/h) combinado con la sequedad. El violín rojo tiene su centro de gravedad claramente más arriba que el gris confirmando que el viento es el vector dinámico que transforma un incendio estático en una amenaza real.

4.1.4 Coherencia del Mapa de Riesgo con la Activación de SAE

Se evaluó la relación entre la variable Riesgo y la probabilidad de alerta, observándose un crecimiento exponencial en zonas de riesgo alto y muy alto.

Figura 14

Gráfico de Distribución de Incendios v/s Activación SAE por Nivel de Riesgo

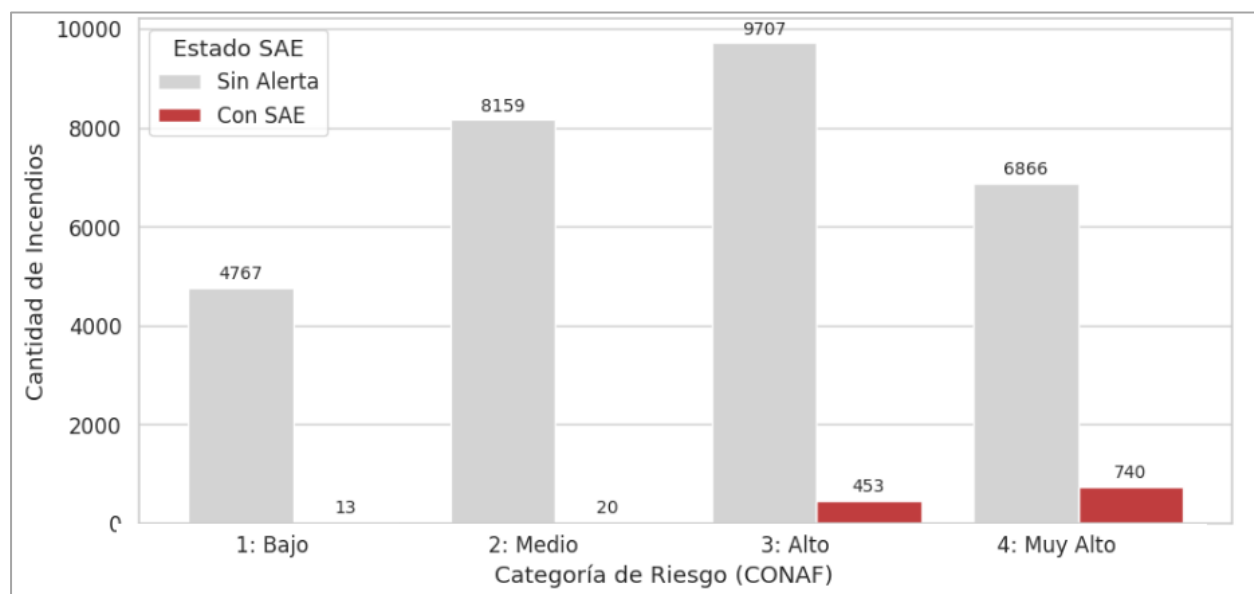
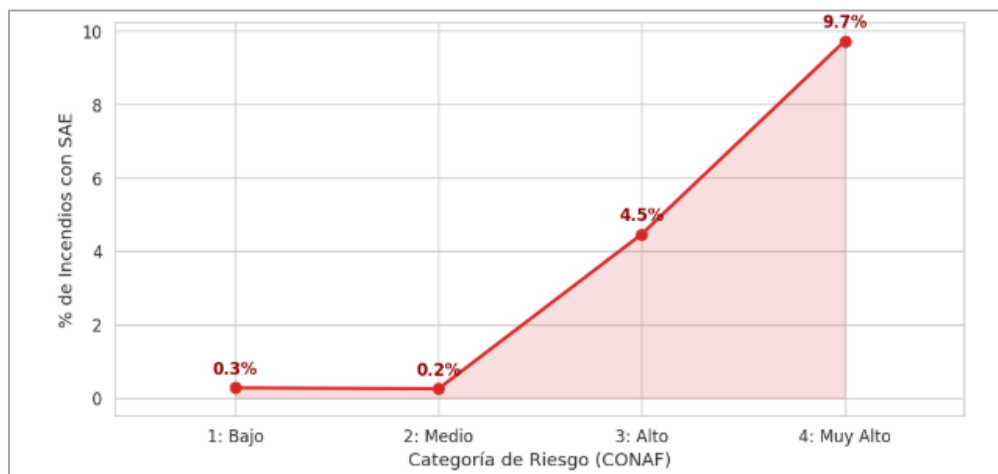


Figura 15

Validación: Probabilidad de SAE según Nivel de Riesgo CONAF



La Figura 14, muestra el gráfico donde se analiza la relación entre la clasificación de riesgo de CONAF y la activación SAE. Este cruce permite validar si la cartografía oficial de amenazas es un predictor eficaz para el modelo de *Machine Learning*.

1. Distribución:

- A. El grueso de los incendios se concentra en el nivel 3 (Alto Riesgo), con 9.707 incendios sin alerta y 453 incendios con alerta SAE. Esto es consistente con zonas de interfaz urbano-rural donde la actividad humana es intensa, generando muchas igniciones, pero no todas catastróficas.
- B. En los niveles 1 y 2 a pesar de existir una distribución considerable de incendios, la cantidad de alertas SAE es despreciable (13 y 20 eventos respectivamente). Esto indica que los incendios en zonas de bajo riesgo se encuentran alejadas a centros poblados, por lo que casi nunca escalan a una emergencia SAE.

2. Probabilidad condicionada de activación: La Figura 15, presenta la probabilidad estimada de activación del SAE condicionada al nivel de riesgo CONAF. Este análisis permite evaluar si la variable espacial derivada del mapa de riesgo entrega una señal consistente y discriminante respecto de la ocurrencia de SAE.

- A. **Riesgo 1 y 2:** La probabilidad de activación es estadísticamente nula ($<0.3\%$). Para el modelo predictivo, esto actuará como un filtro negativo clave: Si el incendio es riesgo 1 o 2 la probabilidad de SAE es casi cero.
- B. **Riesgo 3:** Se observa un quiebre abrupto en la curva la probabilidad salta a un 0.2% en riesgo medio a un 4.5% en riesgo alto. Esto sugiere que el nivel 3 es el verdadero umbral crítico donde la vulnerabilidad territorial comienza a ser clave.

C. Riesgo 4: En el nivel máximo, la tasa de activación se duplica nuevamente hasta el 9.7%. Un incendio en esta zona es 32 veces más propenso a generar una alerta SAE que uno en zona de riesgo 1.

4.2 Evaluación Comparativa del Desempeño Predictivo

Se evaluó la capacidad de generalización de ambos algoritmos en el conjunto de datos de prueba. A continuación, se presenta el reporte detallado de clasificación, desglosando el rendimiento para la clase mayoritaria (Sin SAE) y la clase positiva (SAE = 1)

Tabla 4

Reporte de Clasificación por Clase (Random Forest y XGBoost)

Modelo	Clase	Precisión	Sensibilidad (Recall)	F1-Score	Soporte (Casos)
Random Forest	0 (No Alerta)	0.93	0.95	0.94	1535
	1 (SAE)	0.77	0.70	0.73	368
XGBoost	0 (No Alerta)	0.96	0.91	0.93	1535
	1 (SAE)	0.69	0.83	0.75	368

Como se evidencia en la Tabla 4, existe una clara diferencia en cómo ambos modelos abordan la clase minoritaria (SAE=1)

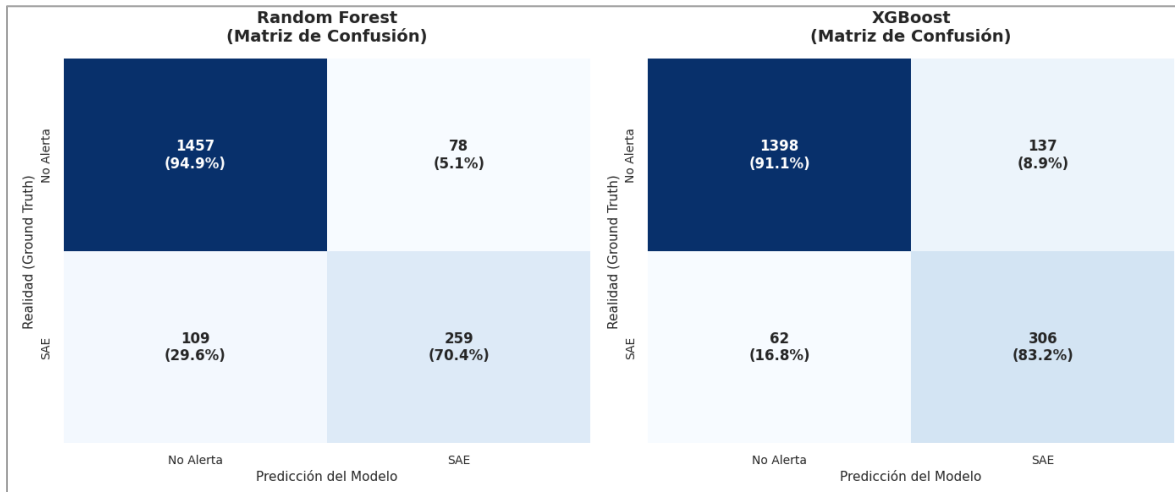
- Random Forest muestra un comportamiento conservador: tiene mayor Precisión (0.77) pero sacrifica la detección de alertas (Recall 0.70)
- XGBoost muestra un comportamiento más sensible: sacrifica puntos de Precisión (baja a 0.69) para maximizar drásticamente la capacidad de encontrar incendios forestales con SAE (Recall sube a 0.83)

Para comprender el impacto operativo de estos resultados, se analiza a continuación la Matriz de Confusión, que detalla la cantidad exacta de incendios detectados y omitidos.

La Figura 16, presenta las matrices de confusión resultantes, las cuales permiten visualizar la distribución exacta de los aciertos y errores de clasificación

Figura 16

Matrices de Confusión Resultantes



4.2.1 Análisis de la Capacidad de Detección (Sensibilidad)

Al analizar el cuadrante de los eventos positivos (fila inferior de las matrices), se observan diferencias sustanciales en el perfil de seguridad de ambos modelos:

- **Random Forest:** Logró identificar correctamente el 70.4% de los incendios que efectivamente requerían SAE (259 eventos). Sin embargo, presentó una tasa de omisión alta, dejando pasar 109 incendios forestales como “No SAE”, (Falsos Negativos), lo cual representa un riesgo operativo alto.
- **XGBoost:** Elevó significativamente la tasa de detección al 83.2% (306 eventos), reduciendo los Falsos Negativos a solo 62 casos.

Interpretación del Recall (83.2%): En términos prácticos este porcentaje indica que de cada 100 incendios forestales reales con SAE, el algoritmo XGBoost logra detectar 83,

mientras que 17 son omitidos. Aunque no es perfecto, representa una mejora sustancial respecto al modelo Random Forest.

Impacto en Seguridad: En la gestión del riesgo de desastres, específicamente en la fase de respuestas, un falso negativo (omitir una alerta SAE necesaria) es el error más penalizado. En términos absolutos, XGBoost logró considerar 47 incendios críticos adicionales que el modelo Random Forest había ignorado. Estos 47 eventos representan oportunidades vitales de evacuación que el modelo XGBoost si fue capaz de advertir.

4.2.2 Análisis de Falsas Alarmas (Precisión)

Como contraparte a su alta sensibilidad, el modelo XGBoost generó una mayor cantidad de Falsos Positivos.

- **XGBoost:** Generó 137 falsas alarmas (8.9% de los casos negativos), resultando en una Precisión del 69%.
- **Random Forest:** Generó solo 78 falsas alarmas (5.1%) manteniendo una Precisión superior.

Una precisión del 69% implica que de cada 10 alertas SAE sugeridas por el modelo XGBoost, aproximadamente 7 corresponden a incendios reales con SAE, mientras que 3 son falsas alarmas o sobre alertamiento (eventos que el modelo clasificó erróneamente).

Debe considerarse que la Precisión es una métrica sensible a la prevalencia de la clase positiva en el conjunto evaluado. El valor 69% reportado para XGBoost es válido para el dataset filtrado con una prevalencia del 19.33% de casos SAE. En el universo original de incendios con una prevalencia del 3.99% este valor sería considerablemente menor. Por esta razón, la Precisión se considera como indicador complementario del costo operacional de falsas alarmas en el contexto del conjunto de análisis. La métrica primaria del estudio es el Recall, que es más estable ante cambios de prevalencia y más relevante para el objetivo de la seguridad pública.

4.2.3 Selección del Modelo Bajo Criterio de Riesgo Operacional

La decisión de selección del modelo definitivo se fundamenta en el principio de precaución aplicado en la Gestión del Riesgo de Desastres. Aunque esta herramienta funciona como un sistema de soporte de decisión, donde la autoridad de SENAPRED tiene la última palabra sobre enviar o no un SAE, es necesario de todas formas minimizar el costo de una recomendación errónea.

- 1. Costo de Falsa Alarma (Falso Positivo):** Es un costo administrativo aceptable. Si el modelo sugiere una alerta errónea, el director de SENAPRED Regional puede descartarla tras verificar con el organismo técnico (CONAF y Bomberos).
- 2. Costo de Omisión (Falso Negativo):** Es un costo alto. Si el modelo no alerta sobre un incendio forestal en la fase de respuesta, se pierde la oportunidad de salvar vidas y reducir pérdidas materiales.

Por tanto, aunque Random Forest presenta una mayor precisión global, se descarta debido a su alta tasa de omisión (29.6%) la cual es incompatible con los estándares de seguridad requeridos.

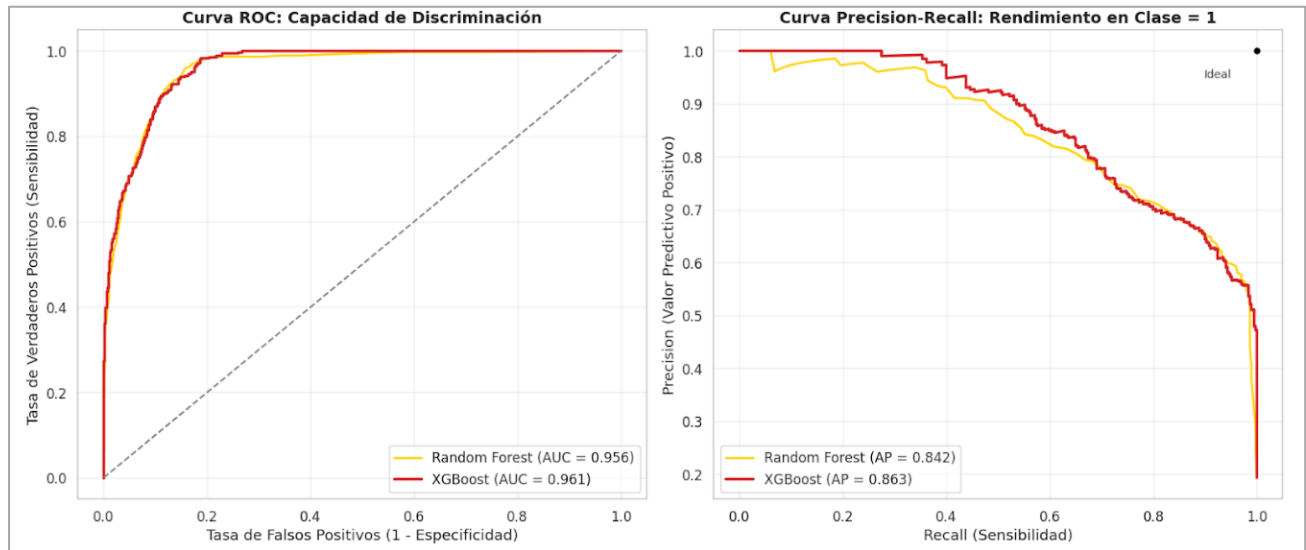
Por consiguiente, se selecciona XGBoost como modelo final, ya que su perfil de riesgo está alineado con la misión institucional de SENAPRED de “Salvar Vidas”. La capacidad de detectar 47 eventos críticos adicionales justifica la carga operativa extra de gestionar las falsas alarmas (Falsos Positivos), priorizando la seguridad por sobre la pureza estadística.

4.2.4 Análisis de Curvas de Rendimientos (ROC y Precisión-Recall)

Para validar la robustez de los modelos más allá de un umbral de corte definido, se analizaron las curvas de rendimiento global. La Figura 17, presenta comparativamente la Curva ROC (capacidad de discriminación) y la Curva Precision-Recall (rendimiento en clase desbalanceada).

Figura 17

Gráfico Comparativo: Curva ROC (capacidad de discriminación) y la Curva Precision-Recall (rendimiento en clase desbalanceada)



A. Análisis de la Curva ROC (Capacidad de Discriminación Global). El gráfico de la izquierda muestra la curva Receiver Operating Characteristic (ROC). Ambos modelos exhiben un rendimiento sobresaliente, acercándose a la esquina superior izquierda (el punto ideal de clasificación perfecta).

- **XGBoost (Línea Roja):** Alcanzó un Área Bajo la Curva (AUC) de 0.961.
- **Random Forest (Línea Amarilla):** Obtuvo un AUC de 0.956

Un AUC superior a 0.95 en ambos casos indica que los algoritmos tienen una probabilidad mayor al 95% de clasificar correctamente a un incendio forestal. Si bien la diferencia es marginal (0.001), XGBoost demuestra una ligera ventaja al distinguir incendios con SAE e incendios sin SAE, lo que indica que se puede alcanzar tasas de Verdaderos Positivos (Sensibilidad) superiores al 90% manteniendo una tasa de Falsos Positivos muy baja (<10%).



B. Análisis de la Curva Precision-Recall (Rendimiento ante desbalance). Dado el fuerte desbalance de clases, la Curva Precision - Recall (gráfico derecho) resulta ser el indicador más fiable, ya que ignora los verdaderos negativos (Incendios Forestales sin SAE) y se focaliza exclusivamente en el desempeño sobre los incendios con SAE.

- **Brecha de calidad:** Línea roja (XGBoost) v/s la línea amarilla (Random Forest) en el rango Recall 0.4 y 0.8, la línea roja se mantiene sistemáticamente por encima. Esto significa que para detectar la misma cantidad de incendios, XGBoost comete menos errores de Falso Positivo (Falsas Alarmas) que Random Forest.
- **Estabilidad:** Si nos fijamos en el extremo derecho (Recall>0.8), cuando se exige al sistema que detecte el 90% de los incendios, Random Forest “colapsa” más rápido (su precisión cae antes que XGBoost)
- **Métricas AP (Average Precision):** El puntaje 0.863 de XGBoost confirma que es un modelo más consistente y fiable para tomar decisiones.

En conclusión, el análisis de las curvas confirma la superioridad de XGBoost. No solo clasifica mejor (AUC 0.961 frente a 0.956 en Random Forest), sino que demuestra ser más certero (AP 0.863). Además, logra capturar la mayor cantidad de alertas reales (minimización de omisiones), manteniendo el incremento de falsas alarmas (Falsos Positivos) dentro de un margen controlado aceptable bajo el principio de precaución.

4.3 Validación de Estabilidad (Cross-Validation)

Para descartar que el rendimiento observado en el conjunto de prueba fuera producto del azar o de una partición de datos favorable, se sometió al modelo seleccionado (XGBoost) a una **validación cruzada estratificada de 10 pliegues** (K=10). Este proceso evaluó el algoritmo en diez subconjuntos de datos distintos e independientes.

Los resultados, resumidos a continuación, conforman la estabilidad estructural del modelo predictivo:

- Sensibilidad Promedio (Recall): 83.04% $\pm$ 3.79%
- AUC Promedio: 0.9668 $\pm$ 0.0056

El modelo obtuvo una sensibilidad promedio de 83.04%. Este valor es estadísticamente consistente con el 83.2% obtenido en el subconjunto de prueba único, lo cual permite **descartar el sobreajuste (overfitting)**.

Desde una perspectiva operativa, para SENAPRED la baja desviación estándar $\pm$ 3.79% ofrece una garantía de fiabilidad, ya que en el peor de los escenarios estadísticos, el modelo mantendrá una tasa de detección cercana al 80% y en los mejores casos superará el 86%. Esta acotada variabilidad confirma que el sistema es altamente confiable y predecible como herramienta de apoyo en la decisión.

El Área Bajo la Curva promedio fue de 0.9668 con una desviación marginal de $\pm$ **0.0056**. Una variación tan baja indica que la capacidad del modelo para distinguir entre incendios con SAE y sin SAE es idéntica en cualquier subconjunto de datos. Esto valida que las variables predictoras (Riesgo, humedad y velocidad del viento) son determinantes dentro del dominio de estudio y no meras correlaciones casuales de un grupo específico de incendios.

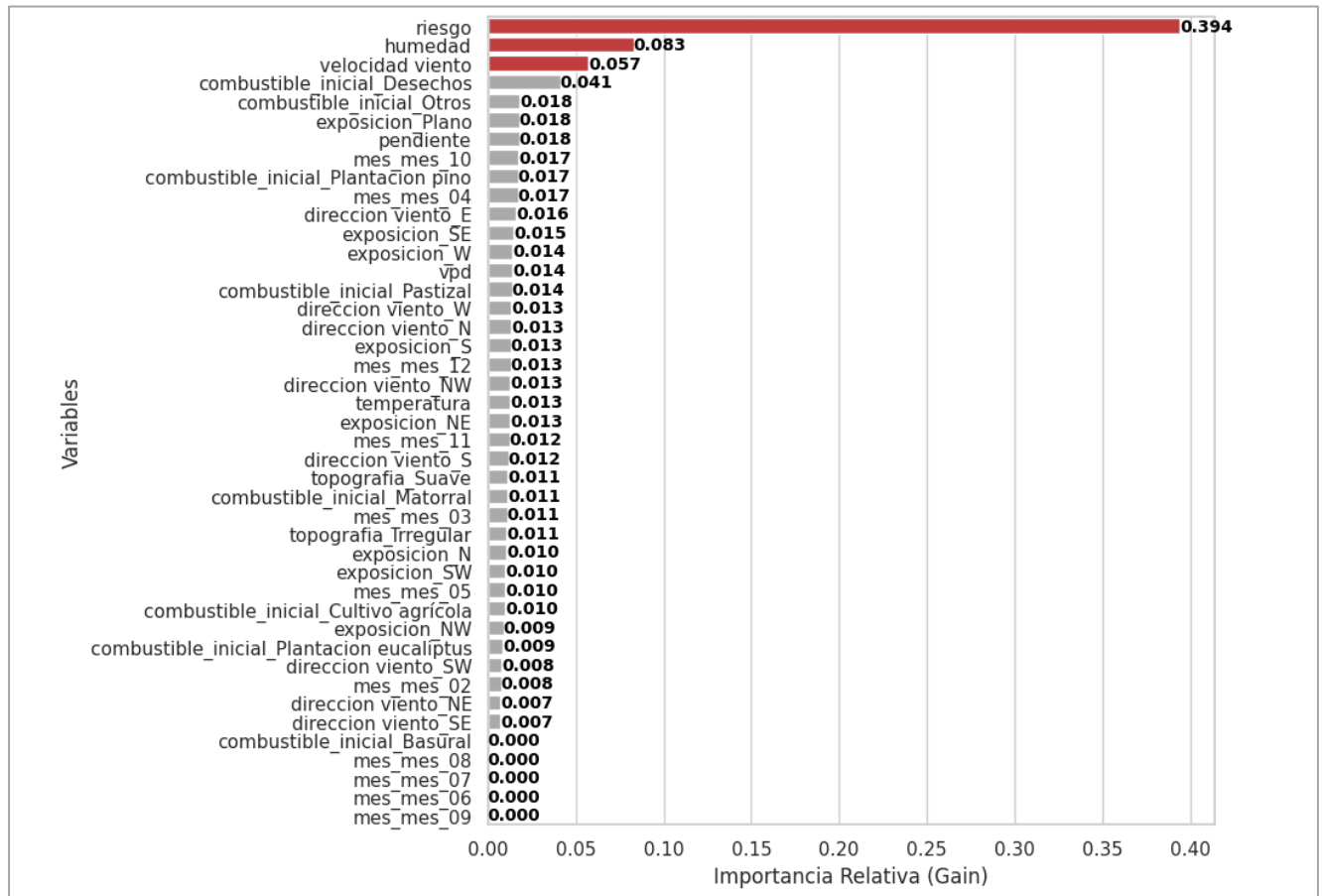
4.4 Análisis de Importancia de Variables

Antes de evaluar la capacidad predictiva global, se analizó qué factores está utilizando el modelo para tomar sus decisiones. Esto permite validar la coherencia del algoritmo con la realidad de los incendios forestales y el envío de mensajería SAE.

La Figura 18 presenta el ranking de importancia de características extraído del modelo XGBoost, cuantificando la ganancia de información que aporta cada variable a la reducción del error.

Figura 18

XGBoost: Ranking de Variables Determinantes para Alerta SAE



En la Figura 18 se desglosa la contribución relativa (Gain) de cada predictor en la construcción de los árboles de decisión del modelo XGBoost.

1. Variable de riesgo (0.394):

La variable de riesgo, basada en la cartografía de riesgo de CONAF, constituye el predictor dominante del modelo, concentrando un 39.4% de la ganancia total, con una importancia relativa aproximadamente cinco veces superior a la del segundo predictor más relevante. Este resultado indica que el moldeo asigna un peso preponderante a las condiciones territoriales de exposición y



vulnerabilidad, actuando como una probabilidad *apriori* sobre la cual las variables meteorológicas ajustan la decisión final.

La alta relevancia asignada a la variable de riesgo territorial es coherente con los patrones espaciales observados en el análisis exploratorio, donde los incendios con activación de SAE se concentraron mayoritariamente en zonas de alta exposición poblacional, reforzando la consistencia entre la evidencia descriptiva inicial y el comportamiento interno del modelo predictivo.

2. Variable humedad (0.083) y velocidad viento (0.057):

En segundo y tercer lugar del ranking de importancia se ubican la humedad relativa y la velocidad del viento respectivamente, consolidándose como variables meteorológicas clave en el modelo. La mayor relevancia de humedad por sobre la velocidad del viento sugiere que la disponibilidad y el estado de secado del combustible constituye un factor condicionante para la propagación del incendio forestal. En este sentido, los resultados del análisis muestran qué condiciones de alta humedad relativa limitan significativamente la generación y propagación del fuego, incluso ante la presencia de viento.

Por su parte la velocidad del viento puede interpretarse como el motor de propagación, actuando una vez que la variable riesgo define el “dónde” y la humedad establece el “cuándo” las condiciones meteorológicas son favorables, determinado así la velocidad e intensidad con que el incendio puede avanzar.

Cabe señalar que la humedad relativa y la temperatura presentan una relación inversa esperable desde el punto de vista físico, lo que evidencia la presencia de cierto grado de colinealidad entre ambas variables. No obstante, los algoritmos basados en árboles de decisión, como XGBoost son robustos frente a la multicolinealidad, ya que permiten capturar relaciones no lineales relevantes sin afectar negativamente el desempeño predictivo del modelo.



En este contexto, la mayor importancia relativa asignada a la humedad indica que esta variable proporciona una señal más directa y consistente para apoyar la decisión de activación de un mensaje SAE. Tal como se evidenció en el análisis exploratorio del subconjunto final del dataset, las variables meteorológicas mantienen relaciones físicas coherentes, lo que refuerza la consistencia entre la estructura de los datos y la interpretación de su relevancia dentro del modelo XGBoost.

3. Combustible inicial desechos (0.041):

Aparece en el cuarto lugar del ranking, por encima de bosque de pino o eucalipto; esto se podría deber a que los incendios en “desechos” (basurales, vertederos ilegales, microbasurales, quemas ilegales, etc.) suelen ocurrir en la interfaz Urbano-Rural, es decir, muy cerca de sectores poblados. El modelo detectó que un incendio forestal, aunque pequeño, tiene una alta probabilidad de requerir una evacuación inmediata por su proximidad a la población.

4. Otras variables:

A partir de este punto del ranking, variables como combustible_inicial, exposición_plano y pendiente presentan valores de importancia inferior a 0.018. Si bien estas variables contribuyen al desempeño del modelo, su importancia es significativamente menor en comparación con las variables principales identificadas previamente.

Este resultado indica que dichas variables cumplen un rol complementario dentro del proceso de modelación, aportando información contextual que permite refinar la clasificación en escenarios específicos, pero sin constituir los factores determinantes en la predicción del envío SAE. En consecuencia, la capacidad explicativa principal del modelo se concentra en un conjunto reducido de variables de mayor importancia relativa, mientras que las restantes variables operan como elementos de ajuste que mejoran la precisión del modelo.

Es relevante precisar que el ranking de importancia derivado de XGBoost constituye una medida de utilidad predictiva dentro del modelo, y no una prueba de causalidad. La Ganancia de Información indica qué variables permiten reducir el error de predicción de forma más eficiente, pero no permite inferir que dichas variables sean causas directas de la activación SAE.

4.5 Ajuste Operativo del Umbral de Decisión

Habiendo validado la estabilidad del modelo, se procedió a la sintonización fina del umbral de decisión (threshold tuning). El umbral estándar de 0.5 no necesariamente representa el punto óptimo para la fase de respuesta en el ciclo del riesgo de desastres.

Cabe precisar que el procedimiento implementado corresponde a threshold tuning o ajuste del umbral de decisión post-entrenamiento, y no a la calibración probabilística en sentido estricto, que implicaría técnicas como Platt Scaling o Isotonic Regression para ajustar las probabilidades estimadas por el modelo. Dado que el umbral fue seleccionado evaluando en trade-off entre Recall y Precisión sobre el mismo conjunto de prueba utilizado para reportar resultados, la estimación podría ser levemente optimista. Este ajuste se presenta como una decisión operativa orientada a privilegiar el Recall en el contexto de seguridad pública, y no como evidencia independiente de mejor generalización del modelo.

La Tabla 5, despliega el comportamiento de las métricas al variar la exigencia de la probabilidad de alerta.

Tabla 5

Comportamiento de las Métricas

Umbral	Recall (Seguridad)	Precisión	F1-Score	Alertas Omitidas (FN)	Falsas Alarmas (FP)
0.10	94.6%	58.0%	0.719	20	252
0.15	94.0%	59.9%	0.732	22	232
0.20	92.4%	61.6%	0.739	28	212
0.25	91.0%	63.1%	0.745	33	196
0.30	89.9%	65.0%	0.755	37	178
0.35	88.6%	66.5%	0.760	42	164
0.40	86.7%	67.6%	0.760	49	153
0.45	85.1%	68.3%	0.758	55	145
0.50	83.2%	69.1%	0.755	62	137
0.55	81.5%	69.9%	0.753	68	129
0.60	79.1%	70.8%	0.747	77	120
0.65	77.2%	71.9%	0.744	84	111
0.70	75.3%	72.3%	0.738	91	106
0.75	72.8%	74.2%	0.735	100	93
0.80	70.7%	76.7%	0.736	108	79
0.85	66.0%	81.8%	0.731	125	54
0.90	61.4%	84.6%	0.712	142	41

Análisis de Sensibilidad del Umbral: Se observa un intercambio claro Recall y Precisión, inversamente proporcional.

- Umbral Estándar (0.50):** Ofrece un Recall de 83.2%, si bien es un rendimiento alto, implica que un 16.8% de las alertas reales serán omitidas.
- Umbral de Seguridad (0.30):** Al reducir la exigencia de probabilidad al 30% el sistema eleva su sensibilidad al 89.9% (aproximadamente 90%). Esto significa que el modelo es capaz de recuperar casi un 7% ($89.9\% - 83.2\% = 6.7\%$) adicional de eventos con alertas reales que antes eran omitidas.

- 3. Costo Operativo:** Esta mejora en seguridad conlleva una disminución de la Precisión al 65.0% Operativamente, esto implica que cada 10 alertas SAE generadas, aproximadamente 6.5 corresponden a eventos SAE reales.

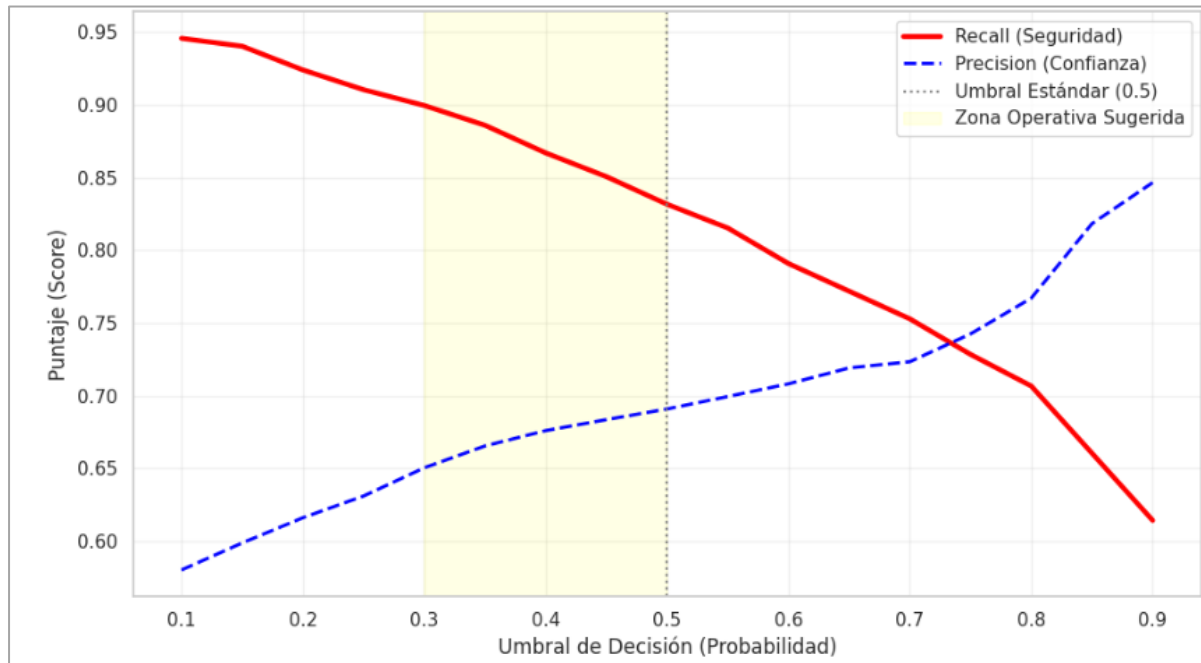
Por tanto, bajo el principio de prevención que rige en emergencia, se selecciona el umbral de 0.30 como el punto de operación definitivo. Esta configuración maximiza la probabilidad de "Salvar Vidas" (minimizando los Falsos Negativos) manteniendo la tasa de falsas alarmas o sobre alertamiento dentro de un rango manejable por los tomadores de decisiones de SENAPRED.

4.5.1 Análisis Visual del Intercambio de Métricas

Para validar gráficamente la elección del punto de corte, se generó el gráfico de "Análisis de Sensibilidad" presentado en la Figura 19, el cual proyecta la evolución inversa de las curvas de seguridad (Recall) y confianza.

Figura 19

Gráfico Análisis de Sensibilidad



En la Figura 19, el gráfico revela una “Zona operativa Sugerida” (franja amarilla) comprendida entre los umbrales 0.30 y 0.50. Dentro de este rango, el comportamiento de los modelos es estable.

- **La línea roja (Recall/Seguridad):** Muestra una pendiente descendente constante. se observa que a partir del umbral 0.50 (línea segmentada vertical) la caída en seguridad se acelera, perdiendo capacidad de detección rápidamente.
- **La Línea Azul (Precisión/Confianza):** Muestra una pendiente ascendente. Mejora a medida que el umbral sube, pero a un ritmo más lento que la pérdida del Recall.

La elección del umbral 0.30 se sitúa justo en el inicio (borde izquierdo) de la zona operativa. En este punto la línea roja (Recall ~ 90%) se mantiene alejada y por encima de la línea azul.

Esto confirma visualmente que la decisión de configuración está sesgada intencionalmente hacia la seguridad. Si nos movemos a la derecha (hacia el umbral de 0.50) las líneas comienzan a converger, lo que significaría sacrificar seguridad para ganar precisión, una estrategia que se ha desechado por la naturaleza de SENAPRED.

CAPÍTULO V: CONCLUSIONES

5.1 Verificación de Hipótesis y Conclusión General

La presente investigación planteó como hipótesis de trabajo que ***el uso de un modelo de Random Forest entrenado con datos históricos de incendios forestales, variables meteorológicas, topográficas y espaciales permiten predecir con alta precisión la necesidad de activar el SAE, constituyendo una herramienta valiosa para la toma de decisiones oportunas en la fase de respuesta en el ciclo del riesgo de desastres.***

A partir de la evidencia empírica obtenida, se acepta la premisa central de la hipótesis, validando que las técnicas de *Machine Learning* son capaces de modelar el proceso de decisión a la activación de alertas SAE en incendios forestales. No obstante, los resultados alcanzados hacen necesario introducir dos precisiones fundamentales respecto del planteamiento original:

1. Del Modelo Random Forest al Modelo XGBoost

Si bien la hipótesis inicial proponía el uso de Random Forest como línea base de investigación, la experimentación comparativa demostró que el algoritmo XGBoost (Extrema Gradient Boosting) presenta un desempeño superior para este tipo de fenómeno. En particular, XGBoost evidenció una mayor robustez frente al desbalance de clases, logrando una mejor capacidad de recuperación de eventos activación real de SAE. En consecuencia, la investigación valida el enfoque en *Machine Learning* pero redirige la elección del algoritmo hacia técnicas de boosting más adecuadas para contextos de riesgo. Es importante delimitar el alcance de esta comparación: ambos modelos fueron entrenados utilizando los hiperparámetros por defecto de sus respectivas librerías, ajustando únicamente los hiperparámetros de manejo del desbalance de clases. Por tanto la superioridad demostrada de XGBoost corresponde a este

dataset filtrado, bajo esta configuración específica, y no constituye evidencia de ventaja algorítmica general.

2. De la Alta Precisión Estadística a la Alta Sensibilidad Operativa

La hipótesis anticipaba una predicción caracterizada por una “alta precisión”. Sin embargo, el análisis desde una perspectiva operativa permitió establecer que, en el contexto de la fase de respuesta del ciclo del riesgo de desastres, la métrica prioritaria no es la precisión estadística, sino la sensibilidad del modelo y el resguardo de la seguridad de la población.

En este sentido, el modelo final fue calibrado no para minimizar falsas alarmas (Falsos Positivos), sino para reducir al máximo los falsos negativos, alcanzando una tasa de detección cercana al 90% (Recall). Esta decisión redefine el concepto de “herramienta valiosa” planteado en la hipótesis, su valor no radica en la ausencia de errores, sino en su capacidad de actuar como un sistema de apoyo confiable que permita anticiparse oportunamente a eventos de alto riesgo a la población. En términos prácticos el modelo desarrollado permite identificar aproximadamente 9 de cada 10 eventos que efectivamente requirieron la activación de un mensaje SAE, aportando a SENAPRED un sistema de soporte de decisión basado en datos que complementa el juicio experto de los organismos técnicos responsables en la emergencia.

5.2 Síntesis de Hallazgos

En relación con los objetivos específicos definidos al inicio de la investigación se destacan los siguientes hallazgos:

En primer lugar, se estableció que la calidad de los datos resulta más determinante que su cantidad. La estrategia de filtrado condicional, orientada a eliminar incendios forestales con superficies menores a 2.6 hectáreas, demostró ser metodológicamente adecuada. Al reducir el ruido generado por eventos de baja relevancia operativa se obtuvo un conjunto de datos más representativo, en el cual las variables espaciales y meteorológicas evidenciaron relaciones más coherentes con la ocurrencia de alertas SAE, permitiendo

al modelo aprender a partir de incendios forestales relevantes y no de anomalías estadísticas.

En segundo lugar, se confirmó que la amenaza preexistente del territorio constituye el factor predictivo más importante. La variable “riesgo” derivada del mapa de riesgo de CONAF se posicionó como el predictor dominante del modelo, validando que la activación del SAE depende estructuralmente de la vulnerabilidad territorial, más allá de las condiciones meteorológicas en el lugar. Asimismo, la coherencia lógica física observada entre variables como humedad relativa, velocidad del viento y temperatura refuerza que el modelo respeta la fenomenología conocida del comportamiento del fuego.

En tercer lugar, la comparación empírica entre algoritmos evidenció que XGBoost obtuvo un mejor desempeño que Random Forest para este caso de estudio, particularmente en la detección de alertas SAE reales, aspecto crítico en la gestión del riesgo de desastres. Cabe precisar que ambos modelos fueron entrenados utilizando los hiperparámetros por defecto de sus respectivas librerías, por lo que este resultado es válido para el dataset filtrado y bajo la configuración implementada, sin que constituya evidencia de superioridad algorítmica general.

5.3 Conclusiones Generales

El análisis de curvas de rendimiento confirmó la solidez del modelo XGBoost, mostrando un desempeño consistente tanto en términos de capacidad discriminativa (AUC) como en escenarios de fuerte desbalance de clases, donde la curva Precision-Recall permitió evaluar de manera más adecuada la relación entre detección y falsas alarmas (Falsos Positivos). Estos resultados fueron respaldados por el proceso de validación cruzada estratificada, el cual evidenció una baja variabilidad en las métricas de desempeño, descartando la presencia de sobre ajustes y confirmando la estabilidad estructural del modelo.

Por último, la investigación permitió establecer que, en contextos de seguridad pública, la métrica más importante no es la precisión global, sino la **sensibilidad del sistema**. La

calibración del umbral de decisión al 30% en lugar del estándar del 50% permitió recuperar aproximadamente un 7% adicional de alertas SAE reales previamente no detectadas.

Este ajuste asume el costo de una menor precisión, bajo el principio preventivo que orienta la gestión del riesgo de desastres resulta preferible asumir alertas SAE preventivas (Falsos Positivos) que enfrentar las consecuencias de una alerta no emitida. De este modo, la selección del umbral del 30% se fundamenta en el principio de prevención, alineando el sistema predictivo con la misión institucional de SENAPRED y consolidando su rol como **herramienta de apoyo a la toma de decisiones**.

CAPÍTULO VI: RECOMENDACIONES

A partir de los hallazgos obtenidos en esta investigación, se proponen las siguientes líneas de acción orientadas a fortalecer la toma de decisiones basada en evidencia en la fase de respuesta ante incendios forestales, consolidando el uso de herramientas analíticas como soporte institucional dentro de SENAPRED.

6.1 Implementación del Modelo Predictivo

6.1.1 Despliegue del modelo XGBoost

Se recomienda la adopción del modelo XGBoost como herramienta de apoyo primario para la evaluación técnica de activación de mensajes SAE ante amenaza de incendios forestales. Su implementación debe concebirse como un sistema de **soporte de decisión**, complementario al juicio experto de los equipos técnicos y a los protocolos de emergencias vigentes, y no como un mecanismo automatizado de activación. Esta limitación en la temporalidad de la información está asociada principalmente a la disponibilidad y actualización de los datos de entrada, y no restricciones propias del modelo predictivo.

El modelo puede integrarse en un *dashboard* institucional que entregue en tiempo casi real la probabilidad estimada de requerir la activación de un mensaje SAE para cada incendio forestal reportado, facilitando el seguimiento continuo de los eventos activos y fortaleciendo la capacidad de análisis temprano en contexto operativos.

6.1.2 Configuración del Umbral Operativo

Se recomienda operar el sistema bajo un umbral de decisión del 30%, asegurando una sensibilidad cercana al 90%. Esta configuración permite maximizar la detección de eventos críticos, reduciendo significativamente los falsos negativos en comparación con el umbral estándar de 50%.

Desde una perspectiva de gestión del riesgo de desastres, este ajuste operativo del umbral se alinea con el principio preventivo, priorizando la protección de la población frente al costo operativo asociado a falsas alarmas.

6.2 Gestión del Dato y Mejora Continua

6.2.1 Actualización de la Variable Riesgo

Dado que la cartografía de riesgo de CONAF se posicionó como el predictor más influyente del modelo, resulta fundamental la actualización periódica y sistemática de esta cobertura territorial. Cambios en expansión urbana e infraestructura crítica pueden alterar significativamente el nivel de exposición poblacional, impactando directamente en la capacidad predictiva del sistema.

La actualización de esta variable “riesgo” debe integrarse dentro de los procesos regulares de gestión de información territorial, asegurando la coherencia entre la realidad territorial y la base de entrenamiento del modelo.

6.2.2 Retroalimentación Continua del Sistema

Se recomienda establecer un mecanismo de retroalimentación de datos, incorporando sistemáticamente cada nuevo evento de incendio forestal, con o sin activación de mensaje SAE a la base histórica utilizada para el entrenamiento.

Este enfoque permitirá implementar un esquema de aprendizaje continuo, facilitando la adaptación del modelo a nuevos patrones (meteorológicos, topográficos, geoespaciales etc.). De este modo, el sistema no solo mantendrá su vigencia en tiempo, sino que podrá evolucionar frente a nuevos escenarios de emergencias.

6.2.3 Profundización del Análisis Exploratorio Temporal

Finalmente, se recomienda profundizar el análisis exploratorio de la distribución temporal de incendios y alertas SAE incorporando medidas de dispersión interanual, tales como: promedios mensuales y barras de error por temporada. Este enfoque permitirá identificar

si la variabilidad entre temporadas sigue patrones estructurales asociados a fenómenos climáticos como El Niño o La Niña, enriqueciendo la caracterización del fenómeno y fortaleciendo la base de conocimiento para futuras versiones del modelo predictivo.

BIBLIOGRAFÍA

- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration: Guidelines for computing crop water requirements* (Irrigation and Drainage Paper No. 56). Food and Agriculture Organization of the United Nations (FAO).
<https://www.fao.org/3/x0490e/x0490e00.htm>
- Amat Rodrigo, J. (2020, octubre). *Árboles de decisión, random forest, gradient boosting y C5.0. Ciencia de Datos*.
https://cienciadedatos.net/documentos/33_arboles_decision_random_forest_gradient_boosting_c50
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Carrasco, R. A., Bueno, I., y Montero, J.-M. (2024). Árboles de clasificación y regresión. En G. Fernández-Avilés y J.-M. Montero (Eds.), *Fundamentos de ciencia de datos con R*. McGraw-Hill. <https://cdr-book.github.io/cap-arboles.html>
- Corporación Nacional Forestal. (2024). Actualización Análisis Nacional de Amenaza y Riesgo por Incendios Forestales. Ministerio de Agricultura.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>

Chile. Ministerio de Transportes y Telecomunicaciones; Subsecretaría de Telecomunicaciones. (2012). *Reglamento para la interoperación y difusión de la mensajería de alerta, declaración y resguardo de la infraestructura crítica de telecomunicaciones e información sobre fallas significativas en los sistemas de telecomunicaciones* [Decreto 60]. 12 de mayo de 2012. <https://bcn.cl/3em1u>.

Chovanec, D., Kollár, B., Halúsková, B., Kubás, J., Pawęska, M., & Ristvej, J. (2025). A component-based approach to early warning systems: A theoretical model. *Applied Sciences*, 15(6), Artículo 3218. <https://doi.org/10.3390/app15063218>

Federal Emergency Management Agency. (s.f.). *Integrated Public Alert & Warning System*. U.S. Department of Homeland Security. <https://www.fema.gov/emergency-managers/practitioners/integrated-public-alert-warning-system>

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2.^a ed.). O'Reilly Media.

Julio, G., Castillo, E. y Pedernera, P. (1995). *Modelación de combustibles*. En *Actas del Taller Internacional sobre Prognosis y Gestión en Control de Incendios Forestales* (pp. 111-127). Universidad de Chile; Proyecto FONDEF.

Ley 21.180. Transformación Digital del Estado. 11 de noviembre de 2019. Ministerio Secretaría General de la Presidencia. Biblioteca del Congreso Nacional de Chile. <https://bcn.cl/33e45>

Ley 21.364. Establece el Sistema Nacional de Prevención y Respuesta ante Desastres, sustituye la Oficina Nacional de Emergencia por el Servicio Nacional de Prevención y Respuesta ante Desastres, y dicta otras disposiciones. 7 de agosto de 2021 (Chile). <https://bcn.cl/2qtdx>

Mileti, D. S. y Sorensen, J. H. (1990). *Communication of Emergency Public Warnings: A Social Science Perspective and State-of-the-Art Assessment* (ORNL-6609). Oak Ridge National Laboratory.

Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. (2024). *Política Nacional de Inteligencia Artificial: Actualización 2024*. Gobierno de Chile.

Oficina de las Naciones Unidas para la Reducción del Riesgo de Desastres y Organización Meteorológica Mundial. (2022). Global status of multi-hazard early warning systems: Target G. <https://www.undrr.org/publication/global-status-multi-hazard-early-warning-systems-target-g>

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2022). Recomendación sobre la ética de la inteligencia artificial. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Protección Civil. (s.f.). *Sistema de Avisos a la Población (Public Warning System)*.

Ministerio del Interior.

<https://www.proteccioncivil.es/coordinacion/redes/ran/public-warning-system>

Ramírez, J., Pérez, M., & González, L.(2021). Predicting wildfire occurrence using Machine Learning techniques. *Journal of Environmental Management*, 270, 110843.

Rokhideh, M., Fearnley, C., & Budimir, M. (2025). Multi-Hazard Early Warning Systems in the Sendai Framework for Disaster Risk Reduction: Achievements, Gaps, and Future Directions. *International Journal of Disaster Risk Science*, 16(1).

<https://doi.org/10.1007/s13753-025-00622-9>

Scikit-learn developers. (s.f.). *sklearn.metrics.average_precision_score*. Scikit-learn.

Recuperado el 4 de febrero de 2026, de https://scikit-learn.org/stable/modules/generated/sklearn.metrics.average_precision_score.html

!

Servicio Nacional de Prevención y Respuesta ante Desastres [SENAPRED]. (2024).

Manual del Sistema de Alerta de Emergencias: Monitoreo, alertamiento y producción de información operacional (Código: MO-MAC-02; Versión 07).

Ministerio del Interior y Seguridad Pública, Gobierno de Chile.

Servicio Nacional de Prevención y Respuesta ante Desastres. (s.f.). *Home*.

<https://web.senapred.cl/>

Smith, A., & Clark, B. (2020). Evac-Cast: An interpretable machine learning framework for real-time evacuation recommendations. *Natural Hazards Review*, 21(3), 04020030.

Tiggeloven, T., Pfeiffer, S., Matanó, A., van den Homberg, M., Thalheimer, L., Reichstein, M. y Torresan, S. (2025). The role of artificial intelligence for early warning systems: Status, applicability, guardrails, and ways forward. *iScience*, 28(113689).
<https://doi.org/10.1016/j.isci.2025.113689>

Turban, E., Sharda, R., Delen, D., y Liang, T.-P. (2010). *Decision Support and Business Intelligence Systems* (9ª ed.). Pearson.
<https://es.scribd.com/document/967645736/Decision-Support-And-Business-Intelligence-Systems-9th-Edition-Dursun-Delen-Efraim-Turban-ebook-pdf-available>

Universidad de Chile. (14 de diciembre de 2022). *¿Qué es el "Factor 30-30-30" que favorece la propagación de incendios?* Noticias Universidad de Chile.
<https://uchile.cl/noticias/200220/que-es-el-factor-30-30-30-que-favorece-la-propagacion-de-incendios>

XGBoost developers. (2025). *XGBoost Parameters*. XGBoost Documentation.
<https://xgboost.readthedocs.io/en/stable/parameter.html>



ANEXOS

Anexo A. Oficio N° 273/2026 – Servicio Nacional de Prevención y Respuesta ante Desastres.

	OFICIO N° 273 ANT.: Solicitud de acceso a información N° AB004T0001847, de fecha 05 de febrero de 2026, ingresa por el Sr. Francisco Javier Páez Ramírez. MAT.: Respuesta a solicitud de información AB004T0001847. SANTIAGO, 16 de Febrero de 2026
DE: DIRECTORA NACIONAL A : FRANCISCO PAEZ RAMIREZ Mediante el documento indicado en el ANT., este organismo ha tomado conocimiento de la solicitud de información pública realizada por usted, en los términos que indica. Al efecto, usted con fecha 05 de febrero de 2026, ingresó al Servicio Nacional de Prevención y Respuesta ante Desastres ("SENAPRED"), la consulta AB004T0001847, cuyo tenor es el siguiente: <i>"Se solicita información correspondiente al registro de envío de mensajería del Sistema de Alerta de Emergencia (SAE) asociada a eventos de incendios forestales, para el periodo 2019-2025. En particular, se requiere contar con los registros históricos que permitan identificar y caracterizar cada mensaje emitido, considerando las siguientes variables: Fecha de emisión del mensaje, Hora de emisión del mensaje, Tipo de mensaje SAE, Texto del mensaje enviado, Región y polígono de difusión asociado al mensaje. La información solicitada tiene por finalidad el análisis histórico y estadístico de la activación del SAE en el contexto de incendios forestales, con fines de investigación y apoyo a la mejora de los procesos de toma de decisiones en la gestión del riesgo de desastres".</i> En primer término, cabe indicar que, con fecha 7 de agosto de 2021, se publicó en el Diario Oficial la Ley N°21.364, que, establece el Sistema Nacional de Prevención y Respuesta ante Desastres, sustituye la Oficina Nacional de Emergencia por el Servicio Nacional de Prevención y Respuesta ante Desastres, y adecúa normas que indica. Sobre el particular, cabe informar que, acorde con lo previsto en el artículo 1° de la referida ley N°21.364, se crea el Sistema Nacional de Prevención y Respuesta ante Desastres ("SINAPRED"), conformado por el "conjunto de entidades públicas y privadas con competencias relacionadas con las fases del ciclo del riesgo de desastres, que se organizan descentralizada o descentralizadamente y de manera escalonada, desde el ámbito comunal, provincial, regional y nacional, para garantizar una adecuada Gestión del Riesgo de Desastres, y comprende las normas, políticas, planes y otros instrumentos y procedimientos atinentes a la Gestión del Riesgo de Desastres". A su vez, según lo previsto en el artículo 19 del referido texto legal se crea el Servicio Nacional de Prevención y Respuesta ante Desastres (SENAPRED) "como un servicio público descentralizado, con personalidad jurídica y patrimonio propio, sometido a la supervigilancia del Presidente de la República a través del Ministerio del Interior y Seguridad Pública mediante la coordinación de la Subsecretaría del Interior, y que será el servicio encargado de asesorar, coordinar, organizar, planificar y supervisar las actividades relacionadas con la Gestión del Riesgo de Desastres del país". En este contexto, es dable informar que la normativa vigente establece la existencia de un Sistema de Alerta Temprana como un conjunto de capacidades necesarias para generar y difundir información de alerta que sea oportuna y significativa, para que las personas, las comunidades y las organizaciones expuestas a alguna amenaza se preparen y actúen de forma adecuada y con suficiente antelación, con el objeto de reducir la posibilidad de que se produzcan pérdidas o daños (artículo 38). En relación con lo anterior, el Sistema de Alerta de Emergencias (SAE) es un sistema de mensajería que tiene por objetivo informar o dar instrucciones de evacuaciones de forma oportuna, clara y precisa, para que las personas, las comunidades y las organizaciones expuestas a alguna amenaza se preparen y actúen de forma adecuada y con suficiente antelación, con el objeto de reducir la posibilidad que se produzcan pérdidas o daños. El SAE es una medida complementaria de refuerzo a otras medidas y acciones de evacuación frente a una emergencia, para alguna de las amenazas que se encuentran preestablecidas para el empleo de esta herramienta en los protocolos internos. Adicionalmente, el SAE se utiliza para el envío de mensajes preventivos para informar con antelación a la población sobre amenazas pronosticas o eventos en desarrollo que los podrían afectar.	



En respuesta a su requerimiento de información, mediante la cual solicitó el registro de envío de mensajería del Sistema de Alerta de Emergencia (SAE) asociada a eventos de Incendios Forestales, para el periodo 2019-2025. Incluyendo información acerca de la Fecha de emisión del mensaje, Hora de emisión del mensaje, Tipo de mensaje SAE, Texto del mensaje enviado, Región y polígono de difusión asociado al mensaje, este Servicio informa que se acoge su solicitud y, en consecuencia, adjunta dicha información requerida.

Se remite respuesta de la solicitud al correo fpaez7@gmail.com.

Saluda atentamente a Ud.,



Alicia Cebrian Lopez
Directora Nacional



Para verificar documento ingresar en la siguiente url <https://validadoc.interior.gob.cl/>

Código Verificación: WHVY8TCgTrjQT7K5Y6LEpA==

FGJ/FOC/MOP/FPC/ARO/aro

ID DOC : 21951353

Distribución:

1. FRANCISCO PAEZ RAMIREZ (Mail: fpaez7@gmail.com)
2. /SENAPRED CENTRAL/Subdirección Operaciones de Emergencias
3. /SENAPRED CENTRAL/Subdirección Operaciones de Emergencias/Departamento del Sistema de Alerta Temprana
4. /SENAPRED CENTRAL/Gabinete Dirección/Opto. Comunicaciones y Difusión/Unidad de Transparencia y Atención Ciudadana
5. /SENAPRED CENTRAL/Gabinete Dirección/Opto. Asesoría Jurídica
6. /SENAPRED CENTRAL/Dirección Nacional
7. FRANCISCO PAEZ RAMIREZ (Mail: fpaez7@gmail.com)
8. SENAPRED CENTRAL/Subdirección de Personas, Administración y Finanzas/Unidad de Oficina de Partes



UNIVERSIDAD
SAN SEBASTIAN

Ilumina el futuro

Anexo B. Carta Oficial N°80/2026 – Corporación Nacional Forestal



CORPORACIÓN NACIONAL FORESTAL
OFICINA CENTRAL
ABA/EBZ/DPN/JGO

CARTA OFICIAL N° 80/2026

SANTIAGO, 12/02/2026

SEÑOR
FRANCISCO PÁEZ RAMÍREZ

Junto con saludar y en respuesta a su requerimiento de información ciudadana N° AR003T-00011047 del 06 de febrero de 2026, mediante la cual solicitó datos históricos de incendios forestales para el período 2019 a 2026, incluyendo información acerca de región, provincia, comuna, nombre del incendio, tipo y nivel de alerta, combustible inicial, superficie afectada, fecha y hora de inicio del incendio, duración del incendio (horas), día de inicio, temperatura, nubosidad, humedad relativa, exposición, dirección del viento, velocidad del viento, topografía, pendiente, latitud y longitud.

Sobre el particular, esta Dirección Ejecutiva indica a usted que se ha enviado a su dirección de correo electrónico, archivo Excel con los antecedentes requeridos en su solicitud de información ciudadana.

Le saluda atentamente,

RODRIGO ANDRÉS ILLESCA ROJAS
DIRECTOR EJECUTIVO
CORPORACIÓN NACIONAL FORESTAL

Incl.: Documento Digital: Solicitud AR003T0011047
Documento Digital: Base Datos IIFF 2018-2025