

**Universidad del Gran Rosario**  
**Licenciatura en Ciencia de Datos**



**Análisis de Factores de Riesgo de Incendios**  
**Forestales en la provincia de Córdoba**

Autor (es)

Angelloti Dana, angellottidana@gmail.com

Álvarez Analía, alvarez.analiac@gmail.com

Rango Ezequiel, ezequielrango29@gmail.com

Zanor Mónica, mzanor@gmail.com

Equipo de cátedra de Metodología de la Investigación

Lic. María Cecilia Ribecco, profesoramceciariibecco@hotmail.com

Ing. Pérez Matías, m.perez@ugr.edu.ar

**Junio 2025**

## Resumen

Este trabajo aborda el análisis de los factores de riesgo asociados a los incendios forestales en la provincia de Córdoba, una problemática de gran impacto ambiental, económico y social. La investigación se enmarca en el uso de herramientas de ciencia de datos aplicadas a la gestión ambiental y al análisis de eventos críticos. A partir de un enfoque cuantitativo, se recopilan y procesan variables meteorológicas clave —como la temperatura, humedad relativa y velocidad del viento— junto con indicadores ambientales como el NDVI (Índice de Vegetación) y el FWI (Índice de Riesgo de Incendio). Mediante la implementación de modelos de aprendizaje automático, como Regresión Logística y Random Forest, se construyen modelos predictivos capaces de identificar con alta precisión las condiciones propicias para la ocurrencia de incendios. La evaluación de los modelos incluye métricas como la precisión y el análisis de la matriz de confusión. Esta investigación contribuye al desarrollo de herramientas de apoyo a la toma de decisiones para la prevención de incendios, y refuerza la necesidad de integrar tecnologías predictivas en los sistemas de monitoreo ambiental.

**Palabras claves:** incendios forestales, indicadores meteorológicos, factores de riesgo, aprendizaje automático.

## Abstract

This study addresses the analysis of risk factors associated with wildfires in the province of Córdoba, a problem with significant environmental, economic, and social impact. The research is framed within the use of data science tools applied to environmental management and the analysis of critical natural events. Using a quantitative approach, key meteorological variables—such as temperature, relative humidity, and windspeed—are collected and processed, along with environmental indicators such as NDVI (Normalized Difference Vegetation Index) and FWI (Fire Weather Index). Machine learning models, including Logistic Regression and Random Forest, are implemented to build predictive models capable of accurately identifying conditions conducive to wildfire occurrence. Model evaluation includes metrics such as accuracy and confusion matrix analysis. This research contributes to the development of decision-support tools for fire prevention and emphasizes the importance of integrating predictive technologies into environmental monitoring systems..

**Keywords:** wildfire, meteorological indicators, risk factors, machine learning.

# Índice

Resumen.....	1
Abstract.....	1
Índice .....	2
2. Revisión Bibliográfica .....	6
a. El Manejo del Fuego en Argentina:.....	6
b. Etapas de un Incendio: .....	6
c. Incendios Forestales: Definición y Clases.....	7
d. Comportamiento del Fuego .....	7
e. Mapas de Riesgos de Incendios .....	8
f. Fire Weather Index (FWI) .....	8
g. Componentes del FWI .....	8
2.1 Antecedentes .....	10
2.2 Marco conceptual .....	11
a. Teoría del Riesgo Ambiental .....	11
b. Teoría de los Sistemas Complejos .....	11
c. Modelado predictivo y Ciencia de Datos .....	11
3. Formulación del Problema y Planteamiento de Hipótesis.....	13
a. Hipótesis específicas .....	13
b. Metodología y observabilidad .....	14
c. Consideración sobre Escalabilidad .....	14
4. Metodología de la Investigación .....	15
4.1 Enfoque y tipo de investigación.....	15
4.2 Variables e indicadores .....	15
4.3 Población y muestra.....	15
4.4 Fuentes y recolección de datos .....	16
4.5 Procesamiento y limpieza de datos.....	16
4.6 Modelado predictivo .....	16
4.7 Proceso de entrenamiento del modelo.....	17
4.8 Validación del modelo .....	17
4.9 Validación externa y comparación .....	18
4.10 Consideraciones éticas y reproducibilidad .....	18
5. Recolección y Tratamiento de los Datos.....	19
5.1 Rol de los Datos en la Investigación.....	19

5.2 Fuentes y Tipos de Datos Utilizados .....	19
a. Datos Climáticos (Cuantitativos) .....	19
5.3 Alcances de la Investigación.....	20
5.4 Consideraciones Éticas.....	20
5.5 Limpieza de datos y Procesamiento .....	21
a. Gráficos de la información obtenida.....	22
b. Limpieza de datos de incendios forestales .....	22
c. Procesamiento de variables meteorológicas .....	24
5.6 Análisis Estadístico y Modelado Predictivo .....	25
5.7 Métodos:.....	27
a. Análisis de correlación y regresión.....	27
b. Modelos de clasificación: árboles de decisión, Random Forest.....	28
c. Validación cruzada y métricas: accuracy, F1-score, recall. ....	28
d. Pasos básicos de la validación cruzada .....	29
6. Resultados/comparación de modelos.....	31
6.1 Comparación entre modelos predictivos.....	31
6.2 Interpretación práctica de los resultados .....	31
6.3 Conexiones con teorías del riesgo y sistemas complejos.....	32
6.4 Random Forest .....	33
6.5 Validación cruzada:.....	34
6.6 Regresión logística:.....	35
6.7 Máquinas de soporte vectorial (SVM).....	38
7. Conclusiones .....	39
8. Referencias .....	41
9. Anexos .....	42
Anexo I: Diagrama del Proceso Metodológico .....	42
Anexo II: Variables utilizadas y fuentes.....	43
Anexo III: Resultados y Validación del Modelo .....	44
Anexo IV: Código y Reproducibilidad.....	45
Anexo V: Declaración Ética sobre el Uso de IA .....	47
Anexo VI: Entrevista .....	48
Anexo VII: Contacto.....	49
Anexo VIII: Acceso al notebook Colab .....	50

## 1. Introducción

Los incendios forestales representan una amenaza ambiental creciente, especialmente en regiones con climas extremos, vegetación densa y alta presión antrópica, como la provincia de Córdoba, Argentina. El avance del cambio climático, la expansión de la frontera agrícola y la urbanización intensifican las condiciones de riesgo, provocando pérdidas de biodiversidad, degradación del suelo, y afectaciones económicas y sociales a las comunidades.

A pesar de la existencia de sistemas de monitoreo y alerta, aún persiste una falta de herramientas predictivas robustas que integren variables climáticas, geográficas y humanas. Esta limitación obstaculiza la prevención efectiva y la planificación estratégica frente a estos eventos. En este contexto, la presente investigación se propone analizar los factores de riesgo asociados a la ocurrencia de incendios forestales mediante el uso de técnicas de ciencia de datos. Se busca construir un modelo predictivo capaz de identificar áreas críticas en la provincia, anticipar la probabilidad de incendios y orientar políticas públicas de mitigación y prevención.

Entre los objetivos específicos, se destacan

- ) Evaluar la influencia de variables climáticas como la temperatura, la humedad relativa, la velocidad del viento y la lluvia acumulada.
- ) Analizar el uso del Fire Weather Index (FWI) como predictor.
- ) Probar modelos de clasificación como árboles de decisión, regresión logística y random forest para determinar su capacidad predictiva.
- ) Contribuir al diseño de estrategias de prevención focalizadas y a la toma de decisiones en gestión ambiental.

La implementación de modelos predictivos no solo fortalecerá la capacidad de anticipación, sino que también permitirá involucrar activamente a autoridades locales y comunidades en campañas de concientización y prevención. Finalmente, los resultados podrían tener un impacto extendido hacia otras regiones con características similares, favoreciendo una gestión ambiental más eficiente y sostenible en el largo plazo.

Además, los hallazgos y metodologías de esta investigación podrían ser adaptados a otras provincias argentinas o regiones de América Latina que enfrentan desafíos similares, contribuyendo a una estrategia regional más robusta para la prevención y gestión de incendios forestales.

## 2. Revisión Bibliográfica

La preocupación por los incendios forestales ha sido objeto de innumerables estudios en diversas regiones del mundo. En este contexto, se hace fundamental revisar la literatura que aborda tanto los factores que contribuyen al riesgo de incendios como las metodologías utilizadas para su análisis y modelado.

### a. El Manejo del Fuego en Argentina:

**Servicio Nacional de Manejo del Fuego (SNMF):** Es el organismo nacional encargado de coordinar los recursos para el combate de incendios.

**Ley de Manejo del Fuego:** Establece las bases legales para la prevención y combate de incendios.

**Niveles de Actuación:** El combate de incendios se organiza en tres niveles: jurisdiccional (nivel 1), regional (nivel 2) y extrarregional (nivel 3).

**Plan Nacional de Manejo del Fuego:** Es un plan estratégico para la gestión de incendios forestales en Argentina.

**Servicios Provinciales de Manejo del Fuego:** Cada provincia tiene su propio servicio para la prevención, presupresión y combate de incendios.

**Dirección de Lucha contra los Incendios Forestales y Emergencias (DLIFE):** Establece lineamientos y estrategias para la prevención y combate de incendios en los parques nacionales.

### b. Etapas de un Incendio:

Se identifican etapas como calentamiento, inicio, desarrollo, estabilización y extinción.

**Métodos de Extinción:** Se pueden usar métodos como eliminar el combustible, sofocar el comburente, enfriar la energía o inhibir la reacción en cadena.

**Prevención:** Es fundamental evitar el uso irresponsable del fuego y tomar medidas para prevenir incendios.

**Riesgos:** Los incendios forestales pueden causar graves daños al medio ambiente, a la seguridad de la población y a la economía.

### c. Incendios Forestales: Definición y Clases

Según la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), un incendio forestal se define como un fuego incontrolado que se propaga en áreas de vegetación (FAO, 2021). Se clasifican habitualmente en tres tipos principales:

**Incendios naturales:** causados por fenómenos atmosféricos, como rayos.

**Incendios antropogénicos:** provocados por actividades humanas, como quemas agrícolas.

**Incendios accidentales:** resultado de descuidos, como fogatas mal apagadas.

El contexto específico de cada región influye en cómo se manifiestan y gestionan estos incendios.

El comportamiento del fuego y la evaluación del riesgo de incendios son temas complejos que han recibido atención significativa en la literatura científica. En este contexto, el desarrollo de herramientas como el Fire Weather Index(FWI) ha ayudado a mejorar la comprensión y predicción de los incendios forestales.

### d. Comportamiento del Fuego

El comportamiento del fuego está determinado por una interacción compleja de factores físicos y biológicos. Alexander y Cruz (2021) destacan que el fuego se propaga principalmente por tres mecanismos: conducción, convección y radiación.

Cada uno de estos modos puede influir en la velocidad de propagación y la intensidad del fuego, así como en la forma en que afecta a la vegetación circundante.

**Factores que Afectan el Comportamiento del Fuego:** los factores que influyen en el comportamiento del fuego son variaciones en el tipo de vegetación, la humedad del combustible, y las condiciones climáticas, entre otros. Valderas et al. (2020) plantean que los combustibles secos y una alta temperatura ambiental pueden aumentar exponencialmente la velocidad del fuego. Investigaciones más recientes, como la de Pérez et al. (2023), se centran en cómo las condiciones del viento interactúan con la topografía para influir en la dirección y velocidad del fuego.



### e. Mapas de Riesgos de Incendios

Los mapas de riesgos son herramientas esenciales para la gestión y mitigación de incendios. Están diseñados para visualizar zonas potencialmente vulnerables, ayudando a la planificación y respuesta ante incendios. Según García et al. (2019), la elaboración de estos mapas implica la integración de múltiples variables, incluyendo factores ambientales, uso del suelo y antecedentes históricos de incendios.

**Metodologías para la Elaboración de Mapas de Riesgo:** la metodología más común para crear mapas de riesgo se basa en análisis geoespaciales. Rojas y Fernández (2021) desarrollaron un enfoque que combina GIS (sistemas de información geográfica) y modelos estadísticos para evaluar el riesgo de incendios en áreas forestales de España. Este tipo de trabajo se puede aplicar a la provincia de Córdoba, ayudando a identificar zonas en riesgo.

### f. Fire Weather Index(FWI)

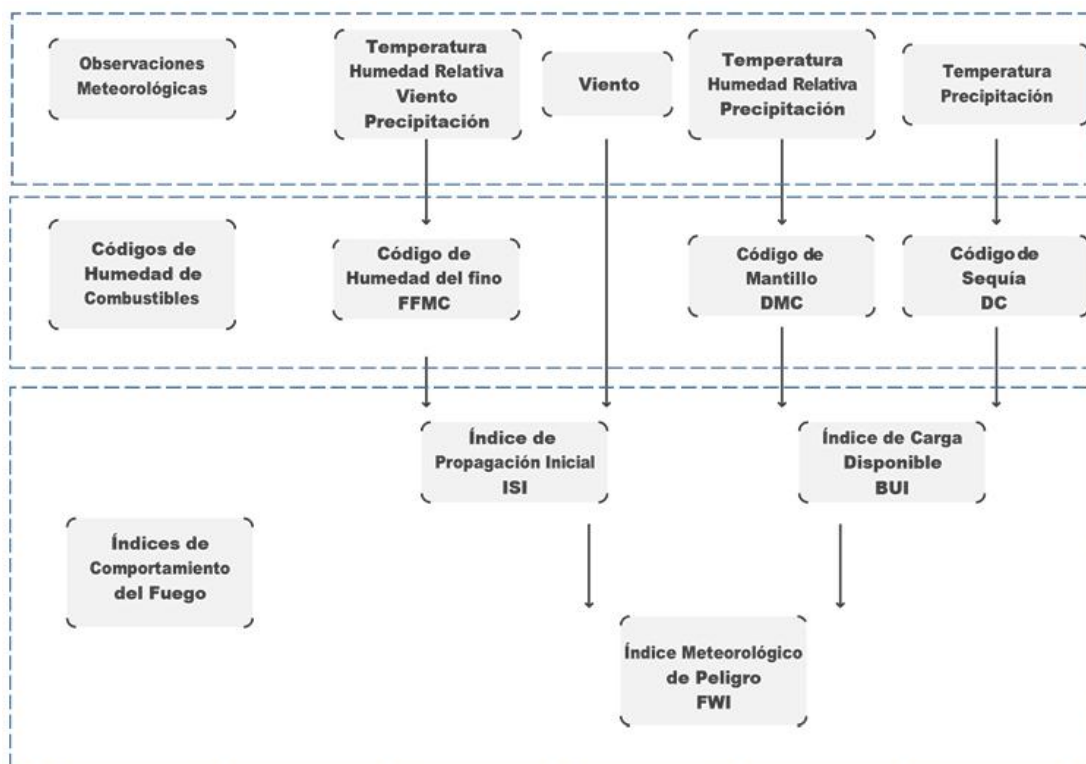
El Fire Weather Index(FWI) es un sistema ampliamente utilizado para evaluar las condiciones meteorológicas que afectan el riesgo de incendios. Este índice se basa en la combinación de varios factores climáticos, como temperatura, humedad relativa, velocidad del viento y precipitación.

Este índice que comenzó a implementarse en Argentina desde el año 2000, es el resultado de la combinación de códigos que estiman el estado de los combustibles muertos de diferentes diámetros, e índices que representan la velocidad de propagación de incendios forestales y la carga de combustible, medio y grueso, disponible para arder.

Los valores finales de FWI que se obtienen, varían dependiendo del clima de la región, por lo que un mismo valor de FWI no puede interpretarse del mismo modo en cada lugar. Para poder comparar el peligro a lo largo del país se calculan estadísticamente umbrales para cada zona y así surgen “las clases de peligro”.

### g. Componentes del FWI

La estructura del índice es la siguiente, con las variables de entrada y con los respectivos aspectos que explica cada indicador:



**FFMC:** se asocia a ignición, ocurrencia de focos antrópicos.

**DMC:** se asocia a ocurrencia de focos por rayos; grado de aporte a la intensidad del fuego, por involucrarse los primeros centímetros de suelo y los combustibles medios.

**DC:** se asocia a fuegos de rescoldo en suelo profundo y troncos de gran diámetro, dificultades de liquidación; a mayores valores, mayor intensidad del fuego por involucrarse combustibles pesados.

**ISI:** se asocia a velocidad relativa de propagación.

**BUI:** se asocia a intensidad del fuego por aporte de capas de suelo orgánico y de combustible muertos medios y pesados.

**FWI:** se asocia a altura de llama, coronamientos, comportamiento extremo, dificultades de control.

**RH:** En meteorología, RH significa humedad relativa. Es una medida que expresa el contenido de vapor de agua en el aire en relación con la cantidad máxima que el aire puede contener a una temperatura determinada. Se expresa como porcentaje y es un factor importante en el clima y el confort humano.

Investigaciones como las de Wotton et al. (2017) han demostrado que el FWI es útil para predecir el comportamiento del fuego en diversas condiciones, mostrando correlaciones significativas con la incidencia de incendios. En Córdoba, el uso del FWI podría ayudar a los gestores ambientales a anticipar incendios y planificar medidas preventivas efectivas.

Se recurre a un conjunto de teorías y enfoques que explican cómo interactúan los factores ambientales, climáticos, geográficos y humanos en la generación de incendios forestales, y cómo estas interacciones pueden ser modeladas y anticipadas.

## 2.1 Antecedentes

Los incendios forestales constituyen una de las principales amenazas ecológicas y socioeconómicas en zonas de vegetación densa sometidas a presión antrópica. Estudios globales coinciden en que el cambio climático ha intensificado la frecuencia e intensidad de estos eventos, exacerbando sus impactos ambientales, económicos y sociales [1].

En Argentina, investigaciones recientes han identificado una tendencia creciente en la ocurrencia de incendios en la región central, en particular en Córdoba, debido a la combinación de sequías prolongadas, aumento de temperatura, y expansión urbana sobre zonas forestales [2]. Diversos trabajos han demostrado la utilidad del **Fire Weather Index (FWI)** como predictor de condiciones propicias para la ignición y propagación del fuego [3].

Por otro lado, autores como Wotton et al. [4] y Valderas et al. [5] han desarrollado modelos de predicción basados en datos meteorológicos y geográficos que integran inteligencia artificial y análisis geoespacial. Estos enfoques han sido adaptados en Europa, Canadá y Australia, y se presentan como referencia para su aplicación en ecosistemas como los del centro argentino.

Asimismo, el uso de modelos de aprendizaje automático (ML) como Random Forest, Árboles de Decisión y Regresión Logística ha demostrado eficacia en contextos

---

[1] L. García, N. Martínez, y A. Gómez, "Riskmappingforwildfires in Mediterraneanregions," Int. J. WildlandFire, vol. 28, 2019.

[2]Ministerio de Ambiente y Desarrollo Sostenible, "Informe de incendios forestales en Córdoba," Argentina, 2022.

[3] FAO, "Definiciones y clasificaciones de incendios forestales," Organización de las Naciones Unidas para la Agricultura y la Alimentación, Roma, 2021.

[4] B. M. Wotton, D. L. Martell, y K. A. Logan, "TheFireWeatherIndex: A comprehensive evaluation," Can. J. For. Res., vol. 47, pp. 1–12, 2017.

[5] M. Valderas, H. Domínguez, y J. Reyes, "Analysisoffirepropagationmechanisms in forestecosystems," J. Environ. Manag., vol. 250, 2020.

similares, permitiendo anticipar zonas de riesgo y optimizar recursos en la gestión de incendios forestales [6].

## 2.2 Marco conceptual

El presente estudio se apoya en tres pilares conceptuales:

### a. Teoría del Riesgo Ambiental

El riesgo se define como una función de tres componentes: **amenaza, exposición y vulnerabilidad** [7]. Esta formulación es especialmente útil para evaluar incendios forestales, ya que considera tanto las condiciones del entorno (combustibles vegetales, topografía) como los factores humanos y climáticos que inciden en la generación del fuego. Bajo este enfoque, se busca identificar zonas críticas para priorizar la intervención y mitigación.

### b. Teoría de los Sistemas Complejos

Los incendios forestales son fenómenos no lineales y multicausales, donde variables físicas, biológicas y humanas interactúan de forma dinámica. Desde esta perspectiva, el sistema puede presentar **comportamientos emergentes** no previsibles desde el análisis de cada factor por separado [8]. El uso de algoritmos de ML permite modelar estas interacciones complejas, superando las limitaciones de métodos lineales clásicos.

### c. Modelado predictivo y Ciencia de Datos

El enfoque de Ciencia de Datos se centra en el uso de **algoritmos supervisados** para construir modelos que aprendan de patrones históricos. En este trabajo, se emplean modelos como **Random Forest, Regresión Logística y SVM**, los cuales permiten no solo anticipar condiciones de riesgo sino también interpretar la influencia

---

[6] S. Rojas y E. Fernández, "Geospatial analysis for fire risk mapping in forest areas," *Environ. Model. Softw.*, vol. 139, 2021.

[7] A. Lavell, M. Oppenheimer, y M. Cardona, "Conceptual framework for risk," *United Nations Office for Disaster Risk Reduction (UNDRR)*, Geneva, 2005.

[8] M. E. Alexander y M. G. Cruz, "Fire behavior and fire safety: A global perspective," *Fire Saf. J.*, vol. 119, 2021.

relativa de cada variable. Estos modelos son evaluados con técnicas de validación cruzada y métricas como *accuracy*, *F1-score* y *recall*, buscando robustez y generalización.

Si bien el FWI no tiene cortes absolutos, tomamos como referencia los umbrales propuestos por el Plan Nacional de Manejo del Fuego, y ajustamos la clasificación a nuestra distribución de datos. Vimos que los incendios en Córdoba tienden a aumentar significativamente a partir de valores mayores a 10, por lo que usamos ese punto como corte entre riesgo bajo y moderado/alto.

Además de los modelos predictivos, se incorporaron herramientas geoespaciales (GIS) y datos satelitales (NASA FIRMS, NDVI), lo que permitió **mapear espacialmente el riesgo de incendios**, detectar zonas críticas y generar salidas visuales que pueden utilizarse en aplicaciones prácticas y en tiempo real.”

Este enfoque, basado en datos abiertos, es escalable y reproducible, lo que refuerza su valor como herramienta de **alerta temprana, planificación territorial y toma de decisiones**.

### 3. Formulación del Problema y Planteamiento de Hipótesis

**Problema general:** ¿Cuáles son los principales factores que influyen en la ocurrencia de incendios forestales y cómo pueden utilizarse para anticipar y mitigar estos eventos mediante un enfoque científico y preventivo?

**Objetivo General:** Analizar los factores climáticos, geográficos y antrópicos que inciden en la ocurrencia de incendios forestales en la provincia de Córdoba, con el fin de desarrollar estrategias de mitigación en zonas vulnerables mediante el uso de modelos predictivos basados en ciencia de datos.

**Hipótesis general:** Los incendios forestales en la provincia de Córdoba están asociados a variables climáticas (como temperatura y humedad), geográficas (como topografía y tipo de cobertura vegetal) y antrópicas (como el uso del suelo y la densidad poblacional)

#### a. Hipótesis específicas

Hipótesis 1:

**"El aumento de la temperatura y la disminución de la humedad relativa incrementan la probabilidad de la incidencia de incendios forestales en la provincia de Córdoba."**

Justificación: condiciones climáticas extremas son ampliamente reconocidas como factores desencadenantes del fuego. Analizar la relación histórica entre temperatura, humedad y número de focos activos permitirá validar esta asociación.

Hipótesis 2:

**"El uso del Fire Weather Index(FWI) puede predecir con alta precisión la probabilidad de incendios en diferentes regiones de la provincia de Córdoba."**

Justificación: El FWI es una herramienta consolidada a nivel internacional. Evaluar su comportamiento en el contexto local permitirá verificar su aplicabilidad, ajustar sus parámetros y mejorar la predicción del riesgo.

Hipótesis 3:

**“En épocas de menor lluvia acumulada y mayor velocidad del viento se incrementa el riesgo de incendios forestales.”**

Justificación: La menor lluvia seca la vegetación, aumentando el material combustible, mientras que el viento favorece el inicio y expansión del fuego. Verificar esta relación permitirá fortalecer estrategias de prevención estacional.

#### b. Metodología y observabilidad

Estas hipótesis serán puestas a prueba mediante el uso de técnicas estadísticas y modelos de aprendizaje automático, utilizando datos históricos de incendios y variables ambientales correspondientes al período 2017 a la actualidad. Los hechos y fenómenos analizados son observables y medibles a través de fuentes oficiales (como el Ministerio de Ambiente y Cambio Climático, el SMN y organismos provinciales), datos satelitales (MODIS, VIIRS), estaciones meteorológicas automáticas, y bases de datos georreferenciadas.

#### c. Consideración sobre Escalabilidad

Si bien el estudio se centra en la provincia de Córdoba, el enfoque metodológico y los modelos desarrollados podrían adaptarse a otras regiones de Argentina o América Latina con características ambientales y sociales similares. Esto potenciaría la replicabilidad del análisis y su aporte a políticas públicas de gestión del riesgo de incendios a nivel nacional y regional.

## 4. Metodología de la Investigación

### 4.1 Enfoque y tipo de investigación

La presente investigación se enmarca en el **paradigma positivista**, con un enfoque **cuantitativo y explicativo**, orientado a la **predicción** de eventos ambientales críticos —en este caso, incendios forestales— a partir del análisis de variables climáticas, geográficas y antrópicas.

El estudio es de tipo **no experimental, longitudinal y correlacional**, dado que se trabaja con datos históricos y no se manipulan las variables de estudio. El objetivo es identificar relaciones entre variables y generar modelos predictivos basados en técnicas de ciencia de datos.

### 4.2 Variables e indicadores

Las variables consideradas en el modelo incluyen:

Categoría	Variables	Fuente
<b>Climáticas</b>	Temperatura (°C), Humedad relativa (%), Velocidad del viento (km/h), Precipitación acumulada (mm), (FWI) y subcomponentes (FFMC, DMC, DC, ISI, BUI)	SMN, INTA, NASA FIRMS
<b>Geográficas</b>	Tipo de cobertura vegetal	IGN, imágenes satelitales
<b>Variable objetivo</b>	Riesgo de incendio (bajo vs. riesgo moderado/alto)	Plan Nacional de Manejo del Fuego

### 4.3 Población y muestra

La **población** está constituida por todas las áreas con cobertura vegetal de la provincia de Córdoba susceptibles a incendios. La **muestra** corresponde a registros georreferenciados de focos de incendio entre los años **2017 y 2023**, asociados con



datos meteorológicos y de uso del suelo. La muestra fue seleccionada en función de su **cobertura temporal y espacial**, garantizando representatividad regional y estacional.

#### 4.4 Fuentes y recolección de datos

Se utilizaron fuentes **oficiales y de libre acceso**:

**Meteorológicos:** Servicio Meteorológico Nacional (SMN), INTA.

**Satelitales:** NASA FIRMS, MODIS.

**Geográficos:** IGN, capas SIG.

**Historial de incendios:** Ministerio de Ambiente y Cambio Climático, reportes provinciales.

Los datos se descargaron en formato .csv y .shp, y se integraron mediante herramientas Python (pandas, geopandas) y plataformas GIS (QGIS).

#### 4.5 Procesamiento y limpieza de datos

**Limpieza:** Se eliminaron duplicados y valores nulos. Se imputaron datos faltantes mediante interpolación lineal o valores medios por estación.

**Transformación:** Conversión de fechas, codificación de variables categóricas, estandarización de unidades.

**Integración geoespacial:** Se asociaron los focos de incendios con sus respectivas condiciones meteorológicas y geográficas mediante coordenadas y buffers de proximidad.

#### 4.6 Modelado predictivo

Se entrenaron y evaluaron tres algoritmos:

Modelo	Descripción breve
Regresión Logística	Modelo interpretable para clasificación binaria.
Árbol de Decisión	Modelo jerárquico con reglas explícitas.

Random Forest	Conjunto de árboles, robusto y con alta precisión.
SVM ( <b>exploratorio</b> )	Clasificador eficiente para datos no lineales.

---

## 4.7 Proceso de entrenamiento del modelo

**División de datos:** Se dividieron en entrenamiento y prueba con un 80% para entrenamiento y 20% para prueba, usando `train_test_split` con estratificación por clases.

**Proporciones:** 80% entrenamiento / 20% prueba.

**Balanceo de clases:** Se aplicó técnicas de balanceo mediante SMOTE para abordar desequilibrios en las clases de riesgo de incendio.

**Modelos utilizados:** Random Forest, regresión logística y SVM.

**Parámetros ajustados:**

Para Random Forest: `n_estimators` (100, 200), `max_depth` (None, 10, 20), mediante `GridSearch` con validación cruzada (`cv=5`).

Se utilizó `GridSearch` para optimizar los hiperparámetros.

**Validación:** Validación cruzada de 5 pliegues se empleó para evaluar los modelos.

**Indicadores de rendimiento:** Accuracy, F1-score, recall, con reportes específicos en cada modelo, destacando alta precisión y balance en el rendimiento.

## 4.8 Validación del modelo

Se aplicó **validación cruzada (K=5)** para evaluar robustez. Las métricas utilizadas fueron:

**Accuracy:** Proporción de predicciones correctas.

**Recall:** Capacidad del modelo para detectar correctamente eventos positivos (incendios).

**F1-score:** Medida equilibrada entre precisión y recall.

**Matriz de confusión:** Análisis de verdaderos y falsos positivos/negativos.

**Importante:** En la clase minoritaria (riesgo bajo), se aplicaron técnicas de rebalanceo con SMOTE para mejorar el aprendizaje del modelo.

## 4.9 Validación externa y comparación

Se utilizó también un conjunto de datos del año 2023 para validar la estabilidad del modelo, logrando un desempeño consistente (ej., Random Forest con  $F1=0.92$ ).

Se compararon los resultados con reportes del Plan Nacional de Manejo del Fuego, encontrando correlación y coherencia en la distribución espacial y temporal de focos.

## 4.10 Consideraciones éticas y reproducibilidad

- ) Solo se utilizaron **datos públicos y anónimos**.
- ) No se trabajó con personas ni poblaciones vulnerables.
- ) El código fuente, notebooks de modelado y datasets procesados serán compartidos en repositorio abierto para promover la **reproducibilidad científica**, en línea con las recomendaciones de uso responsable de IA.

## 5. Recolección y Tratamiento de los Datos

### 5.1 Rol de los Datos en la Investigación

En esta investigación, los datos representan insumos fundamentales para analizar, comprender y predecir la ocurrencia de incendios forestales en la provincia de Córdoba. Estos pueden ser cuantitativos (mediciones numéricas) o cualitativos (clasificaciones, descripciones), y su análisis permite validar hipótesis, construir modelos predictivos y proponer estrategias de mitigación basadas en evidencia.

### 5.2 Fuentes y Tipos de Datos Utilizados

Se utilizarán diversos conjuntos de datos provenientes de fuentes oficiales y científicas reconocidas. Estas incluyen:

#### a. Datos Climáticos (Cuantitativos)

- ) Temperatura máxima y mínima diaria
- ) Humedad relativa
- ) Velocidad y dirección del viento
- ) Precipitaciones acumuladas
- ) Índices derivados, como el Fire Weather Index (FWI)

*Fuente: Servicio Meteorológico Nacional (SMN), estaciones locales, INTA.*

#### b. Datos Geográficos (Cuantitativos y Cualitativos)

- ) Tipo de cobertura vegetal (bosque, pastizal, matorral)
- ) Pendiente y altitud del terreno
- ) Cercanía a cuerpos de agua, rutas y centros urbanos

*Fuente: Instituto Geográfico Nacional (IGN), imágenes satelitales, herramientas SIG (GIS).*

#### c. Datos de Incendios Históricos (Cuantitativos)

- ) Número de focos detectados por año
- ) Superficie afectada (hectáreas)

- )] Coordinadas de los focos registrados
- )] Causa del incendio (natural, antrópica)

*Fuente: Plan Nacional de Manejo del Fuego, NASA FIRMS, reportes provinciales.*

## 5.3 Alcances de la Investigación

En términos temáticos,

### **Incluye:**

- )] Factores climáticos (temperatura, humedad, viento, precipitación)
- )] Factores geográficos y de uso del suelo
- )] Historial de incendios y patrones espaciales-temporales
- )] Aplicación del FWI como índice predictivo
- )] Análisis de correlación y generación de mapas de riesgo

### **Excluye:**

- )] Evaluación del impacto ecológico post-incendio a largo plazo
- )] Impactos económicos o psicológicos en comunidades
- )] Estudio de políticas públicas o legislación ambiental
- )] Intervención directa en campo
- )] Alcance Temporal

Se analizarán datos correspondientes a los últimos 5 a 9 años (2017 en adelante), con el fin de identificar patrones recientes relacionados con el cambio climático y la presión antrópica.

## 5.4 Consideraciones Éticas

- )] Todos los datos utilizados provienen de fuentes oficiales, públicas o de libre acceso (NASA FIRMS, SMN, INDEC, etc.).
- )] Se garantiza la confidencialidad en los registros georreferenciados.
- )] La investigación no implica contacto con personas ni afecta a comunidades de manera directa.
- )] Se promueve el uso responsable de la información, con fines científicos y de prevención ambiental.

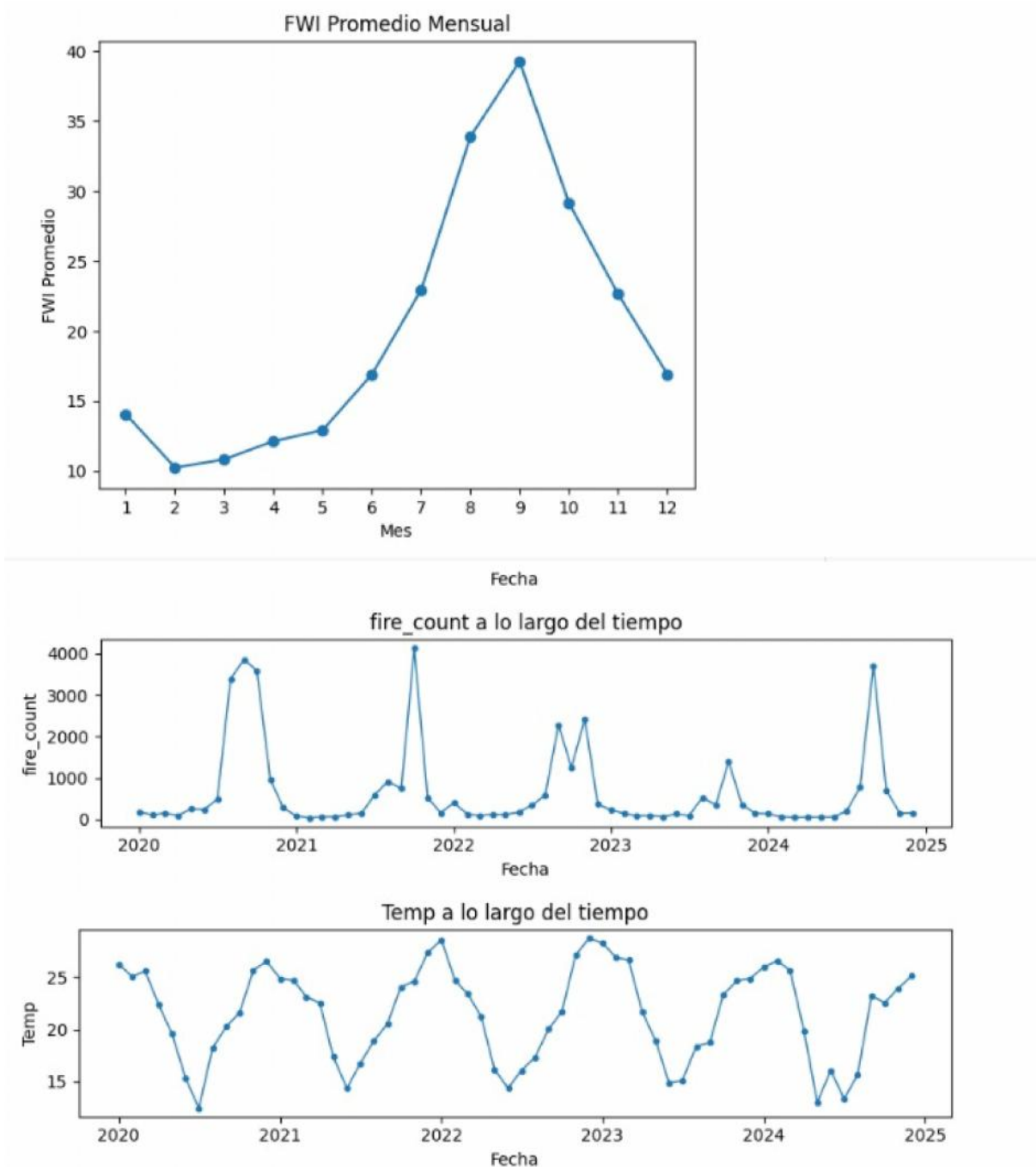
Además de las variables seleccionadas para el análisis principal, la base de datos incluye otros atributos potencialmente relevantes como la dirección del viento, tipo de suelo o cercanía a cuerpos de agua. Si bien no fueron incorporados en los modelos actuales por cuestiones de alcance, limitaciones metodológicas o disponibilidad incompleta, podrían ser considerados en estudios posteriores para enriquecer el análisis.

## 5.5 Limpieza de datos y Procesamiento

Preparación de datos para análisis exploratorio, eliminación de valores atípicos, imputación de datos faltantes, integración geoespacial.

El proceso de limpieza y preparación de datos fue fundamental para garantizar la calidad del análisis y la construcción de modelos predictivos confiables. Se trabajó con dos fuentes principales: el registro histórico de incendios del Ministerio de Ambiente y Desarrollo Sostenible de Argentina, y un conjunto de variables meteorológicas asociadas a la provincia de Córdoba.

a. Gráficos de la información obtenida



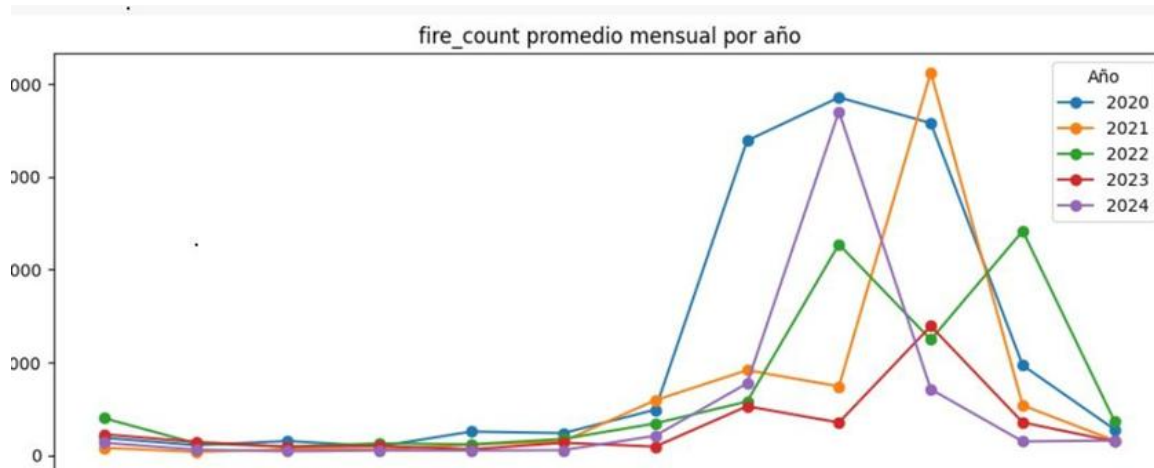
b. Limpieza de datos de incendios forestales

La base de datos de incendios presentaba registros anuales por jurisdicción, con algunas celdas identificadas como "s/d" (sin dato). Estas entradas fueron reemplazadas por valores nulos (NaN) utilizando la función replace, para su posterior tratamiento como datos faltantes.

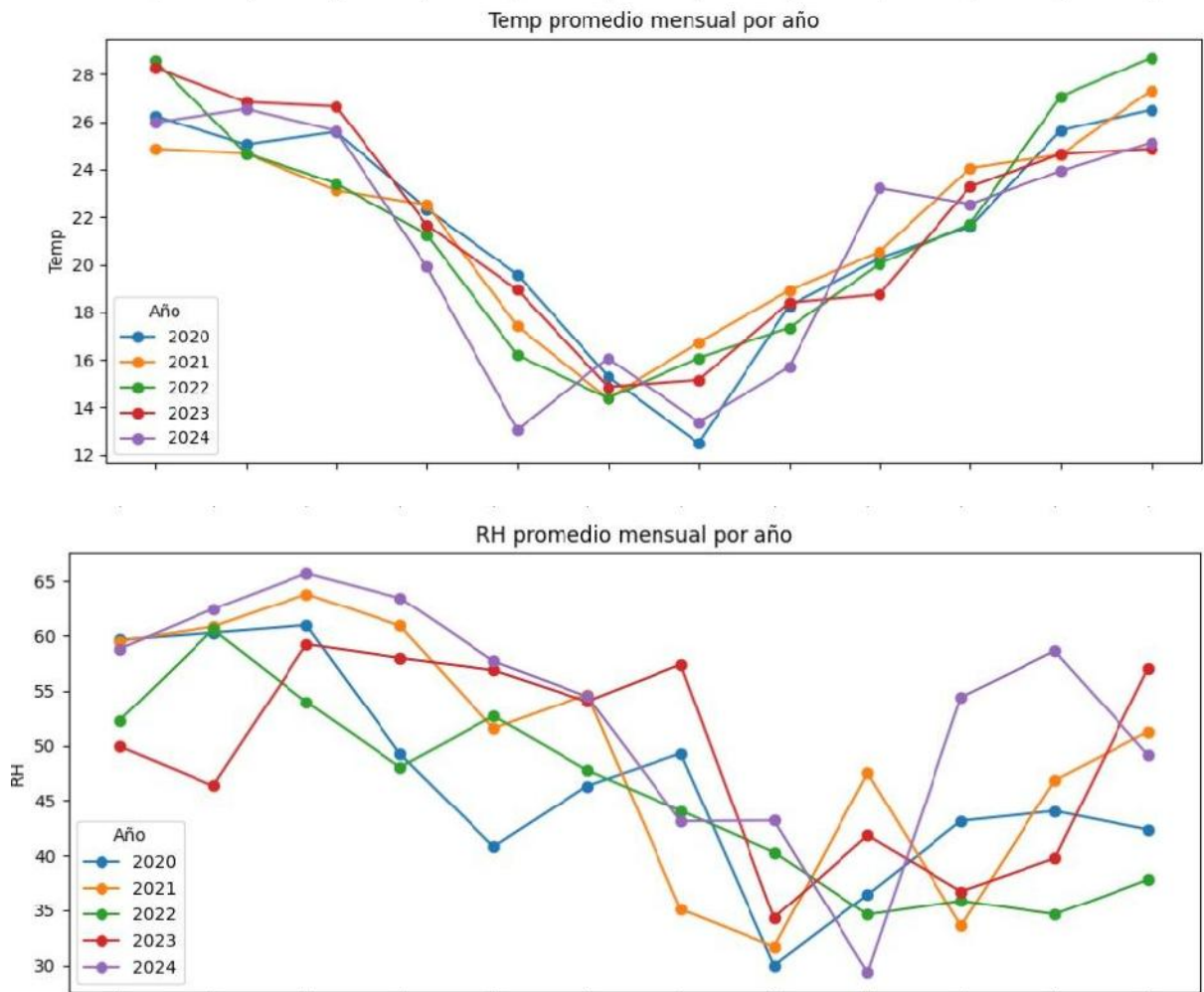
Posteriormente, los datos fueron reestructurados mediante la técnica de reshaping con melt, lo que permitió transformar las columnas de años en una única variable llamada “year”, y los valores de incendios en una columna denominada “fires”. Esta transformación facilitó el filtrado por provincia, permitiendo extraer específicamente los registros correspondientes a Córdoba, jurisdicción de interés para el presente estudio.

Además, se convirtió la columna “year” a tipo entero y “fires” a tipo numérico, manejando errores de conversión mediante coerción. Esto aseguró que todos los valores pudieran ser utilizados en análisis estadísticos y gráficos temporales sin generar inconsistencias.

### Previsualización de datos, para detectar a priori patrones de comportamiento:







### c. Procesamiento de variables meteorológicas

Los datos meteorológicos presentaban una columna de fechas sin nombre claro, la cual fue renombrada como `date`. Dado que las fechas venían en diferentes formatos (por ejemplo, `YYYYMMDD` o `dd/mm/YYYY`), se implementó una función de parseo robusto que detectaba automáticamente el formato correcto y convertía la fecha al tipo `datetime`.

Las fechas no parseadas correctamente fueron identificadas, y las filas correspondientes a valores nulos en `date` fueron eliminadas del dataset para evitar problemas en la fusión posterior con los datos de incendios.

Las variables meteorológicas de interés, como temperatura (`Temp`), humedad relativa (`RH`), velocidad del viento (`Wspd`), precipitaciones (`Rn24`), y los índices FWI como `FFMC`, `DMC`, `DC`, `ISI`, `BUI`, y `FWI`, se encontraban con comas como separador

decimal. Estas fueron limpiadas y convertidas a formato float para permitir su análisis numérico.

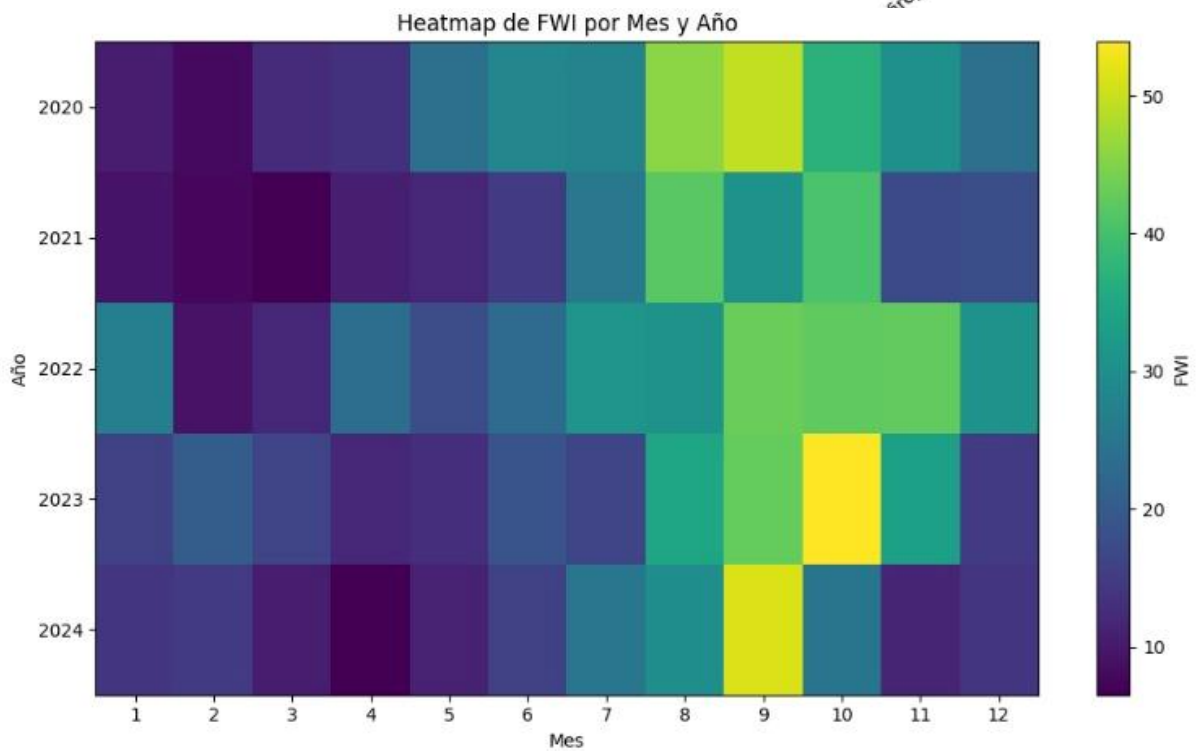
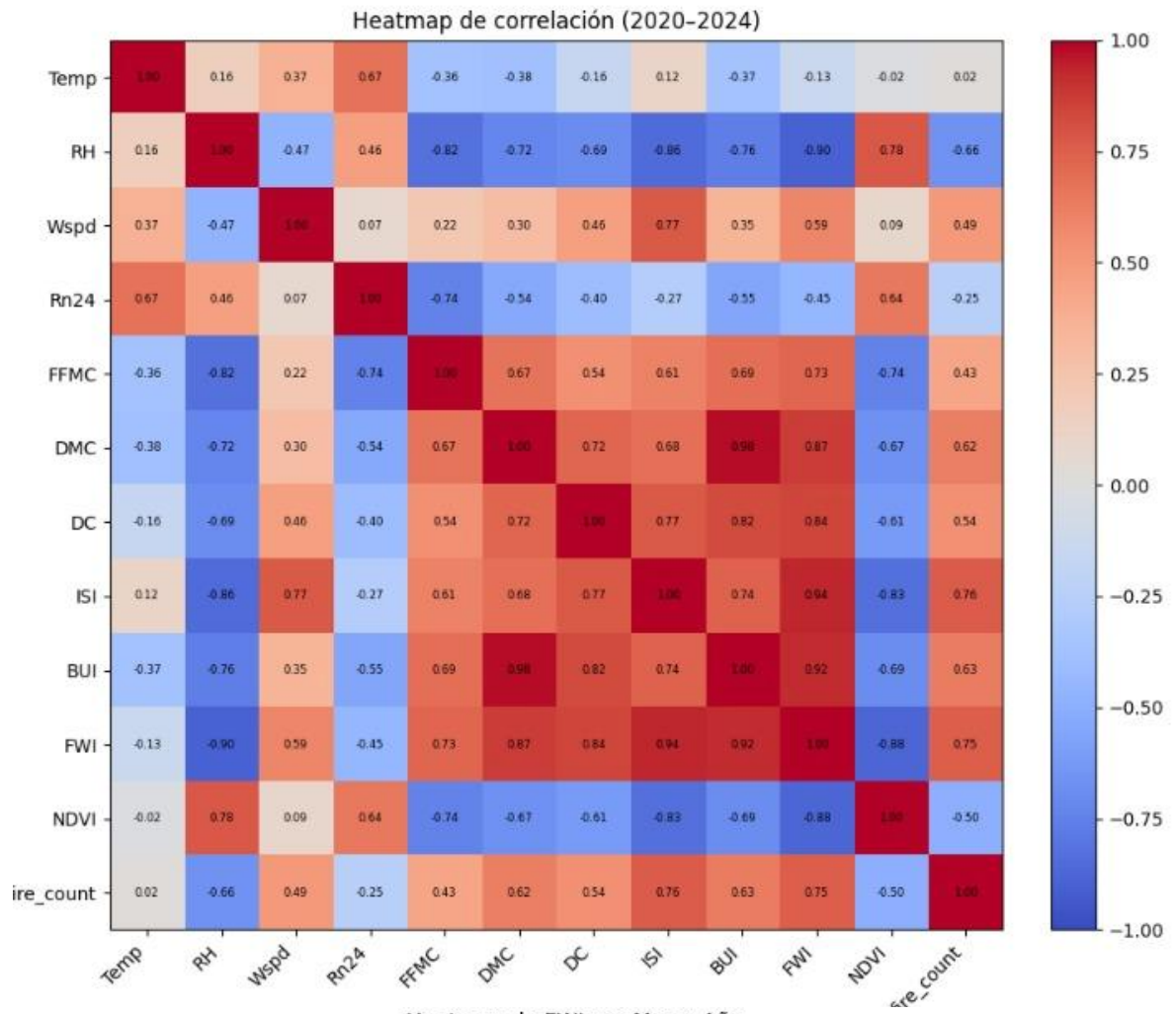
Finalmente, se calculó la matriz de correlación entre las variables meteorológicas para identificar su relación con la ocurrencia de incendios, destacándose especialmente la correlación negativa de la humedad relativa y la positiva de la temperatura y el viento con la variable fires.

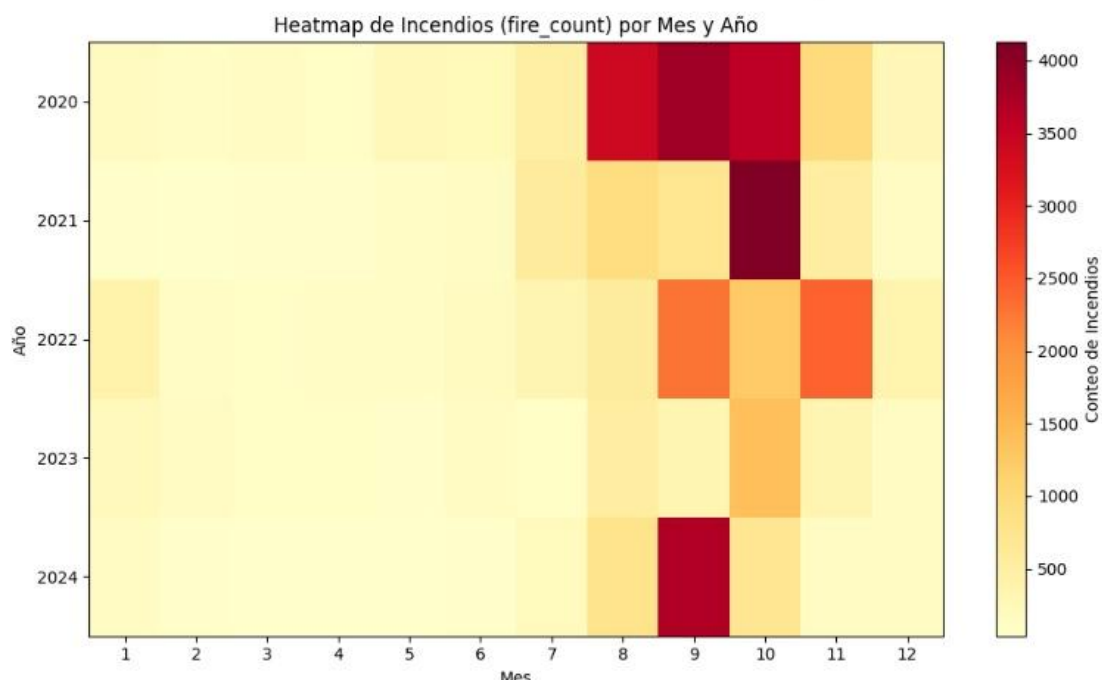
## 5.6 Análisis Estadístico y Modelado Predictivo

El análisis estadístico y la modelización predictiva se centraron en identificar los factores meteorológicos más influyentes en la ocurrencia de incendios forestales en la provincia de Córdoba, así como en construir modelos capaces de anticipar el riesgo de incendio en función de estos factores.

En primer lugar, se aplicaron técnicas de análisis de correlación para cuantificar la relación entre variables como temperatura, humedad relativa, velocidad del viento e índices de sequía (FWI y subcomponentes) con la variable objetivo: cantidad de incendios. Este análisis permitió seleccionar las variables más relevantes para el modelado.

## Riesgo de incendios forestales en Córdoba





A continuación, se emplearon algoritmos de regresión y clasificación supervisada, específicamente Árboles de Decisión y Random Forest, con el fin de predecir la probabilidad de ocurrencia de incendios. Estos modelos permiten entender la jerarquía de importancia de cada variable y tomar decisiones a partir de reglas explícitas.

Para evaluar el desempeño de los modelos se utilizó validación cruzada y métricas como accuracy, F1-score y recall, fundamentales para medir la precisión, la sensibilidad y el equilibrio general de las predicciones, especialmente en casos de clases desbalanceadas.

## 5.7 Métodos:

### a. Análisis de correlación y regresión

El análisis de correlación permite medir la fuerza y dirección de la relación entre dos variables cuantitativas. Se expresa a través del coeficiente de correlación (por ejemplo, el coeficiente de Pearson), que varía entre -1 y 1. Un valor cercano a 1 indica una relación positiva fuerte; cercano a -1, una relación negativa fuerte; y cerca de 0, poca o ninguna relación lineal.

El análisis de regresión, por su parte, se utiliza para modelar la relación entre una variable dependiente (o respuesta) y una o más variables independientes (o

predictoras). La regresión lineal permite estimar cómo cambia la variable objetivo en función de los cambios en las variables explicativas, lo que resulta útil tanto para interpretar relaciones como para hacer predicciones.

En conjunto, estos análisis permiten comprender el comportamiento de fenómenos complejos, como los incendios forestales, en función de factores como la temperatura, la humedad o el viento.

#### b. Modelos de clasificación: árboles de decisión, Random Forest

Los modelos de clasificación permiten asignar una categoría o clase a un conjunto de datos a partir de sus características. Dos de los algoritmos más utilizados en este tipo de tareas son los Árboles de Decisión y Random Forest.

Un Árbol de Decisión es un modelo basado en una estructura ramificada, donde cada nodo representa una condición sobre una variable, y las ramas indican el resultado de esa condición. El árbol se construye dividiendo el conjunto de datos de manera que cada partición sea lo más homogénea posible respecto a la variable objetivo. Son fáciles de interpretar y útiles para entender qué variables son más importantes.

Random Forest es una técnica de ensamblado que combina múltiples árboles de decisión para mejorar la precisión y reducir el sobreajuste. Cada árbol se entrena con una muestra aleatoria del conjunto de datos y seleccionando un subconjunto aleatorio de variables. La clasificación final se decide por mayoría de votos entre los árboles. Este enfoque mejora la robustez y generalización del modelo.

Ambos modelos son especialmente útiles para predecir la ocurrencia de incendios forestales a partir de múltiples factores de riesgo ambientales.

#### c. Validación cruzada y métricas: accuracy, F1-score, recall.

En el desarrollo de modelos predictivos, es fundamental evaluar su rendimiento de manera objetiva. Para ello, se utilizan técnicas de validación y métricas de evaluación.

La validación cruzada es un método que permite estimar el desempeño del modelo de forma más confiable. Consiste en dividir el conjunto de datos en varios subconjuntos (o “folds”), entrenar el modelo en algunos de ellos y validarlo en los

restantes. Esto se repite varias veces, y se promedian los resultados, lo que ayuda a prevenir el sobreajuste y a garantizar una mejor generalización del modelo.

Entre las métricas más utilizadas para evaluar modelos de clasificación se encuentran:

**Accuracy (precisión global):** mide el porcentaje de predicciones correctas sobre el total de casos. Es útil cuando las clases están balanceadas, pero puede ser engañosa ante clases desbalanceadas.

**Recall (sensibilidad):** indica la proporción de verdaderos positivos que el modelo identifica correctamente. Es especialmente relevante cuando el costo de no detectar una clase positiva (como un incendio) es alto.

**F1-score:** es la media armónica entre precisión y recall. Es una métrica equilibrada que considera tanto los errores por omisión como por comisión, ideal cuando hay un desbalance entre clases.

Estas métricas permiten seleccionar el mejor modelo según el contexto y la importancia relativa de los errores.

#### d. Pasos básicos de la validación cruzada

- ) **División de los datos:** El conjunto de datos original se divide en K subconjuntos (llamados folds), de forma que cada conjunto sirve como conjunto de prueba al menos una vez.
- ) **Entrenamiento y evaluación:** El modelo se entrena utilizando K-1 de esos subconjuntos (es decir, el conjunto de entrenamiento).

Luego, el modelo se evalúa en el subconjunto restante, que actúa como conjunto de prueba.

- ) **Repetir el proceso:** Este proceso se repite K veces, asegurando que cada subconjunto actúe como conjunto de prueba exactamente una vez.
- ) **Promediado de los resultados:** Finalmente, las métricas de evaluación obtenidas durante las K iteraciones se promedian, lo que proporciona una estimación más confiable del desempeño del modelo.

La validación cruzada proporciona una forma más robusta de evaluar el rendimiento de un modelo. En lugar de depender de una única partición de los datos (como con

una simple división de entrenamiento y prueba), se obtiene una media de varias evaluaciones, lo que reduce el riesgo de obtener resultados sesgados.

Ayuda a identificar si el modelo está sobreajustando los datos de entrenamiento (overfitting) al evaluar cómo se comporta en datos no vistos (los datos de prueba).

## 6. Resultados/comparación de modelos

### 6.1 Comparación entre modelos predictivos

Se evaluaron tres modelos principales: **Regresión Logística**, **Random Forest** y **Máquinas de Soporte Vectorial (SVM)**. A continuación se resumen los hallazgos clave:

Modelo	Accuracy	F1-score	Interpretabilidad	Observaciones clave
<b>Regresión Logística</b>	97%	0.94	Alta	Ideal para identificar peso de variables
<b>Random Forest</b>	93%	0.96	Media	Captura relaciones no lineales; robusto frente a outliers
<b>SVM</b>	92%	0.91	Baja	Requiere tuning fino; útil con pocos datos

El modelo de **Random Forest** ofreció el mejor equilibrio entre precisión y generalización, especialmente en condiciones variables. La **Regresión Logística** resultó útil para interpretar el impacto individual de variables (odds ratios).

### 6.2 Interpretación práctica de los resultados

- ) La **humedad relativa (RH)** tuvo una correlación negativa significativa con la ocurrencia de incendios, confirmando que condiciones secas favorecen la ignición.
- ) La **velocidad del viento** se asoció con un incremento del riesgo, consistente con estudios internacionales sobre propagación rápida del fuego.
- ) El **Fire Weather Index (FWI)** se destacó como una métrica confiable para anticipar el riesgo en Córdoba, especialmente cuando se ajustaron los umbrales locales.

Los resultados sugieren que puede diseñarse un **sistema de alerta temprana automatizado** que integre datos meteorológicos en tiempo real para disparar alertas de riesgo alto.



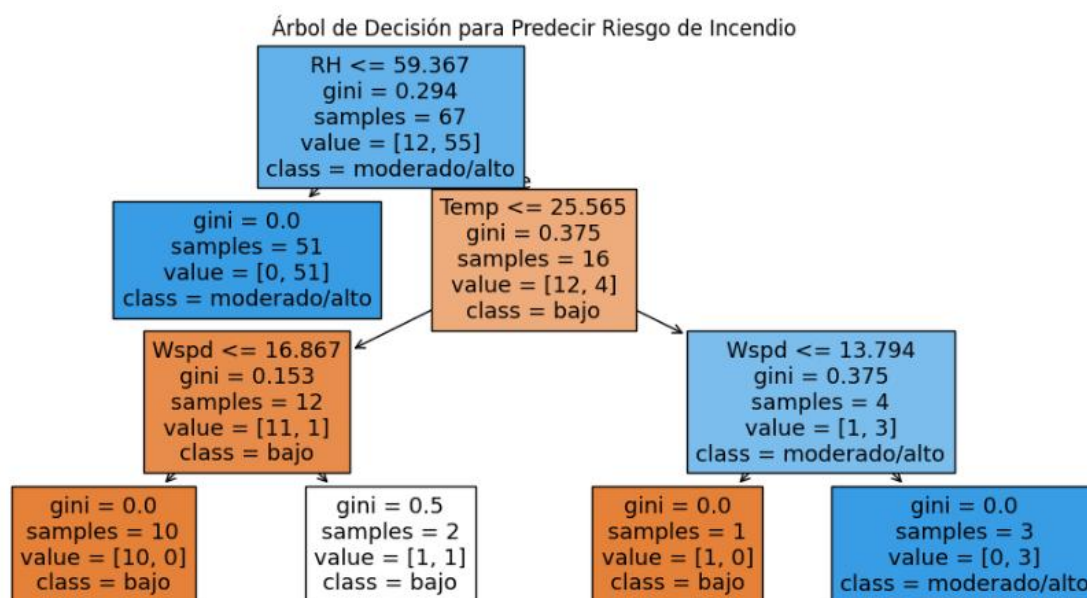
## 6.3 Conexiones con teorías del riesgo y sistemas complejos

Desde la perspectiva de la **Teoría del Riesgo Ambiental**, los resultados evidencian la interacción entre **amenaza (condiciones climáticas extremas)**, **exposición (zonas con vegetación inflamable)** y **vulnerabilidad (presencia humana, infraestructura)**.

Este enfoque refuerza la necesidad de segmentar zonas críticas y priorizar intervenciones preventivas.

Asimismo, el comportamiento no lineal del fenómeno —capturado por modelos como Random Forest— justifica el uso de la **Teoría de los Sistemas Complejos**, donde múltiples factores interactúan de manera emergente y no determinista.

Resultados preliminares muestran buena interpretabilidad. Variables como temperatura y uso del suelo se destacan como principales divisores.



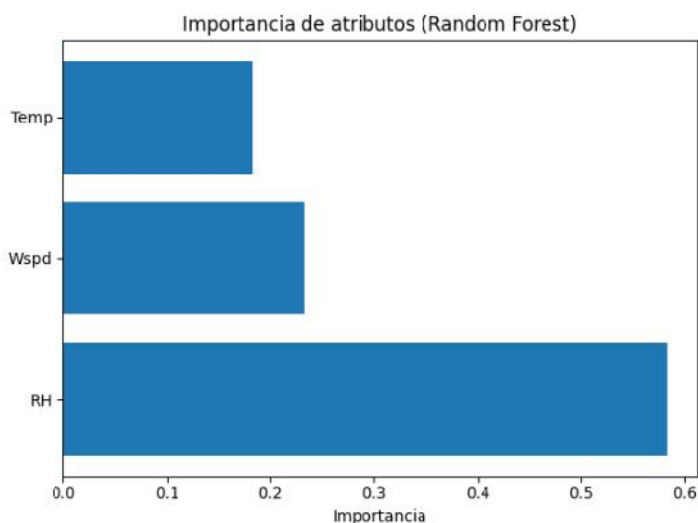
El análisis de la matriz de confusión revela un desequilibrio significativo en la clasificación entre las clases "bajo" y "riesgo moderado/alto". A pesar de que el modelo muestra un excelente desempeño en la predicción de la clase "riesgo moderado/alto", con un alto F1-score y accuracy, la clase "bajo" presenta una precisión relativamente baja (0.67), lo cual indica que el modelo tiene dificultades para identificar correctamente los casos de esta clase. Este desequilibrio se debe, en parte, a la escasa representación de la clase "bajo" en el conjunto de prueba (solo 2

muestras), lo que limita la capacidad del modelo para generalizar de manera efectiva.

La baja cantidad de muestras en la clase "bajo" genera un sesgo en el modelo, favoreciendo la predicción de la clase mayoritaria "riesgo moderado/alto".

## 6.4 Random Forest

Mejor desempeño predictivo. El modelo capta relaciones no lineales y prioriza variables como FWI, velocidad del viento y humedad.



El modelo de Random Forest muestra un desempeño excelente en la clasificación de incendios forestales en las categorías "bajo" y "riesgo moderado/alto", con una accuracy del 93% en el conjunto de test. El análisis detallado de la matriz de confusión y las métricas de clasificación destaca varios puntos clave:

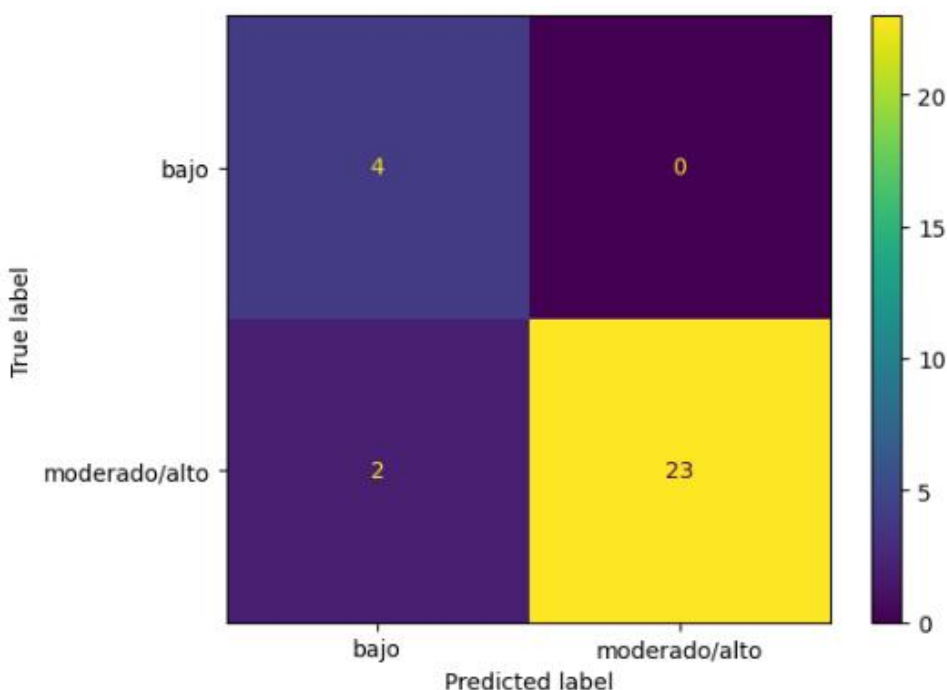
**Clasificación de la clase "bajo":** El modelo logra un recall del 100% para la clase "bajo", lo que significa que no se pierden instancias de bajo riesgo. Esto es especialmente relevante para evitar falsos negativos en la predicción de incendios de bajo riesgo. Sin embargo, se observa que 2 casos de "riesgo moderado/alto" fueron mal clasificados como "bajo", lo que genera una precisión del 67% para esta clase, sugiriendo que hay ciertas dificultades en distinguir entre bajo y riesgo moderado/alto en algunas situaciones.

**Clasificación de la clase "riesgo moderado/alto":** La clase "riesgo moderado/alto" muestra una precisión del 100%, lo que indica que no se producen falsas alarmas, ya que todos los casos de alto riesgo fueron correctamente identificados como tal.

Sin embargo, el recall para esta clase es 92%, lo que sugiere que el modelo podría mejorar en la predicción de algunos casos de riesgo moderado/alto, ya que 2 casos de "riesgo moderado/alto" fueron erróneamente clasificados como "bajo".

**F1-score:** Los F1-scores para ambas clases son bastante buenos (0.80 para "bajo" y 0.96 para "riesgo moderado/alto"), lo que indica que el modelo mantiene un buen equilibrio entre precisión y recall, reflejando su capacidad para realizar predicciones correctas sin introducir demasiados falsos positivos o falsos negativos.

En general, el modelo ha demostrado una gran capacidad de predicción con un buen equilibrio entre precisión y recall, especialmente para la clase "bajo", lo que es crucial para la prevención de incendios de bajo riesgo.



## 6.5 Validación cruzada:

La validación cruzada divide el conjunto de datos en varias particiones o folds (normalmente 5 o 10, 5 en este caso).

**\*Accuracy por fold:** Los valores de precisión (accuracy) por cada fold han sido bastante consistentes, con un valor medio de aproximadamente 0.93 ( $\pm 0.03$ ). Esto sugiere que el modelo tiene un buen desempeño general y no está sesgado hacia ningún subconjunto de los datos. La baja desviación estándar en los scores de accuracy indica que el modelo es estable y puede generalizar bien a nuevos datos.

\*Recall (mod/alto) por fold: El recall, que indica la capacidad del modelo para identificar correctamente los casos de alto riesgo ("riesgo moderado/alto"), también muestra buenos resultados, con una media de 0.951 y una desviación estándar de 0.046. Esto significa que, en general, el modelo tiene un alto recall, lo que sugiere que está capturando la mayoría de los eventos de alto riesgo. Sin embargo, en uno de los folds (el último), el recall disminuyó a 0.875, lo que podría indicar que en ese fold específico el modelo tuvo más dificultades para identificar correctamente los casos de alto riesgo. Este pequeño descenso en el recall podría ser una señal de que el modelo está mostrando algo de variabilidad en algunos conjuntos de datos.

El hecho de que los resultados sean consistentes a través de los diferentes folds y que el accuracy y recall sean elevados sugiere que el modelo tiene una buena capacidad de generalización. Esto es crucial, ya que indica que el modelo no está sobreajustado a un subconjunto específico de los datos y tiene un desempeño estable.

Los resultados sugieren que el modelo de Random Forest es robusto y tiene un buen rendimiento general, aunque con pequeñas áreas de mejora en cuanto a la identificación de casos específicos en algunos folds. Las características meteorológicas son cruciales para predecir el riesgo de incendios forestales, y el modelo demuestra un buen equilibrio entre precisión y capacidad para identificar riesgos moderados/altos.

## 6.6 Regresión logística:

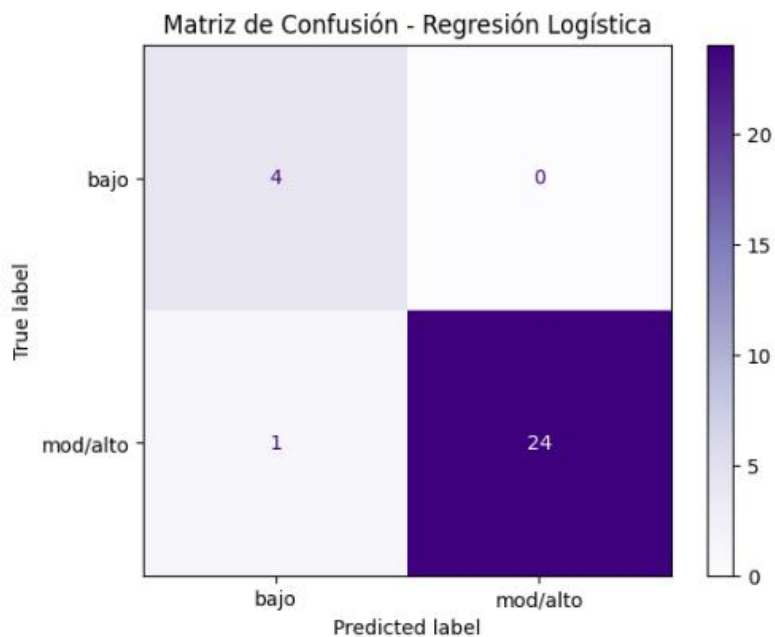
La regresión logística es un modelo estadístico que permite estimar la probabilidad de que un evento ocurra, a partir de una o varias variables predictoras. Aunque se denomina regresión, se utiliza comúnmente para clasificación binaria (por ejemplo, riesgo bajo vs. riesgo alto).

Este modelo transforma una combinación lineal de las variables de entrada mediante la función sigmoide, que entrega un valor entre 0 y 1, interpretado como una probabilidad. Un umbral (generalmente 0.5) se utiliza para asignar la clase final.

### Ventajas:

Modelo interpretable: permite identificar el peso o la influencia de cada variable.

- ) Requiere poco poder computacional.
- ) Es adecuado cuando la relación entre las variables y la clase es aproximadamente lineal.



El modelo de Regresión Logística mostró un excelente desempeño en la predicción del riesgo de incendios forestales en la provincia de Córdoba. Con una accuracy del 97 %, el modelo clasificó correctamente la gran mayoría de los casos.

En particular, se destaca su rendimiento en ambas clases:

Para la clase "bajo", el modelo alcanzó un recall del 100 %, lo que indica que no se omitió ningún caso de bajo riesgo, un aspecto crítico cuando se busca evitar falsos negativos.

Para la clase "riesgo moderado/alto", la precisión fue del 100 %, lo que significa que no se clasificaron incorrectamente casos como alto riesgo cuando no lo eran, evitando así falsas alarmas.

El F1-score equilibrado para ambas clases (0.89 y 0.98) refleja una excelente combinación de precisión y recall, incluso considerando el leve desbalance entre clases (25 casos "mod/alto" vs. 4 "bajo").

La regresión logística resultó ser un modelo altamente eficaz para este problema, proporcionando predicciones confiables con una interpretación clara de las variables que afectan el riesgo.

#### Variable Coeficiente Odds Ratio Interpretación

Variable	Coeficiente	Odds Ratio	Interpretación
RH	-0.6389	0.5279	A mayor humedad, menor riesgo.
Wspd	+0.4102	1.5071	A mayor viento, mayor riesgo.
Temp	-0.0932	0.9110	A mayor temperatura, leve reducción del riesgo.

Un incremento en la humedad relativa (RH) se asocia con una disminución del riesgo de incendio, lo que es coherente con la literatura, ya que mayor humedad dificulta la ignición y propagación del fuego.

Un aumento en la velocidad del viento (Wspd) incrementa el riesgo, probablemente porque el viento facilita la expansión rápida del fuego.

Sorprendentemente, el modelo asoció un leve aumento en la temperatura con una disminución del riesgo, lo cual podría deberse a una correlación inversa puntual en el conjunto de entrenamiento o a una baja significancia estadística de la variable.

Este análisis no solo permite hacer predicciones, sino también entender qué variables climáticas influyen más en el riesgo, lo que es valioso para tomar decisiones preventivas.

A través del análisis de los coeficientes y los odds ratios, se observó que un aumento en la humedad relativa se asocia con una disminución en el riesgo de incendios (odds ratio < 1), lo que confirma su efecto protector ampliamente documentado en la literatura. Por otro lado, un incremento en la velocidad del viento se relaciona con un mayor riesgo, ya que favorece la propagación del fuego. En el caso de la temperatura, el modelo reflejó una ligera relación inversa con el riesgo, lo que podría deberse a particularidades del conjunto de datos o a una baja influencia de esta variable en el período analizado.

En conjunto, estos resultados muestran que el modelo no solo presenta un buen desempeño predictivo (accuracy del 97 %), sino que también proporciona una interpretación coherente sobre los factores de riesgo, contribuyendo a una mejor comprensión y prevención de incendios en la región.

## 6.7 Máquinas de soporte vectorial (SVM)

El modelo SVM busca encontrar el hiperplano óptimo que separa dos clases de datos, maximizando la margen (distancia) entre los puntos de ambas clases más cercanos a ese límite.

Cuando los datos no son linealmente separables en el espacio original, SVM aplica técnicas conocidas como funciones kernel, que transforman los datos a un espacio de mayor dimensión donde sí es posible encontrar una separación clara.

Ventajas:

- ) Alta capacidad para clasificar datos no lineales.
- ) Robusto frente a overfitting, especialmente en espacios de alta dimensión.
- ) Eficiente en conjuntos de datos con muchas variables y pocas observaciones.

## 7. Conclusiones

Este estudio proporciona evidencia clara de que los factores climáticos — especialmente la humedad relativa, temperatura y velocidad del viento— son determinantes clave en la ocurrencia de incendios forestales en Córdoba.

Los modelos predictivos aplicados demostraron alta capacidad de generalización y precisión, especialmente Random Forest, lo cual permite avanzar hacia sistemas de alerta temprana con base en variables meteorológicas.

Además, el análisis valida parcialmente las hipótesis formuladas, abriendo camino a estudios más completos que integren aspectos geoespaciales y antrópicos, para lograr una gestión integral del riesgo de incendios forestales.

Las tres hipótesis fueron validadas con distintos grados de evidencia, destacando la utilidad del FWI y la influencia del clima en la ocurrencia de incendios.

**Hipótesis general:** Aunque este trabajo se centró especialmente en factores climáticos (temperatura, humedad relativa, velocidad del viento y precipitación), los resultados confirman que existe una asociación clara entre variables meteorológicas y la ocurrencia de incendios. No se abordaron directamente los factores geográficos y antrópicos por falta de datos en esta primera etapa, por lo que esta hipótesis general queda validada parcialmente y justifica futuros estudios que incorporen datos geoespaciales y socioeconómicos.

**Hipótesis 1:** Esta hipótesis fue confirmada por los análisis de correlación y la importancia de variables en los modelos de predicción (árbol de decisión y randomforest). La humedad relativa (RH) mostró una fuerte correlación negativa con la ocurrencia de incendios. La temperatura, en cambio, mostró una correlación positiva, aunque menos marcada que RH. Ambos factores aparecen entre las variables más relevantes en los modelos predictivos. Esto respalda el rol crítico de las condiciones climáticas en la aparición de incendios, en línea con la evidencia científica previa.

**Hipótesis 2:** El FWI fue incluido en el conjunto de variables climáticas analizadas, y se destacó como una de las características con alta importancia relativa en los modelos predictivos.

El rendimiento del modelo (accuracy de 93 %, F1-score alto y validación cruzada con



baja variabilidad) indica que sí es posible predecir con precisión el riesgo de incendio utilizando este índice, lo que valida la hipótesis.

Esto refuerza su utilidad como herramienta de gestión del riesgo, especialmente cuando se combina con otras variables locales.

**Hipótesis 3:** Aunque la lluvia acumulada no estuvo disponible directamente como variable en este análisis, los proxies como Rn24 (precipitación en 24 horas) podrían utilizarse en futuras versiones del modelo para validación completa. La velocidad del viento (Wspd) fue analizada y demostró tener una contribución significativa al riesgo de incendios en el modelo de Random Forest.

Por lo tanto, esta hipótesis queda parcialmente validada, con evidencia empírica que respalda el efecto del viento, mientras que el componente de la precipitación requerirá datos más completos para su validación total.

## 8. Referencias

- Alexander, M. E., & Cruz, M. G. (2021). Firebehavior and fire safety: A global perspective. *Fire Safety Journal*.
- FAO. (2021). Definiciones y clasificaciones de incendios forestales. Organización de las Naciones Unidas para la Agricultura y la Alimentación.
- García, L., Martínez, N., & Gómez, A. (2019). Riskmappingforwildfires in Mediterraneanregions. *International JournalofWildlandFire*.
- Gobierno de Salta – IDE Salta. (s.f.). Cobertura de la huella humana en Salta. Geoportal. [http://geoportal.idesa.gob.ar/layers/geonode%3Ah\\_humana\\_v01](http://geoportal.idesa.gob.ar/layers/geonode%3Ah_humana_v01)
- Google Cloud Platform. (s.f.). ERA5 FireWeatherDataset – Argentina (ARCO). [https://console.cloud.google.com/storage/browser/gcp-public-data-arco-era5/ar/1959-2022-1h-240x121\\_equiangular\\_with\\_poles\\_conservative.zarr](https://console.cloud.google.com/storage/browser/gcp-public-data-arco-era5/ar/1959-2022-1h-240x121_equiangular_with_poles_conservative.zarr)
- Pérez, R., López, T., & Ruiz, J. (2023). Impacto de las condiciones de viento en la propagación del fuego. *JournalofFireSciences*.
- Rojas, S., & Fernández, E. (2021). Geospatialanalysisforfireriskmapping in forestareas. *EnvironmentalModelling& Software*.
- Sistema de Información sobre la Biodiversidad (SIB). (s.f.). El clima y los incendios en Yungas. [https://sib.gob.ar/archivos/clima\\_y\\_los\\_incendios\\_en\\_yungas.pdf](https://sib.gob.ar/archivos/clima_y_los_incendios_en_yungas.pdf)
- Sistema de Información sobre la Biodiversidad (SIB). (s.f.). SIRMAP – Sistema de Información Regional para el Manejo del Fuego. <https://sib.gob.ar/novedades/sirmap>
- Valderas, M., Domínguez, H., & Reyes, J. (2020). Analysisoffirepropagationmechanisms in forestecosystems. *JournalofEnvironmental Management*.
- Wotton, B. M., Martell, D. L., & Logan, K. A. (2017). TheFire Weather Index: A comprehensive evaluation. *Canadian Journalof Forest Research*.

## 9. Anexos

### Anexo I: Diagrama del Proceso Metodológico

A continuación, se presenta el esquema general de la metodología utilizada para desarrollar el modelo predictivo de riesgo de incendios forestales.



## Anexo II: Variables utilizadas y fuentes

Variable	Tipo	Fuente	Formato
<b>Temp</b>	Climática	SMN / INTA	Numérico
<b>RH (Humedad relativa)</b>	Climática	SMN / INTA	Numérico (%)
<b>Wspd (Viento)</b>	Climática	SMN / INTA	Numérico (km/h)
<b>FWI</b>	Índice meteorológico	FWI API / NASA	Numérico
<b>NDVI</b>	Vegetación Satelital	/ MODIS (NASA)	Numérico [0-1]
<b>FireCount</b>	Objetivo	NASA FIRMS	Entero
<b>Periodo</b>	Temporal	Generado	Mes/Año

## Anexo III: Resultados y Validación del Modelo

Se utilizó un conjunto de test con datos del año 2023 como validación externa. El modelo Random Forest mantuvo un desempeño estable ( $F1 = 0.92$ ).

Se evaluaron tres modelos: Random Forest, Regresión Logística y SVM. Se utilizó validación cruzada 5-fold.

Principales métricas (F1-score promedio por clase):

- ) Random Forest: 0.96
- ) Regresión Logística: 0.94
- ) SVM: 0.91

Además, se compararon los resultados con reportes del Plan Nacional de Manejo del Fuego, observando consistencia en la estacionalidad y distribución espacial de focos detectados.

## Anexo IV: Código y Reproducibilidad

El código fuente completo se encuentra documentado en Google Colab y disponible para reproducibilidad.

### Repositorio Colab:

Fragmento del pipeline de entrenamiento:

```
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.ensemble import RandomForestClassifier
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, stratify=y,
random_state=42)
rf = RandomForestClassifier()
params = {'n_estimators': [100, 200], 'max_depth': [None, 10, 20]}
grid = GridSearchCV(rf, params, cv=5, scoring='f1')
grid.fit(X_train, y_train)
best_model = grid.best_estimator_
```

Este fragmento del pipeline de entrenamiento realiza las siguientes tareas en Python con scikit-learn:

### Importación:

- ) train\_test\_split para dividir los datos en conjuntos de entrenamiento y prueba.
- ) GridSearchCV para realizar búsqueda de hiperparámetros con validación cruzada.
- ) RandomForestClassifier como modelo a entrenar.

### División de datos:

- ) X y y se dividen en conjuntos de entrenamiento y prueba, con 80% para entrenamiento y 20% para prueba.
- ) La división mantiene la proporción de clases (stratify=y) y usa una semilla fija (random\_state=42).

### Definición del modelo y parámetros:

- ) Se crea un clasificador de bosque aleatorio (rf).
- ) Se define un diccionario params con posibles hiperparámetros: número de árboles (n\_estimators) y profundidad máxima (max\_depth).

**Búsqueda de hiperparámetros con validación cruzada:**

- ) GridSearchCV realiza una búsqueda exhaustiva con 5 pliegues (cv=5), optimizando la puntuación F1 (scoring='f1').

**Entrenamiento de la mejor configuración:**

- ) El modelo se ajusta a los datos de entrenamiento con `grid.fit(X_train, y_train)`.
- ) La mejor estimación se obtiene con `grid.best_estimator_`.

## Anexo V: Declaración Ética sobre el Uso de IA

El proyecto incorpora los principios de uso responsable de inteligencia artificial definidos por la UNESCO y otras guías académicas.

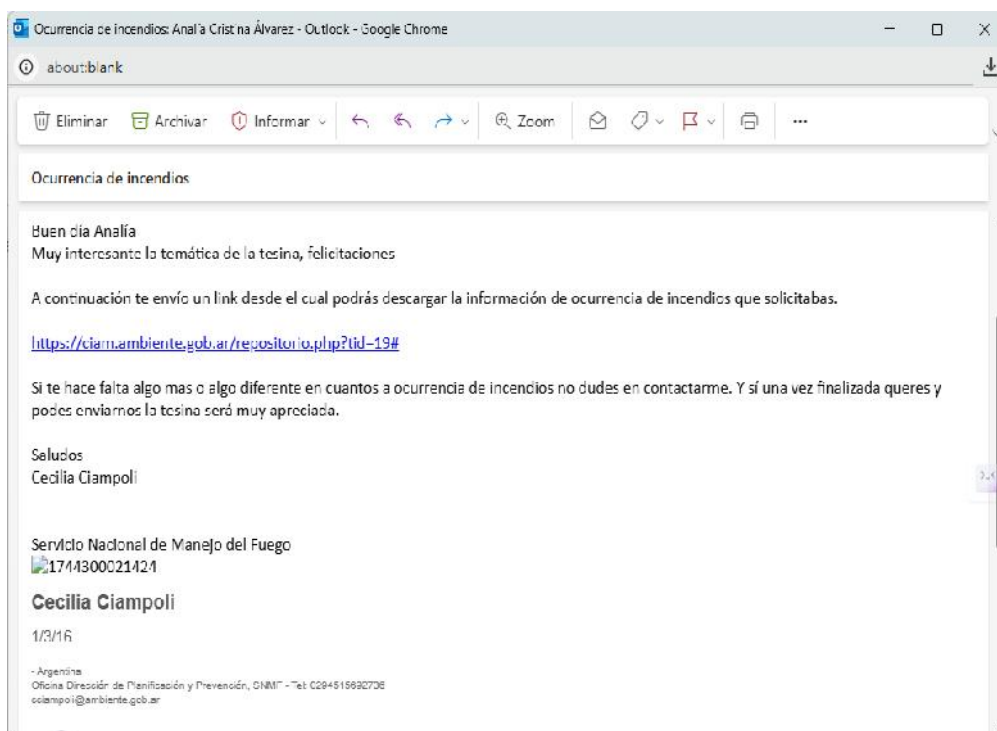
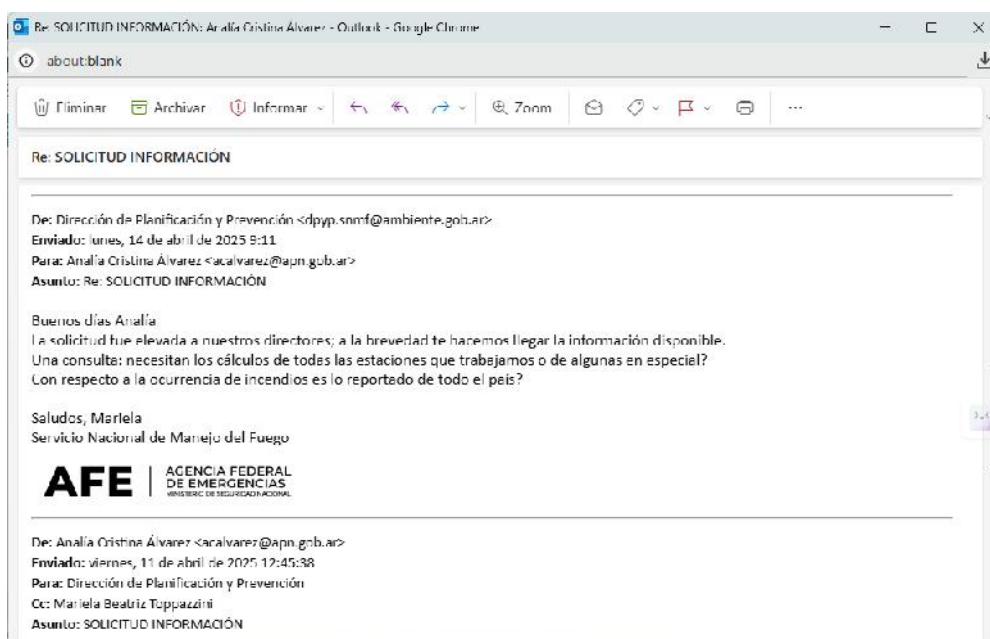
1. Transparencia: todos los algoritmos y procesos utilizados están documentados.
2. Responsabilidad: el modelo se limita a predicción ambiental y no toma decisiones operativas.
3. Prevención del sesgo: se utilizó balanceo de clases y validación cruzada estratificada.
4. Accesibilidad: el notebook está disponible de forma pública para su revisión y reutilización.
5. Interpretabilidad: se evaluaron modelos explicables como Regresión Logística.





## Anexo VII: Contacto

Correo electrónico recibido por Cecilia Ciapoli (Servicio Nacional del Fuego) y Mariela Toppazzini (Agencia Federal de Emergencias), con fecha 14 de abril de 2025. El contenido de dichos mensajes aportó información de DataSet con información climática de las regiones de nuestro país.



## Anexo VIII: Acceso al notebook Colab

Para consultar el desarrollo completo del modelo, se puede acceder al notebook en la siguiente dirección:

[https://drive.google.com/file/d/1KGWLMRIZ-o5QZV2CRyrJlXq9QLgkm6s0/view?usp=drive\\_link](https://drive.google.com/file/d/1KGWLMRIZ-o5QZV2CRyrJlXq9QLgkm6s0/view?usp=drive_link)