DISEÑO DE UN SISTEMA DE SOFTWARE PARA EL MONITOREO METEOROLÓGICO Y LA GENERACIÓN DE ALERTAS TEMPRANAS DE DESASTRES NATURALES EN EL ÁREA METROPOLITANA DEL VALLE DE ABURRÁ

AUTORES

JORGE EDUARDO ARANGO GAVIRIA,
JUAN PABLO ARANGO GAVIRIA

INSTITUCIÓN UNIVERSITARIA PASCUAL BRAVO

FACULTAD DE INGENIERÍA

INGENIERÍA DE SOFTWARE
TECNOLOGÍA EN DESARROLLO DE SOFTWARE

MEDELLÍN
2026

DISEÑO DE UN SISTEMA DE SOFTWARE PARA EL MONITOREO METEOROLÓGICO Y LA GENERACIÓN DE ALERTAS TEMPRANAS DE DESASTRES NATURALES EN EL ÁREA METROPOLITANA DEL VALLE DE ABURRÁ

AUTORES

JORGE EDUARDO ARANGO GAVIRIA,
JUAN PABLO ARANGO GAVIRIA

ASESOR

JAVIER ALBERTO SALDARRIAGA CANO
MAGISTER EN ENTORNOS VIRTUALES DE EDUCACIÓN

TRABAJO DE GRADO PARA OPTAR AL TÍTULO DE TECNÓLOGO EN DESARROLLO DE SOFTWARE

INSTITUCIÓN UNIVERSITARIA PASCUAL BRAVO

FACULTAD DE INGENIERÍA
INGENIERÍA DE SOFTWARE
TECNOLOGÍA EN DESARROLLO DE SOFTWARE

MEDELLÍN
2026

Índice

1 Resumen 8
2 Abstract 9
3 Glosario 10
4 Introducción 13
5 Planteamiento del problema 15
5.1 Descripción 15
5.2 Formulación 16
6 Justificación 17
7 Objetivos 19
7.1 Objetivo general 19
7.2 Objetivos específicos 19
8 Marco teórico 21
8.1 Teoría general de sistemas 21
8.2 Sistemas de alerta temprana y gestión del riesgo 21
8.3 Aprendizaje automático para predicción meteorológica 22
8.4 Open-Meteo como fuente de datos hidrometeorológicos 22
8.5 Visualización geoespacial para gestión del riesgo 23
8.6 Variables hidrometeorológicas relevantes para predicción de riesgos 23
8.7 Antecedentes 23
8.7.1 Antecedentes geográficos 23
8.7.2 Antecedente académico 25
8.7.3 Antecedente tecnológico 27
8.8 Marco legal y normativo 28
8.8.1 Ley 1523 de 2012 Política Nacional de Gestión del Riesgo de Desastres 28
8.8.2 Decreto 1974 de 2013 Conformación del SNGRD 29
8.8.3 Ley 1581 de 2012 Protección de datos personales 29
8.8.4 Ley 23 de 1982 y Decisión Andina 351 de 1993 29
8.8.5 ISO/IEC 25010:2011 Calidad del producto de software 30
8.8.6 Términos de uso de Open-Meteo 30
8.9 Tecnologías y herramientas utilizadas en el desarrollo del sistema 31
8.9.1 Python 3.x 31
8.9.2 XGBoost 31
8.9.3 Pandas y Numpy 32
8.9.4 Scikit-learn 32

8.9.5 Requests 33
8.9.6 Open-Meteo API 33
8.9.7 Folium 33
8.9.8 Seaborn y Matplotlib 34
8.9.9 Google Colab 34
8.9.10 SQLite 35
8.9.11 Plotly 35
9 Marco conceptual 37
9.1 Sistema de monitoreo meteorológico 37
9.2 Inteligencia artificial 37
9.3 Sistema de alerta temprana (SAT) 37
9.4 XGBoost 38
9.5 Redes de sensores ambientales y APIs meteorológicas 38
9.6 Análisis predictivo por sectores 38
10 Metodología 39
10.1 Tipo de proyecto 39
10.2 Metodología de desarrollo 39
10.3 Cronograma de actividades 40
11 Resultados del desarrollo del proyecto 40
11.1 Levantamiento de información 40
11.2 Análisis de requerimientos 41
11.2.1 Requisitos de usuario 41
11.2.2 Requerimientos funcionales 43
11.2.3 Requerimientos no funcionales 44
11.3 Diseño del sistema 44
11.3.1 Arquitectura general 44
11.3.2 Variables por módulo 45
11.3.3 Factores geográficos por sector 46
11.3.4 Diseño del dashboards de visualización de resultados 47
11.4 Desarrollo de la aplicación 51
11.4.1 Módulo de predicción de deslizamientos 51
11.4.2 Módulo de predicción de inundaciones 52
11.4.3 Módulo de predicción de incendios 52
11.4.4 Exportación y persistencia de datos 52
11.5 Pruebas y métricas de evaluación 57
11.6 Implementación y visualización de resultados 58
11.7 Análisis de resultados por sector 60

11.8 Resultados obtenidos por módulo 62
11.8.1 Resultados dashboard módulo deslizamientos 62
11.8.2 Resultados dashboard módulo incendios 63
11.8.3 Resultados dashboard módulo inundaciones 64
12 Conclusiones 65
13 Recomendaciones 67
14 Referencias bibliográficas 68

Índice de ilustraciones

Ilustración 1. Cronograma de actividades 40
Ilustración 2. Variables incendios 45
Ilustración 3. Dashborad resumen superior 48
Ilustración 4. Dashboard resumen central - Mapa general 49
Ilustración 5. Dashboard top 5 con mayor o menor riesgo 49
Ilustración 6. Dashboard grafico riesgo por sector 50
Ilustración 7. Dashboard gráficos de distribución 50
Ilustración 8. Dashboard gráfico con relación de variables 51
Ilustración 9 Tabla de datos open-meteo parte I 54
Ilustración 10 Tabla de datos open-meteo parte II 54
Ilustración 11 Tabla SQL riesgo por sector deslizamiento 55
Ilustración 12 Tabla SQL riesgo por sector inundaciones 56
Ilustración 13 Tabla SQL riesgo por sector incendio 56
Ilustración 14. Riesgo por sectores 58
Ilustración 15. Riesgo por cada sector 59
Ilustración 16. Riesgo general Valle de Aburra 59
Ilustración 17. Comparativa de variables 60

Índice de tablas

Tabla 1 Tabla requisitos de usuario 43
Tabla 2. Requisitos funcionales 44
Tabla 3. Requisitos funcionales 44
Tabla 4. Variables de incendios 46
Tabla 5. Factores por sector 47
Tabla 6. Métricas de evaluación 57
Tabla 7. Resultados módulo deslizamientos 62
Tabla 8. Resultados módulo incendio 63
Tabla 9. Resultados módulo inundaciones 64

1 Resumen

El área metropolitana del Valle de Aburrá presenta condiciones geográficas y climáticas que la hacen especialmente vulnerable a eventos hidrometeorológicos extremos como deslizamientos, inundaciones e incendios, por lo que ante la ausencia de una herramienta de software integrada que permita analizar y visualizar el riesgo por sector de forma automatizada, el presente proyecto desarrolla un sistema de software modular compuesto por tres módulos de predicción independientes. Cada módulo extrae datos históricos y en tiempo real desde la API Open-Meteo, construye variables hidrometeorológicas derivadas y entrena modelos de aprendizaje autónomo como XGBoost con series temporales horarias desde 2020 hasta 2026, para los modelos se incorporan factores geográficos específicos por sector como la pendiente, drenaje, inestabilidad del suelo, vegetación y exposición al viento con el fin de producir estimaciones del riesgo, estos resultados se visualizan mediante mapas interactivos Folium codificados por color y gráficos comparativos por sector. Los tres módulos presentan métricas de Accuracy y ROC-AUC superiores a 0.90, evidenciando la viabilidad técnica del sistema como herramienta de apoyo a la gestión del riesgo para instituciones como el SIATA, Bomberos y Defensa Civil en el Valle de Aburrá.

Palabras clave: sistema de alerta temprana, aprendizaje automático, XGBoost, Open-Meteo, gestión del riesgo, Valle de Aburrá, deslizamientos, inundaciones, incendios, visualización geoespacial.

2 Abstract

The Metropolitan Area of the Aburrá Valley, due to its geographical and climatic characteristics, faces high vulnerability to extreme hydrometeorological events such as landslides, floods, and wildfires. Given the lack of integrated tools for automated sector-based risk analysis, this research develops a modular software system focused on predicting and visualizing these threats. The implemented methodology integrates real-time and historical data extraction from the Open-Meteo API, the processing of hourly time series (spanning 2020-2026), and the training of classification models based on the XGBoost algorithm. A key innovation of the system is the incorporation of local geographic factors—including slope, drainage, soil instability, vegetation cover, and wind exposure—enabling a territorially contextualized risk estimation. The results, represented through interactive Folium maps and sectoral comparative analyses, demonstrate outstanding performance with Accuracy and ROC-AUC metrics exceeding 0.90 across all three modules. It is concluded that the system possesses high technical viability as a decision-support tool for risk management organizations such as SIATA, Firefighters, and Civil Defense within the region.

Keywords: early warning system, machine learning, XGBoost, Open-Meteo, risk management, Aburrá Valley, landslides, floods, wildfires, geospatial visualization.

3 Glosario

API: Interfaz de programación que permite la comunicación entre sistemas de software mediante protocolos definidos, en este proyecto, Open-Meteo se consume como API REST para la obtención de datos meteorológicos.

Accuracy: Métrica de evaluación de modelos de clasificación que mide la proporción de predicciones correctas sobre el total de predicciones realizadas.

Dashboards: Interfaz visual que consolida información de múltiples fuentes en gráficos, indicadores y mapas para facilitar la toma de decisiones.

Folium: Librería Python que permite la generación de mapas interactivos basados en Leaflet.js, utilizada en este proyecto para la visualización geoespacial del riesgo.

Gradient Boosting: Técnica de ensamble de modelos de aprendizaje automático que construye árboles de decisión de forma secuencial, corrigiendo los errores del modelo anterior.

IDEAM: Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia. Entidad responsable del monitoreo hidrometeorológico nacional.

Índice de saturación del suelo: Variable derivada que cuantifica la capacidad del suelo para absorber agua adicional, calculada como el producto de la humedad del suelo y la lluvia acumulada.

Índice de sequedad: Variable derivada que combina temperatura, baja humedad relativa y baja humedad del suelo para estimar el riesgo de ignición de incendios.

Open-Meteo: API meteorológica de código abierto que provee datos históricos y pronósticos para cualquier coordenada geográfica mundial, sin requerir autenticación para uso no comercial.

Plotly: Librería Python de visualización interactiva que permite generar gráficos en formato HTML, se utiliza para construir componentes Dashboard como indicador de velocímetro, gráficos de barras, gráfico de torta y diagramas de dispersión.

Requerimiento funcional: Especificación de una función o comportamiento que el sistema debe realizar.

Requerimiento no funcional: Especificación de una cualidad o restricción del sistema, como rendimiento, modularidad o mantenibilidad.

ROC-AUC: Métrica que evalúa la capacidad discriminativa de un modelo de clasificación binaria con valores cercanos a 1.0 indican alta capacidad predictiva.

SIATA: Sistema de Alerta Temprana del Valle de Aburrá. Red de monitoreo ambiental e hidrológico operada por el Área Metropolitana del Valle de Aburrá.

XGBoost: Algoritmo de aprendizaje automático supervisado basado en gradient boosting sobre árboles de decisión, reconocido por su eficiencia y precisión en tareas de clasificación con datos tabulares.

4 Introducción

Los fenómenos hidrometeorológicos extremos representan uno de los principales desafíos para la gestión del riesgo en zonas urbanas con topografía compleja, en el caso del área metropolitana del Valle de Aburrá, la cual está conformada por Medellín y diez municipios aledaños, concentra una población superior a los tres millones de habitantes en un territorio caracterizado por laderas empinadas, una densa red de quebradas afluentes del río Medellín y altos niveles de urbanización informal en zonas de alta pendiente, por lo que esta combinación de factores convierte a la región en un territorio especialmente vulnerable a eventos de deslizamiento, inundación e incendio ante episodios de precipitación intensa o condiciones de sequedad prolongada.

Actualmente, el monitoreo meteorológico regional se realiza mediante redes de sensores operadas por entidades como el SIATA y el IDEAM, cuyos datos son procesados en gran parte de forma manual y consultados a través de plataformas institucionales segmentadas. Esta fragmentación dificulta el análisis integrado y oportuno de la información en situaciones de emergencia, generando una brecha tecnológica concreta, la ausencia de una herramienta de software que integre datos de múltiples variables meteorológicas, los analice mediante técnicas de inteligencia artificial y presente los resultados de forma visual e interactiva para apoyar la toma de decisiones institucional.

El presente proyecto responde a esta necesidad mediante el desarrollo de un sistema de software modular compuesto por tres módulos de predicción independientes para deslizamientos, inundaciones e incendios; el sistema utiliza la API Open-Meteo como fuente de datos hidrometeorológicos históricos y en tiempo real, implementa modelos de clasificación XGBoost entrenados con series temporales horarias de seis años, e incorpora factores geográficos específicos por sector para contextualizar territorialmente las estimaciones de riesgo además los resultados se visualizan mediante mapas interactivos y gráficos comparativos por sector en el Valle de Aburrá.

El presente documento se estructura en nueve secciones, las secciones 1 y 2 presentan el planteamiento del problema y la justificación del proyecto, la sección 3 define los objetivos generales y específicos, las secciones 4 y 5 desarrollan el marco teórico y el marco conceptual que sustentan el sistema. La sección 6 describe la metodología de desarrollo adoptada, la sección 7 presenta el desarrollo del proyecto, incluyendo el levantamiento de información, el análisis de requerimientos, el diseño, el desarrollo de los módulos, las pruebas de evaluación y la visualización de resultados, en las secciones 8 y 9 se presentan las conclusiones y recomendaciones y para finalmente concluir con las referencias bibliográficas y los anexos con la descripción estructural de cada módulo.

5 Planteamiento del problema

5.1 Descripción

Los fenómenos meteorológicos extremos, como inundaciones, deslizamientos e incendios, representan una amenaza que está en aumento para las comunidades urbana, por lo cual el alcance de este problema se manifiesta de manera diferente según el territorio, y es en el área metropolitana del Valle de Aburrá es donde se presenta una gran vulnerabilidad debido a la gran diferencia de ambientes en cada una de las zonas.

Medellín y su área metropolitana presentan condiciones geográficas, climáticas y urbanas que la hacen estar muy expuesta debido a su topografía montañosa, el alto nivel de urbanización en laderas y la extensa red hídrica conformada por quebradas y el río Medellín configuran un entorno en el que los episodios de precipitación intensa pueden desencadenar situaciones de emergencia en cuestión de horas. Por lo que, en estos casos, el monitoreo oportuno de las condiciones meteorológicas y ambientales constituye un factor crítico para la gestión del riesgo y la protección de la vida humana.

Actualmente, el monitoreo meteorológico en el Valle de Aburrá se lleva a cabo mediante entidades como el SIATA y el IDEAM los cuales operan redes de sensores distribuidos en la región, incluyendo pluviómetros, estaciones meteorológicas y sensores de humedad del suelo. Los datos recolectados son consultados por técnicos y tomadores de decisiones a través de plataformas institucionales, y en algunos casos se procesan de forma manual para elaborar informes de riesgo o activar protocolos de alerta.

No obstante, este proceso presenta limitaciones dado que los datos provienen de distintas redes de monitoreo dispersas en diferentes fuentes y formatos, lo que dificulta su integración y análisis en conjunto, además el procesamiento manual de la información implica demoras que pueden resultar críticas en situaciones de emergencia, por lo que estos modelos de análisis tradicionales presentan baja capacidad predictiva frente a eventos de rápida evolución, y las herramientas de visualización disponibles no siempre permiten una

interpretación ágil e integrada de la información para la toma de decisiones.

Esta situación es la que revela una gran brecha tecnológica al entender que no existe una solución de software suficientemente integrada que permita reunir datos meteorológicos de múltiples fuentes, analizarlos de manera automatizada mediante técnicas avanzadas y presentarlos de forma visual e interactiva para apoyar la gestión del riesgo, por lo que el uso de técnicas de inteligencia artificial representa una posibilidad concreta de fortalecimiento tecnológico, al permitir procesar grandes volúmenes de datos, identificar patrones de comportamiento y estimar niveles de riesgo con mayor precisión y oportunidad que los métodos convencionales.

Ante este panorama, el presente proyecto se plantea como una respuesta de ingeniería de software orientada a mejorar el análisis, la visualización y el uso estratégico de la información para la gestión del riesgo en el área metropolitana del Valle de Aburrá. No se trata de sustituir los sistemas existentes, sino de desarrollar una herramienta complementaria que integre y potencie la información disponible, contribuyendo a que las instituciones competentes dispongan de mejores insumos tecnológicos para la toma de decisiones.

5.2 Formulación

¿Cómo desarrollar un sistema de software que integre datos hidrometeorológicos del Valle de Aburrá provenientes de fuentes como Open-Meteo, aplique técnicas de inteligencia artificial para estimar el nivel de riesgo de deslizamientos, inundaciones e incendios por sectores, y visualice los resultados mediante un dashboards interactivo que apoye la toma de decisiones en la gestión del riesgo?

6 Justificación

Desde la perspectiva tecnológica, el proyecto responde a una necesidad concreta de integración de datos hidrometeorológicos que actualmente se encuentran dispersos en múltiples fuentes. La API Open-Meteo, utilizada en el presente proyecto, permite acceder de forma gratuita y estructurada a datos históricos y en tiempo real de variables como temperatura, humedad relativa, precipitación, humedad del suelo y velocidad del viento, con una cobertura para cualquier punto del Valle de Aburrá.

Los módulos de predicción desarrollados como deslizamientos, inundaciones e incendios implementan modelos XGBoost entrenados con series temporales horarias desde 2020 hasta 2026, procesando variables derivadas como lluvia acumulada en 24h y 72h, índices de saturación del suelo e índices de sequedad, por lo que estos modelos alcanzan métricas de precisión superiores al 90%, evidenciando que la integración de IA con datos meteorológicos estructurados es una solución técnicamente viable y eficiente para la estimación de riesgo por sectores en el área metropolitana.

Este proyecto constituye un trabajo de grado pertinente dentro del programa de l tecnología en desarrollo de software, pues integra de manera práctica competencias del área de ingeniería de software, desde el diseño de módulos de procesamiento de datos, implementación de modelos de machine learning mediante la librería XGBoost, desarrollo de visualizaciones geoespaciales con Folium, y estructuración de un sistema modular con Python. Además, el proyecto no se limita a replicar una solución existente, sino que construye un sistema propio adaptado a las condiciones geográficas y sociales del Valle de Aburrá, con

factores geográficos específicos como lo son la pendiente, el drenaje, la vegetación y la inestabilidad del suelo según la probabilidad por cada sector.

Desde el ámbito social, el sistema desarrollado aporta una herramienta que puede apoyar la gestión del riesgo y la toma de decisiones de instituciones como el SIATA, Bomberos, Defensa Civil y alcaldías locales, de los 16 a 21 sectores urbanos y rurales del Valle de Aburrá modelados en el sistema se incluyen comunas con alta vulnerabilidad histórica ante deslizamientos e inundaciones, como Popular, Villa Hermosa, San Javier y Buenos Aires, por lo que al momento de visualizar el riesgo estimado en mapas interactivos y gráficos por sector, la herramienta facilita que los tomadores de decisiones puedan priorizar recursos y comunicar alertas de manera más oportuna, sin reemplazar los sistemas institucionales existentes.

En síntesis, el presente proyecto responde a una necesidad tecnológica concreta y actual el proveer a las instituciones del Valle de Aburrá de una herramienta de software que transforme datos meteorológicos abiertos en estimaciones de riesgo accionables por sector, reduciendo la brecha entre la disponibilidad de información y la capacidad de respuesta oportuna ante eventos hidrometeorológicos extremos.

7 Objetivos

7.1 Objetivo general

Desarrollar un sistema de software que integre datos hidrometeorológicos del Valle de Aburrá provenientes de la API Open-Meteo y utilice técnicas de inteligencia artificial para estimar el nivel de riesgo de deslizamientos, inundaciones e incendios por sector, visualizando los resultados mediante mapas interactivos y gráficos que apoyen la gestión del riesgo.

7.2 Objetivos específicos

Identificar y recopilar datos hidrometeorológicos históricos y en tiempo real provenientes de la API Open-Meteo para el área del Valle de Aburrá, procesando variables como precipitación, temperatura, humedad relativa, humedad del suelo y presión superficial en series de tiempos específicas.

Diseñar e implementar módulos de inteligencia artificial basados en el algoritmo XGBoost para la estimación del riesgo de deslizamientos, inundaciones e incendios, incorporando variables derivadas como lluvia acumulada, índices de saturación del suelo e índices de sequedad.

Desarrollar un sistema de visualización geoespacial mediante mapas interactivos y gráficos de barras que permita representar el nivel de riesgo estimado por sector en el área

metropolitana del Valle de Aburrá.

Evaluar el desempeño de los modelos desarrollados mediante métricas de clasificación y validar la coherencia de las predicciones por sector con los factores geográficos conocidos de cada zona.

Integrar los módulos de predicción y visualización en una solución de software funcional que permita la estimación dinámica del riesgo a partir de datos meteorológicos actuales.

8 Marco teórico

8.1 Teoría general de sistemas

La Teoría General de Sistemas, derivada de los trabajos de Ludwig von Bertalanffy (1968) y consolidada en décadas posteriores, constituye el fundamento conceptual para el análisis de sistemas complejos abiertos como los sistemas de monitoreo ambiental. La TGS enfatiza la interdependencia de las partes, la estructura organizativa y la totalidad del sistema, características esenciales para comprender un sistema integrado de monitoreo meteorológico con inteligencia artificial, en la gestión contemporánea de desastres, esta aproximación facilita la identificación de retroalimentaciones y relaciones causa-efecto que impactan en la detección y respuesta ante eventos extremos.

8.2 Sistemas de alerta temprana y gestión del riesgo

Los sistemas de alerta temprana se conciben como componentes críticos de la gestión de riesgos, estructurados en cuatro pilares: monitoreo continuo, análisis predictivo, comunicación efectiva y capacidad de respuesta (UNDRR, 2025). La incorporación de inteligencia artificial en estos sistemas permite la integración de grandes volúmenes de datos heterogéneos y la generación automática de alertas con mayor precisión y oportunidad. En el contexto del presente proyecto, el sistema desarrollado opera como un componente de apoyo a la decisión institucional, generando estimaciones de riesgo por sector sin reemplazar los mecanismos de comunicación pública del SIATA u otras entidades competentes.

8.3 Aprendizaje automático para predicción meteorológica

Los avances en aprendizaje automático han permitido desarrollar modelos predictivos basados en técnicas supervisadas que superan los métodos numéricos tradicionales en velocidad y precisión para pronósticos hidrometeorológicos (Atmospheric Environment, 2024). En el presente proyecto se utiliza XGBoost (Extreme Gradient Boosting), un algoritmo de ensamble basado en árboles de decisión con gradient boosting desarrollado por Chen y Guestrin (2016), que ha demostrado resultados destacados en tareas de clasificación con datos tabulares, estos modelos se entrenan con datos horarios históricos de 2020 a 2026, utilizando un esquema de división temporal 80/20 para garantizar la validez de los resultados sobre datos no vistos.

8.4 Open-Meteo como fuente de datos hidrometeorológicos

Open-Meteo es una API meteorológica de código abierto que provee datos históricos y pronósticos para cualquier ubicación geográfica del mundo, sin requerir autenticación para uso no comercial (Open-Meteo, 2024). En el contexto de este proyecto, se utilizan dos endpoints: el archivo histórico para el entrenamiento de los modelos con datos desde el 1 de enero de 2020, y el endpoint de pronóstico para la estimación de riesgo en tiempo real por sector, por lo que las variables extraídas incluyen temperatura a 2 m, humedad relativa, precipitación, presión superficial, velocidad del viento a 10 m y humedad del suelo en la capa de 0-7 cm.

8.5 Visualización geoespacial para gestión del riesgo

La visualización de datos geoespaciales constituye un componente esencial en los sistemas de apoyo a la decisión para gestión del riesgo (iScience, 2025). En el presente proyecto se utiliza Folium, una librería Python que permite generar mapas interactivos basados en Leaflet.js, con marcadores codificados por color según el nivel de riesgo estimado, adicionalmente se utilizan Seaborn y Matplotlib para la generación de gráficos de barras que permiten comparar el nivel de riesgo entre sectores de forma clara y ordenada.

8.6 Variables hidrometeorológicas relevantes para predicción de riesgos

El análisis de riesgo por tipo de evento requiere la identificación de las variables hidrometeorológicas más relevantes para cada fenómeno, en el caso de los deslizamientos, las variables críticas son la lluvia acumulada en 24h y 72h, la humedad del suelo y el índice de saturación, para las inundaciones urbanas, los acumulados de lluvia en 3h, 6h, 12h y 24h, combinados con la humedad del suelo, determinan la capacidad de infiltración y el riesgo de excedentes hídricos y con respecto a los incendios, el índice de sequedad y la velocidad del viento son los factores determinantes (Nature Communications, 2025).

8.7 Antecedentes

8.7.1 Antecedentes geográficos

Con el fin de presentar los antecedentes correspondientes al enfoque que presentamos para nuestro proyecto de manera local tenemos a el SIATA es operado por el Área

Metropolitana del Valle de Aburrá y constituye el referente institucional más directo para este proyecto, se enfoca en el monitoreo hidrometeorológico en tiempo real mediante una red de más de 300 sensores distribuidos en la región, incluyendo pluviómetros, sensores de nivel de ríos, estaciones meteorológicas y cámaras de videovigilancia. Las variables monitoreadas incluyen precipitación, nivel hídrico de quebradas, temperatura y humedad relativa, por lo que sus resultados se comunican a través de alertas públicas codificadas por colores y boletines institucionales dirigidos a entidades como Bomberos, Defensa Civil y las alcaldías, por lo que la mayor diferencia con el presente proyecto radica en que el SIATA opera desde la infraestructura física de sensores y la comunicación reactiva de alertas, sin incorporar modelos de aprendizaje automático para la estimación predictiva de riesgo por sector. Por lo que nuestro sistema desarrollado actúa como una capa complementaria de análisis predictivo que puede operar con datos de fuentes abiertas como Open-Meteo, sin depender de la red de sensores físicos.

Otro actor que opera de manera local y se constituye como referente hacia el proyecto es el DAGRD es la entidad del municipio de Medellín encargada de coordinar la respuesta ante emergencias y de gestionar el conocimiento del riesgo en el territorio, al publicar esquemas de amenaza por deslizamientos e inundaciones para las comunas de Medellín, elaborada con base en estudios geológicos, históricos y de modelación hidráulica. Sin embargo, estas herramientas son de naturaleza estática lo que reflejan condiciones de riesgo estructural del territorio pero no incorporan variables meteorológicas dinámicas ni modelos que actualicen la estimación de riesgo en función de las condiciones climáticas actuales, por lo que el proyecto llega a complementar este enfoque al generar estimaciones de riesgo que varían con los datos meteorológicos en tiempo real, permitiendo diferenciar el nivel de

amenaza entre sectores según las condiciones del momento.

Al momento de analizar los actores a nivel nacional nos encontramos con el IDEAM es la autoridad técnico-científica de Colombia en materia de información hidrometeorológica y ambiental, esta opera una red nacional de estaciones de monitoreo y publica modelos de pronóstico del tiempo, alertas por crecientes súbitas y boletines de riesgo climático, además ha incorporado técnicas de predicción computacional para el pronóstico de precipitación extrema a escala nacional, por lo que a diferencia con el presente proyecto es de escala y propósito en el caso del IDEAM trabaja a escala nacional y regional con resolución espacial amplia, mientras que este sistema opera a escala de sectores urbanos y rurales del Valle de Aburrá, incorporando factores geográficos locales como la pendiente, el drenaje y la cobertura vegetal por zona, lo que permite una estimación de riesgo territorialmente contextualizada que el IDEAM no provee en sus análisis.

8.7.2 Antecedente académico

Continuando con el contexto de los antecedentes en especial sobre el tema de predicción de deslizamientos con aprendizaje automático se pudo investigar sobre un trabajo en especial realizado, Dikshit et al. (2020), en su investigación "Pathways to advance machine learning-based landslide early warning systems", publicada en la revista Landslides, revisaron el estado del arte de los sistemas de alerta temprana para deslizamientos basados en ML, estos autores evaluaron algoritmos como Random Forest, SVM y redes neuronales aplicados a datos de lluvia acumulada, humedad del suelo y características geomorfológicas, sus resultados mostraron que los modelos basados en árboles de decisión superan a los

modelos lineales en presencia de variables no lineales como la saturación del suelo. El presente proyecto adopta XGBoost, una evolución del gradient boosting sobre árboles de decisión, e incorpora variables similares como la lluvia acumulada en 24h y 72h y el índice de saturación, diferenciándose por la escala por sector que se aplica al Valle de Aburrá y el uso de la API Open-Meteo como única fuente de datos abierta.

En consecuencia, de los antecedentes y el tema de predicción de inundaciones urbanas y mediante la investigación realizada encontramos este articulo donde nos habla sobre Wang et al. (2022), en el artículo "Urban flood prediction under heavy precipitation", publicado en el Journal of Hydrology, entrenaron modelos XGBoost con datos horarios de precipitación, humedad del suelo y nivel hídrico para predecir inundaciones en zonas urbanas de China. Alcanzaron un ROC-AUC de 0.92 y destacaron que los acumulados de lluvia en ventanas cortas como periodos de entre 3h y 6h son las variables con mayor poder predictivo. El presente proyecto replica esta estrategia con características de análisis como los acumulados de 3h, 6h, 12h y 24h y obtiene métricas comparables como los son la ROC-AUC > 0.94, diferenciándose por la incorporación de datos en tiempo real desde Open-Meteo y por la visualización del riesgo en mapas interactivos Folium con resolución de cuadrícula de 0.01°.

Si profundizamos hacia el lado de la predicción de incendios con índices de sequedad se pudo obtener el siguiente sistema desarrollado Jain et al. (2020), en "A review of machine learning applications in wildfire science and management" (Environmental Research Letters), sistematizaron el uso de ML para la predicción de incendios forestales, estos autores identificaron que los modelos basados en gradient boosting, combinados con índices de sequedad derivados de temperatura, humedad relativa y humedad del suelo, son los de mayor desempeño para clasificación binaria de riesgo de ignición. El presente proyecto construye

un índice de sequedad propio el cual es temperatura × 0.4 + (100 - humedad) × 0.3 + (1 - suelo) × 100 × 0.3 y lo integra como variable central del módulo de incendios, diferenciándose por el ajuste de factores geográficos de vegetación y exposición al viento por sector, que no aparecen en la revisión de Jain et al.

8.7.3 Antecedente tecnológico

Al seguir presentando los hallazgos de nuestra investigación nos encontramos desde el lado tecnológico con una gran herramienta llamada Aurora es un modelo de inteligencia artificial desarrollado por Microsoft Research y presentado en 2024, entrenado con más de un millón de horas de datos atmosféricos. Según la WMO (2025), Aurora supera a los modelos numéricos tradicionales de predicción del tiempo en precisión para pronósticos de precipitación extrema a escala global, pero sin embargo, Aurora opera como un modelo de pronóstico meteorológico general y no está diseñado para la estimación de riesgo de desastres específicos por sector urbano, del cual es el enfoque específico del proyecto el cual se enfoca en la clasificación de riesgo de deslizamientos, inundaciones e incendios para sectores concretos del Valle de Aburrá, incorporando factores geográficos locales que un modelo global como Aurora no considera.

Adicional como herramienta que pudimos lograr mediante nuestra investigación fue Open-Meteo es una API meteorológica de código abierto que integra datos de modelos numéricos globales como ERA5 de Copernicus/ECMWF para proveer series históricas horarias desde 1940 y pronósticos en tiempo real para cualquier coordenada geográfica, ya que a diferencia de APIs comerciales como OpenWeatherMap o Tomorrow.io, no requiere autenticación para uso no comercial y ofrece variables de humedad del suelo que son

esenciales para los modelos de deslizamiento e inundación, por lo que su uso en el presente proyecto elimina la dependencia de infraestructura de sensores propios, lo que reduce los costos de implementación, y permite replicar el sistema en cualquier región del país con coordenadas geográficas conocidas, además esta herramienta es la que más viabilidad a el proyecto al eliminar las dependencias antes mencionadas.

8.8 Marco legal y normativo

El desarrollo del presente sistema de software está enmarcado en un conjunto de normas nacionales e internacionales que orientan tanto la gestión del riesgo de desastres como el uso de información ambiental, la protección de datos y las buenas prácticas en el desarrollo de software, por lo que a continuación se explica cada norma en función del proyecto.

8.8.1 Ley 1523 de 2012 Política Nacional de Gestión del Riesgo de Desastres

Esta ley se establece el marco legal para la gestión del riesgo en Colombia, crea el Sistema Nacional de Gestión del Riesgo de Desastres (SNGRD) y se define el conocimiento del riesgo, la reducción del riesgo y el manejo de desastres como los tres procesos fundamentales de la política pública, por lo que en su artículo 4, la ley reconoce la información hidrometeorológica como un insumo esencial para el conocimiento del riesgo y la toma de decisiones. El presente proyecto se alinea con este marco al desarrollar una herramienta de software que contribuye al conocimiento del riesgo mediante la estimación automatizada del nivel de amenaza por sector, apoyando las decisiones de entidades como el

SIATA, Bomberos y el DAGRD sin reemplazar sus sistemas institucionales.

8.8.2 Decreto 1974 de 2013 Conformación del SNGRD

Este decreto reglamenta la integración de entidades del orden nacional, departamental y municipal en el sistema de gestión del riesgo, el cual establece que las entidades como el IDEAM y el SIATA son actores clave en la generación de información técnica para la gestión del riesgo, por lo que el presente sistema se concibe como una herramienta complementaria que puede integrarse al flujo de información de estas entidades, procesando datos de fuentes abiertas y generando estimaciones de riesgo que enriquecen el análisis territorial.

8.8.3 Ley 1581 de 2012 Protección de datos personales

Esta ley regula el tratamiento de datos personales en Colombia, aunque el presente proyecto no recolecta ni procesa datos personales de ciudadanos, pero opera con datos meteorológicos y geográficos de fuentes abiertas y es pertinente en cuanto establece el principio de finalidad y transparencia en el uso de información, ya que el sistema cumple este principio al usar datos de Open-Meteo bajo su licencia de uso no comercial abierto y al no almacenar información que identifique personas.

8.8.4 Ley 23 de 1982 y Decisión Andina 351 de 1993

La Ley 23 de 1982 y la Decisión 351 de la Comunidad Andina protegen los derechos de autor sobre obras intelectuales, incluyendo el software, por lo que los notebooks de Python, los modelos entrenados y el código de visualización desarrollados en este proyecto

constituyen una obra intelectual original de los autores, además el uso de librerías de código abierto como XGBoost, Folium , Pandas y Scikit-learn es compatible con estas normas, y su uso en un proyecto académico sin fines comerciales se ajusta a los términos de cada licencia.

8.8.5 ISO/IEC 25010:2011 Calidad del producto de software

Este estándar internacional define las características de calidad de un sistema de software, incluyendo funcionalidad, rendimiento, usabilidad, fiabilidad, mantenibilidad y portabilidad. El presente proyecto incorpora varios de estos atributos en sus requerimientos no funcionales el requerimiento RNF-01 establece un tiempo de respuesta, el RNF-04 garantiza la modularidad del sistema o un enfoque en el mantenimiento de la aplicación y el RNF-03 asegura la reproducibilidad de resultados ante los mismos datos de entrada enfocado en la fiabilidad, teniendo claro que con todo esto se le da cumplimiento a este estándar que orienta el desarrollo hacia un producto de software confiable y evaluable objetivamente.

8.8.6 Términos de uso de Open-Meteo

La API Open-Meteo está disponible bajo una licencia de uso libre para aplicaciones no comerciales, con atribución requerida al proveedor, por lo que el uso en el presente proyecto cumple con estos términos al tratarse de un proyecto académico sin fines comerciales por lo que esta licencia permite el uso del endpoint histórico y del endpoint de pronóstico, ambos utilizados en los módulos de predicción.

8.9 Tecnologías y herramientas utilizadas en el desarrollo del sistema

Esta sección se busca presentar las tecnologías seleccionadas para el desarrollo del sistema, explicando su función específica dentro del proyecto, su aporte técnico y la razón de su elección frente a otras alternativas disponibles.

8.9.1 Python 3.x

Python es el lenguaje de programación principal del sistema, para iniciar a analizar su función nos enfocares en tres razones este lenguaje tiene el ecosistema más amplio de librerías para ciencia de datos y machine learning disponible en cualquier lenguaje con el fin de mencionar algunos podemos hablar de NumPy, Pandas, y XGBoost, su sintaxis legible que facilita el mantenimiento del código, y su compatibilidad nativa con entornos de ejecución como Google Colab, que permiten combinar código, resultados y documentación en un mismo archivo, teniendo todo lo anterior en cuenta por eso nos decidimos por Python actúa como el orquestador de todo el flujo desde la extracción de datos, procesamiento, entrenamiento de modelos y generación de visualizaciones.

8.9.2 XGBoost

XGBoost es el algoritmo de aprendizaje automático central del sistema, es un método de ensamble basado en árboles de decisión con gradient boosting, desarrollado por Chen y Guestrin (2016). Se eligió sobre alternativas como Random Forest y regresión logística porque maneja eficientemente datos con relaciones entre variables en nuestro caso como la relación entre lluvia acumulada y riesgo de deslizamiento, además tolera datos faltantes sin

preprocesamiento adicional, provee importancia de variables que permite interpretar qué factores son más determinantes para cada tipo de riesgo, en el sistema, se instancian tres modelos XGBClassifier independientes, uno por cada amenaza, configurados con 300–350 estimadores, profundidad máxima de 6, learning rate de 0.05 y submuestreo de 0.9 para prevenir sobreajuste.

8.9.3 Pandas y Numpy

Pandas es la librería de manipulación de datos estructurados de Python, en el proyecto se usa para cargar la respuesta JSON de la API Open-Meteo en un DataFrame, construir las variables de lluvia, índices de saturación y medias móviles, además dividir los datos en conjuntos de entrenamiento y prueba, y preparar los datos de entrada para el modelo. Además, NumPy complementa a Pandas en operaciones matemáticas vectorizadas, como el cálculo del índice de sequedad y los umbrales percentílicos para la generación de etiquetas por lo que ambas librerías son estándar de la industria en proyectos de ciencia de datos y no tienen alternativas equivalentes en términos de rendimiento y madurez.

8.9.4 Scikit-learn

Scikit-learn provee las herramientas de evaluación de modelos usadas en el proyecto como lo son accuracy_score, roc_auc_score, precision_score, recall_score y classification_report para medir el desempeño de cada módulo XGBoost sobre el conjunto de prueba, también se usa para la división temporal de los datos con el fin de respetar el orden cronológico, por esto se eligió sobre herramientas de evaluación propias de XGBoost porque ofrece una interfaz unificada y además compatible con cualquier clasificador de scikit-learn, facilitando comparaciones futuras con otros algoritmos.

8.9.5 Requests

La librería requests de Python es el cliente HTTP usado para consultar la API Open-Meteo lo cual nos permite construir las peticiones GET al endpoint histórico y al endpoint de pronóstico con los parámetros de coordenadas, variables y rango de fechas, y gestionar la respuesta JSON, se eligió sobre urllib por su sintaxis más limpia y por incluir manejo automático de redirecciones y errores de conexión.

8.9.6 Open-Meteo API

Open-Meteo es la fuente de datos hidrometeorológicos del sistema, es la que nos provee series temporales horarias de temperatura a 2 m, humedad relativa, precipitación, presión superficial, velocidad del viento a 10 m y humedad del suelo en la capa de 0–7 cm, con cobertura global y datos históricos desde 1940, su elección frente a alternativas comerciales como OpenWeatherMap Pro o Tomorrow.io se justifica por tres factores, el principal es su acceso libre sin autenticación para uso no comercial, disponibilidad de la variable de humedad del suelo la cual es esencial para los módulos de deslizamiento e inundación, de las cuales no se presentan en otras APIs gratuitas y datos históricos horarios sin costo, que son indispensables para el entrenamiento de los modelos con series entre los años 2020 y 2026.

8.9.7 Folium

Folium es la librería Python para la generación de mapas interactivos basados en Leaflet.js, en el sistema se usa para producir los mapas de riesgo por sector en formato

HTML, con marcadores circulares codificados por color como el verde para riesgo bajo (< 40%), naranja para riesgo medio (40–70%) y rojo para riesgo alto (> 70%), además para el módulo de inundaciones genera además un mapa de calor en cuadrícula de resolución 0.01° sobre la región metropolitana, se decidió elegir Folium sobre alternativas como Plotly Maps o GeoPandas porque genera archivos HTML autocontenidos que pueden abrirse en cualquier navegador sin servidor, lo que es coherente con la naturaleza de prototipo académico del sistema y facilita la verificación de resultados.

8.9.8 Seaborn y Matplotlib

Seaborn y Matplotlib son las librerías de visualización estática del sistema, en el caso de Matplotlib es la base de graficación de Python y Seaborn es una capa de alto nivel construida sobre ella que simplifica la creación de gráficos estadísticos. En el proyecto se usan con el fin de generar los gráficos de barras horizontales ordenados por nivel de riesgo descendente para cada módulo, construir la matriz de variables del módulo de deslizamientos, y visualizar las métricas de evaluación de los modelos, se eligieron sobre Plotly porque estas librerías pueden producir gráficos en formato PNG de alta resolución adecuados para inserción en el informe, sin dependencias adicionales de servidor o navegador.

8.9.9 Google Colab

Es el entorno de desarrollo y ejecución del sistema Google Colab nos presenta grandes herramienta sin instalación local, por lo que este entorno permite combinar en un mismo archivo las celdas de código Python, los resultados numéricos, los gráficos generados y la documentación en Markdown, lo que facilita tanto el desarrollo iterativo como la revisión

y verificación del sistema por parte de evaluadores, por todo lo anterior se eligió sobre un script Python tradicional porque permite ejecutar el sistema por secciones, visualizar los resultados inmediatamente y documentar cada paso del proceso dentro del mismo archivo fuente.

8.9.10 SQLite

SQLite es un motor de base de datos relacional liviano integrado de forma nativa en Python mediante la librería estándar sqlite3, lo que significa que no requiere instalación adicional ni configuración de servidor para su funcionamiento. En el sistema desarrollado, cada módulo genera al finalizar su ejecución una base de datos SQLite en formato de archivo portable, que persiste dos tablas estructuradas, la primera almacena la serie temporal horaria completa utilizada para el entrenamiento, con todas las variables brutas extraídas de Open-Meteo y las variables derivadas construidas durante el procesamiento; la segunda almacena los resultados de predicción por sector, incluyendo el porcentaje de riesgo calculado y el nivel de alerta asignado. Se eligió SQLite sobre alternativas como PostgreSQL o MySQL porque opera como un archivo local sin dependencia de servicios externos, es completamente compatible con el entorno Google Colab sin configuración adicional, y permite que los resultados sean consultables mediante SQL por usuarios técnicos o exportables a herramientas de hoja de cálculo por usuarios no técnicos, ampliando el alcance institucional del sistema sin incrementar su complejidad de despliegue.

8.9.11 Plotly

Plotly es una librería Python de visualización interactiva que permite generar gráficos en formato HTML con JavaScrip, de modo que el resultado puede abrirse y explorarse en

cualquier navegador sin requerir servidor ni ejecución de código adicional. En el sistema desarrollado, Plotly se utiliza exclusivamente para la construcción de los dashboards de resultados de cada módulo, donde genera cuatro tipos de componentes, el indicador de velocímetro tipo gauge que muestra el riesgo promedio del área metropolitana en una escala de 0 a 100 con zonas codificadas por color, los gráficos de barras horizontales ordenados por nivel de riesgo descendente para todos los sectores, el gráfico circular de distribución de niveles de alerta y los diagramas de dispersión que relacionan las variables meteorológicas principales con el porcentaje de riesgo calculado. Se eligió Plotly para los componentes del dashboard porque produce visualizaciones interactivas que el usuario puede explorar mediante zoom, filtros y tooltips informativos al pasar el cursor sobre cada sector, lo que enriquece significativamente la experiencia de consulta frente a los gráficos estáticos en formato PNG que generan las otras librerías. La combinación de Plotly para el dashboard y Seaborn/Matplotlib para los gráficos de análisis del modelo responde a las necesidades diferenciadas de cada contexto desde la interactividad para la presentación de resultados institucionales y alta resolución para la documentación técnica del informe.

9 Marco conceptual

9.1 Sistema de monitoreo meteorológico

Conjunto integrado de sensores, fuentes de datos y herramientas de software que permiten la recopilación, procesamiento y análisis continuo de datos hidrometeorológicos en tiempo real. En el presente proyecto, este sistema se implementa mediante la integración de la API Open-Meteo como fuente principal de datos, complementada con factores geográficos específicos por sector del Valle de Aburrá.

9.2 Inteligencia artificial

Conjunto de algoritmos y técnicas computacionales que permiten construir modelos capaces de identificar patrones en datos históricos y realizar estimaciones sobre eventos futuros. En este proyecto, la IA se materializa en tres modelos XGBClassifier independientes, uno para cada tipo de amenaza como en los deslizamientos, inundaciones e incendios, entrenados con variables hidrometeorológicas horarias y variables derivadas diseñadas para capturar los mecanismos físicos de cada fenómeno.

9.3 Sistema de alerta temprana (SAT)

Mecanismo integrado de vigilancia, previsión y comunicación orientado a reducir el impacto de eventos adversos mediante la anticipación y la toma de decisiones oportuna. El sistema desarrollado en este proyecto opera como un componente de apoyo a la decisión institucional, estimando niveles de riesgo por sector sin reemplazar los mecanismos de comunicación pública del SIATA u otras entidades competentes.

9.4 XGBoost

Algoritmo de aprendizaje automático supervisado basado en el ensamble de árboles de decisión con gradient boosting (Chen & Guestrin, 2016). Destaca por su eficiencia computacional, capacidad de manejo de datos faltantes y alto rendimiento en tareas de clasificación con datos tabulares, en los modelos desarrollados se configura con 300-350 estimadores, profundidad máxima de 6, tasa de aprendizaje de 0.05 y submuestreo de 0.9.

9.5 Redes de sensores ambientales y APIs meteorológicas

Infraestructuras de recopilación de datos que integran dispositivos físicos con plataformas digitales de acceso. Open-Meteo combina modelos numéricos de predicción del tiempo y análisis climático para ofrecer datos con resolución horaria, siendo utilizada en este proyecto como fuente principal tanto para entrenamiento histórico como para predicción en tiempo real.

9.6 Análisis predictivo por sectores

Proceso de estimación del nivel de riesgo diferenciado para unidades geográficas específicas, incorporando tanto las condiciones meteorológicas actuales como factores geográficos propios de cada sector. En el sistema desarrollado, cada sector del Valle de Aburrá tiene asignados factores específicos de pendiente, drenaje, inestabilidad del suelo, vegetación y exposición al viento, que modulan la predicción base del modelo para producir estimaciones generadas según el riesgo especial en cada territorio

10 Metodología

10.1 Tipo de proyecto

El presente trabajo corresponde a un proyecto de tipo aplicado-tecnológico, orientado al desarrollo de una solución de software funcional que responde a una necesidad concreta de gestión del riesgo en el área metropolitana del Valle de Aburrá, la investigación se apoya en datos reales provenientes de la API Open-Meteo y en técnicas de aprendizaje automático para la construcción de modelos predictivos validados empíricamente.

10.2 Metodología de desarrollo

El proyecto adoptó una metodología de desarrollo incremental, estructurada en fases que permitieron construir y validar cada módulo de forma independiente antes de su integración, dado esto las fases del proceso fueron las siguientes:

Como primera fase se realiza el levantamiento de información identificando fuentes de datos hidrometeorológicos disponibles y seleccionando a Open-Meteo como fuente principal por su cobertura, resolución horaria y acceso libre, luego se pasa a la segunda fasae donde se realiza el procesamiento de datos con pasos como la extracción, limpieza y construcción de variables derivadas para cada tipo de amenaza como lluvia acumulada, índices de saturación y sequedad, medias móviles, en la siguiente fase se desarrollo los modelos, donde se diseña e implementa los tres modelos XGBoost, incluyendo la definición de etiquetas de entrenamiento y la partición temporal 80/20, en la fase de validación se enfoca en temas como la evaluación de modelos con métricas de clasificación como Accuracy, ROC-AUC,

Precision, Recall sobre el conjunto de prueba y en la fase 5 se presenta la visualización e integración del desarrollo de mapas interactivos Folium y gráficos Seaborn/Matplotlib, e integración de los módulos en el sistema completo.

10.3 Cronograma de actividades

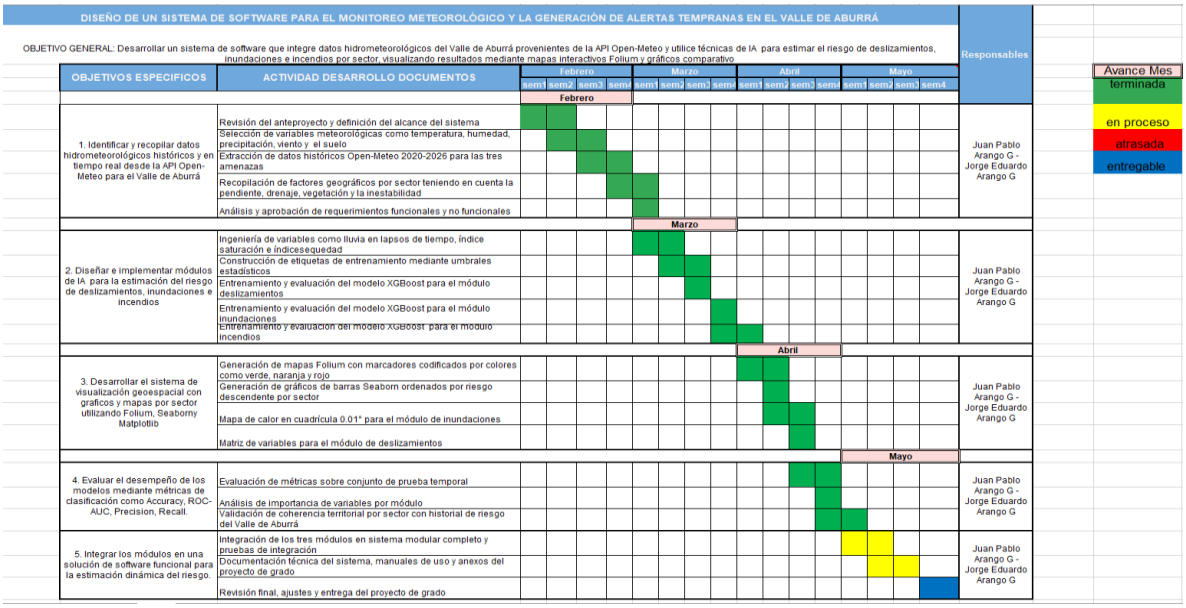


Ilustración 1. Cronograma de actividades

11 Resultados del desarrollo del proyecto

11.1 Levantamiento de información

La principal fuente de datos del sistema es la API Open-Meteo, seleccionada por su acceso libre, resolución horaria, cobertura global y disponibilidad de datos históricos desde 2020. Por lo que se consultaron dos endpoints, el archivo histórico para el entrenamiento de los modelos con datos desde el 1 de enero de 2020 hasta mayo de 2026, y el endpoint de pronóstico para la estimación de riesgo en tiempo real en el módulo de inundaciones.

Las coordenadas base utilizadas para la extracción de datos históricos fueron latitud 6.2442 y longitud -75.5812, correspondientes al centro del área metropolitana del Valle de Aburrá. Por lo que, para el módulo de inundaciones en tiempo real, se consultaron las coordenadas individuales de cada sector.

El dataset de entrenamiento consolidado para el módulo de deslizamientos alcanza aproximadamente 52.000 registros horarios distribuidos entre los 31 sectores analizados, abarcando el período enero 2020 y mayo 2026, para el proceso de limpieza de datos se incluyó la verificación de valores nulos en las variables críticas de precipitación y humedad del suelo, y la validación de rangos físicos esperados para cada variable. No se encontraron valores nulos en las variables principales, lo que confirma la robustez de la API Open-Meteo como fuente de datos para el período analizado.

11.2 Análisis de requerimientos

11.2.1 Requisitos de usuario

ID	Nombre del requisito	Descripción
RU-001	Consulta automática de datos meteorológicos	El sistema debe obtener automáticamente datos históricos y en tiempo real de temperatura, humedad relativa, precipitación, presión superficial, velocidad del viento y humedad del suelo para el Valle de Aburrá desde una fuente abierta, sin intervención manual del usuario
RU-002	Estimación de riesgo de deslizamiento por sector	El sistema debe calcular el nivel de riesgo de deslizamiento para cada uno de los 31 sectores del Valle de Aburrá, considerando la lluvia acumulada, la humedad del suelo y las características geográficas propias de cada zona como la pendiente y el drenaje

RU-003	Estimación de riesgo de inundación en tiempo real	El sistema debe estimar el riesgo de inundación por sector combinando datos meteorológicos en tiempo real con las características de drenaje urbano de cada zona, mostrando riesgo cero cuando las condiciones de lluvia no superan el umbral mínimo establecido
RU-004	Estimación de riesgo de incendio por sector	El sistema debe calcular el nivel de riesgo de incendio para cada sector, considerando la temperatura, la humedad relativa, la velocidad del viento y la cobertura vegetal propia de cada zona
RU-005	Visualización del riesgo en mapa interactivo	El sistema debe mostrar los resultados de riesgo de cada módulo en un mapa del Valle de Aburrá, donde cada sector esté representado por un marcador con color verde, naranja o rojo según el nivel de amenaza calculado, y que pueda consultarse desde un navegador sin instalar un software
RU-006	Identificación de los sectores con mayor y menor riesgo	El sistema debe mostrar de forma destacada cuál es el sector con mayor nivel de riesgo y cuál es el de menor riesgo en cada módulo, incluyendo las variables meteorológicas que explican ese resultado
RU-007	Comparación del riesgo entre sectores mediante gráficos	El sistema debe presentar gráficos que permitan comparar el nivel de riesgo entre los 31 sectores de forma ordenada, facilitando la identificación rápida de las zonas más vulnerables en cada tipo de amenaza
RU-008	Incorporación de características propias de cada sector	El sistema debe ajustar la estimación de riesgo de cada sector según sus condiciones geográficas particulares como la pendiente, el drenaje, la inestabilidad del suelo, la cobertura vegetal y la exposición al viento, de modo que dos sectores con las mismas condiciones meteorológicas puedan tener niveles de riesgo diferentes
RU-009	Reporte de confiabilidad del modelo	El sistema debe informar qué tan confiables son sus predicciones mediante métricas de evaluación como precisión, sensibilidad y capacidad, para que el usuario

		pueda valorar lo pertinente de los resultados
RU-010	Exportación de resultados para consulta posterior	El sistema debe guardar los resultados de cada ejecución en archivos que puedan ser consultados posteriormente por usuarios técnicos mediante consultas de base de datos y por usuarios no técnicos mediante herramientas de hoja de cálculo

Tabla 1 Tabla requisitos de usuario

11.2.2 Requerimientos funcionales

Número del requisito	Función	Descripción
RF-01	Extracción de datos de Open-Meteo	El sistema debe consultar la API Open-Meteo para obtener datos históricos y en tiempo real de variables hidrometeorológicas por coordenadas geográficas
RF-02	Procesamiento de variables derivadas	El sistema debe calcular variables como lluvia acumulada en intervalos de hora, índices de saturación del suelo e índices de sequedad
RF-03	Estimación de riesgo por sector	El sistema debe calcular un porcentaje de riesgo para cada sector del Valle de Aburrá, aplicando factores geográficos específicos sobre la predicción base del modelo
RF-04	Generación de mapas interactivos	El sistema debe generar mapas interactivos Folium con marcadores por color según el nivel de riesgo como verde, naranja y rojo representado por sector
RF-05	Generación de gráficos comparativos	El sistema debe generar gráficos de barras horizontales ordenados por nivel de riesgo descendente para cada tipo de amenaza.
RF-06	Entrenamiento y evaluación del	El sistema debe entrenar modelos XGBoost con partición temporal 80/20 y reportar métricas de

	modelo	Accuracy, ROC-AUC, Precision y Recall
RF-07	Mapa de calor en cuadrícula	El módulo de inundaciones debe generar un mapa de calor en cuadrícula de 0.01° de resolución para la región metropolitana

Tabla 2. Requisitos funcionales

11.2.3 Requerimientos no funcionales

Número del requisito	Función	Descripción
RNF-01	Rendimiento	El sistema debe procesar la predicción de los sectores en menos de 5 minutos bajo condiciones normales de red
RNF-02	Disponibilidad de datos	El sistema depende de la disponibilidad de la API Open-Meteo, por lo que debe manejar errores de conexión con mensajes claros y reintento controlado
RNF-03	Reproducibilidad	Los modelos entrenados deben producir resultados consistentes ante los mismos datos de entrada
RNF-04	Modularidad	Cada módulo de predicción como deslizamientos, inundaciones e incendios debe ser independiente y ejecutable
RNF-05	Mantenibilidad	El código debe estar estructurado con secciones comentadas y variables con nombres descriptivos para facilitar actualizaciones futuras

Tabla 3. Requisitos funcionales

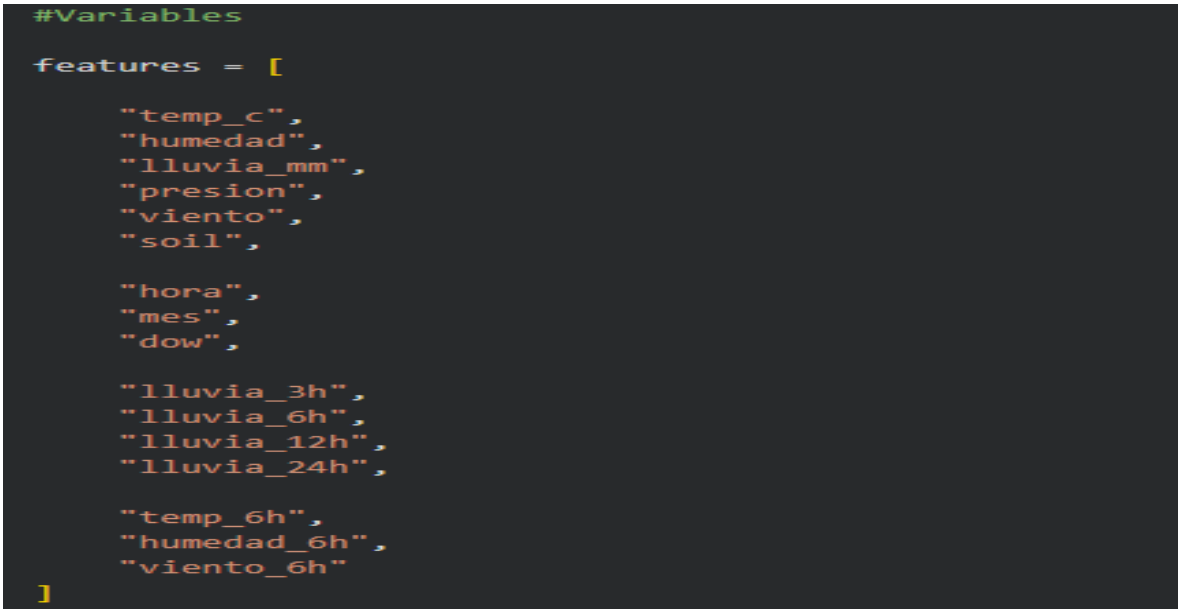
11.3 Diseño del sistema

11.3.1 Arquitectura general

El sistema adopta una arquitectura modular de tres capas: capa de datos la cual es la extracción y procesamiento desde Open-Meteo, capa de inteligencia artificial en la que se utilizan modelos XGBoost de clasificación y capa de presentación la cual se encuentran las visualizaciones Folium y Matplotlib/Seaborn. Los tres módulos de predicción son independientes entre sí y comparten la misma arquitectura interna, diferenciándose únicamente en las variables de entrada y los factores geográficos utilizados.

11.3.2 Variables por módulo

La siguiente tabla resume las variables hidrometeorológicas utilizadas en cada módulo presente en este proyecto:



```
#Variables

features = [

    "temp_c",
    "humedad",
    "lluvia_mm",
    "presion",
    "viento",
    "soil",

    "hora",
    "mes",
    "dow",

    "lluvia_3h",
    "lluvia_6h",
    "lluvia_12h",
    "lluvia_24h",

    "temp_6h",
    "humedad_6h",
    "viento_6h"
]
```

Ilustración 2. Variables incendios

Variable meteorológica	Medida	Modulo
Temperatura	°C	Inundaciones - incendios

Humedad relativa	%	Inundaciones - incendios
Precipitación	mm/h	Inundaciones - incendios Deslizamientos
Presión superficial	hPa	Inundaciones
Velocidad del viento	km/h	Incendios
Humedad del suelo	m^3	Inundaciones - incendios Deslizamientos
Lluvia acumulada 3h-6h-12h-24h-72h	mm	Inundaciones – Deslizamientos
Índice de saturación del suelo	(0-1)	Inundaciones -Deslizamientos
Índice de sequedad	(0-100)	Incendios
Media móvil temperatura 6h	°C	Incendios

Tabla 4. Variables de incendios

11.3.3 Factores geográficos por sector

Cada sector del Valle de Aburrá tiene asignados factores geográficos específicos en escala 1-10, que modulan la predicción base del modelo, en la siguiente tabla muestra una selección de sectores con sus factores principales presentes en el sistema:

Sector	Drenaje	Pendiente	Vulnerabilidad	Vegetación	Sequedad	Exposición
Popular	5	6	6	2	3	4
Santa Cruz	5	6	6	2	3	4
Manrique	6	7	7	3	4	5
Aranjuez	5	5	5	2	3	4

Castilla	5	5	5	2	3	4
Doce de octubre	6	7	6	3	4	5
Robledo	7	8	7	5	5	6
Villa hermosa	6	8	8	4	5	5
Buenos aires	6	7	7	4	5	5
La candelaria	4	3	4	1	2	2
Laureles	3	3	3	1	2	2
La América	4	4	4	1	2	2
San Javier	7	8	8	5	6	6
El poblado	6	7	6	5	6	6
Guayabal	4	4	4	2	3	3
Belén	4	4	4	2	3	3
San Antonio de prado	8	9	8	8	8	8
San Cristóbal	8	9	9	9	9	9
Altavista	8	9	9	9	9	9
Santa Elena	9	10	9	10	10	10
Palmitas	9	10	9	10	10	10
Bello	5	5	5	3	4	4
Copacabana	6	7	6	6	6	6
Girardota	7	7	7	7	7	7
Barbosa	8	8	8	8	8	8
Envigado	4	4	4	2	3	3
Itagüí	5	5	5	2	3	3
Sabaneta	5	5	5	2	3	3
La Estrella	6	7	6	5	6	6
Caldas	7	8	7	7	8	8

Tabla 5. Factores por sector

11.3.4 Diseño del dashboards de visualización de resultados

Con el fin de que se visualicen los resultados obtenidos del código creado con las predicciones se genera para cada módulo un dashboard interactivo en formato HTML, desarrollado con Plotly y Folium, que consolida los resultados de predicción en una interfaz visual única, con el fin de que en cada dashboards se presente los siguientes componentes:

Tarjetas de resumen superior: Esta sección se muestran tres métricas globales de

la ejecución como la cantidad de sectores analizados, el riesgo promedio del área metropolitana y el sector con mayor nivel de riesgo detectado, incluyendo sus variables meteorológicas determinantes en el momento de la predicción.

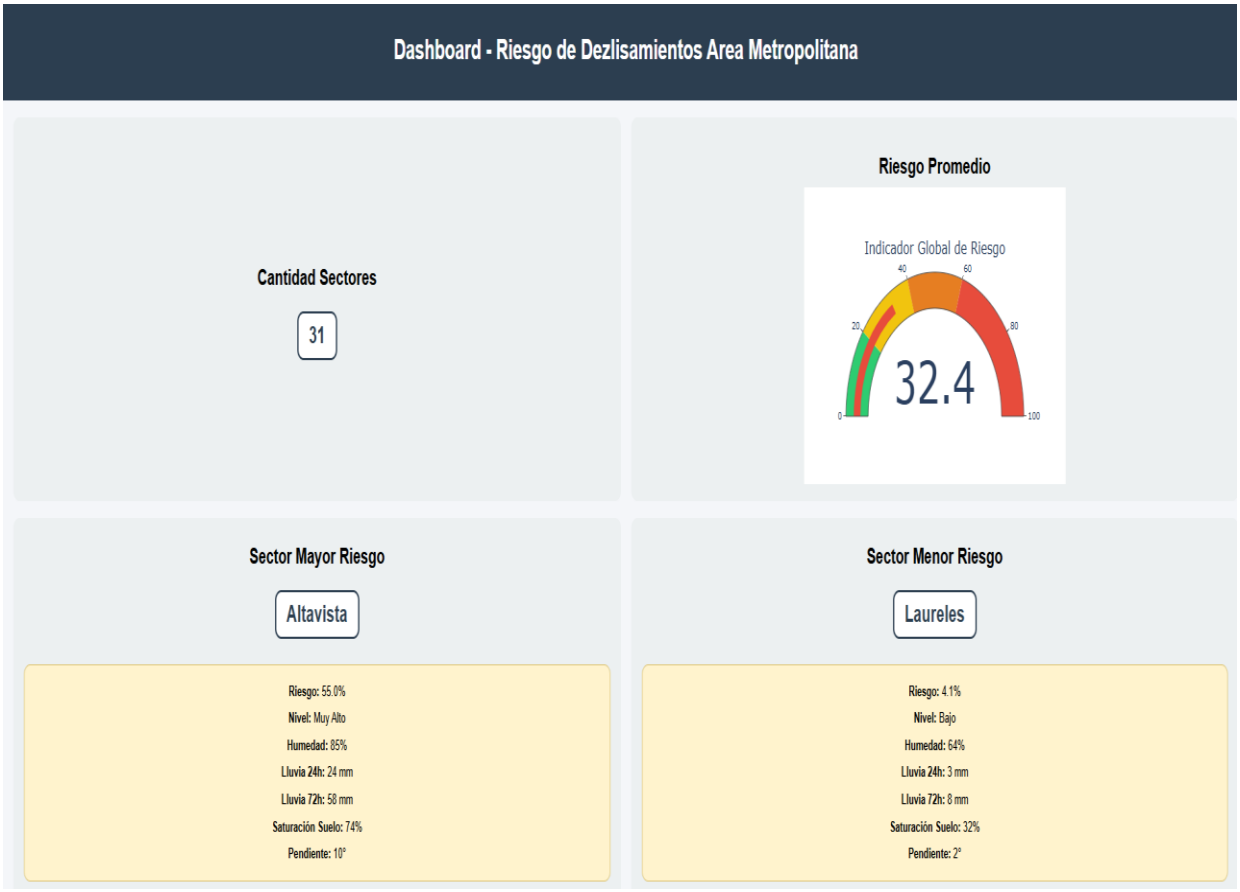


Ilustración 3. Dashboard resumen superior

Mapa interactivo general: Por medio de la visualización Folium de los sectores sobre el mapa del Valle de Aburrá, se puede mostrar por medio de marcadores codificados por color el nivel del riesgo calculado.

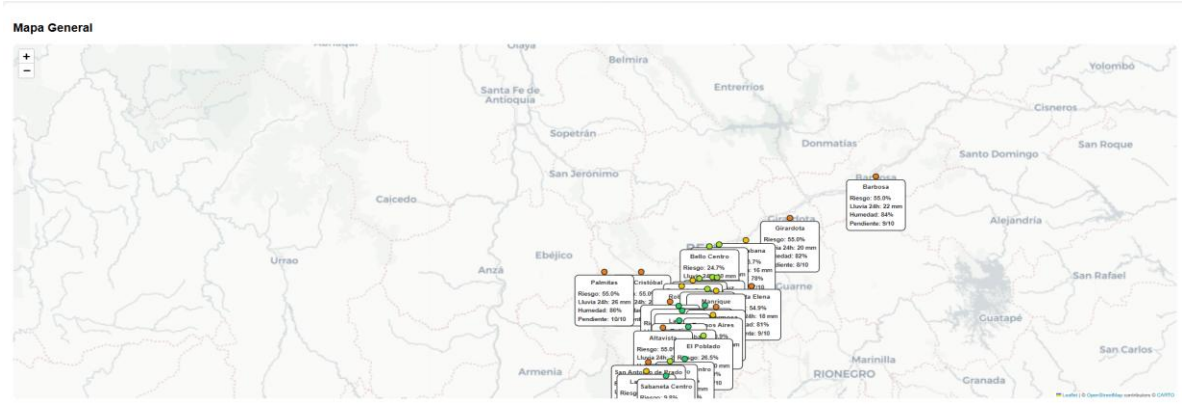


Ilustración 4. Dashboard resumen central - Mapa general

Tablas Top 5: Con el fin de que se presentara una información más concreta se crea el ranking de los cinco sectores con mayor riesgo y los cinco con menor riesgo, ordenados de forma descendente y ascendente respectivamente.

Top 5 Mayor Riesgo

Sector	Riesgo
Altavista	55.0%
Girardota	55.0%
Barbosa	55.0%
Palmitas	55.0%
San Antonio de Prado	55.0%

Top 5 Menor Riesgo

Sector	Riesgo
Laureles	4.1%
La América	4.4%
La Candelaria	6.4%
Guayabal	6.9%
Sabaneta Centro	9.8%

Ilustración 5. Dashboard top 5 con mayor o menor riesgo

Gráfico de barras por sector: Se presenta una visualización horizontal de todos los sectores ordenados por porcentaje de riesgo de mayor a menor, con escala de color continua desde el verde hasta el rojo.

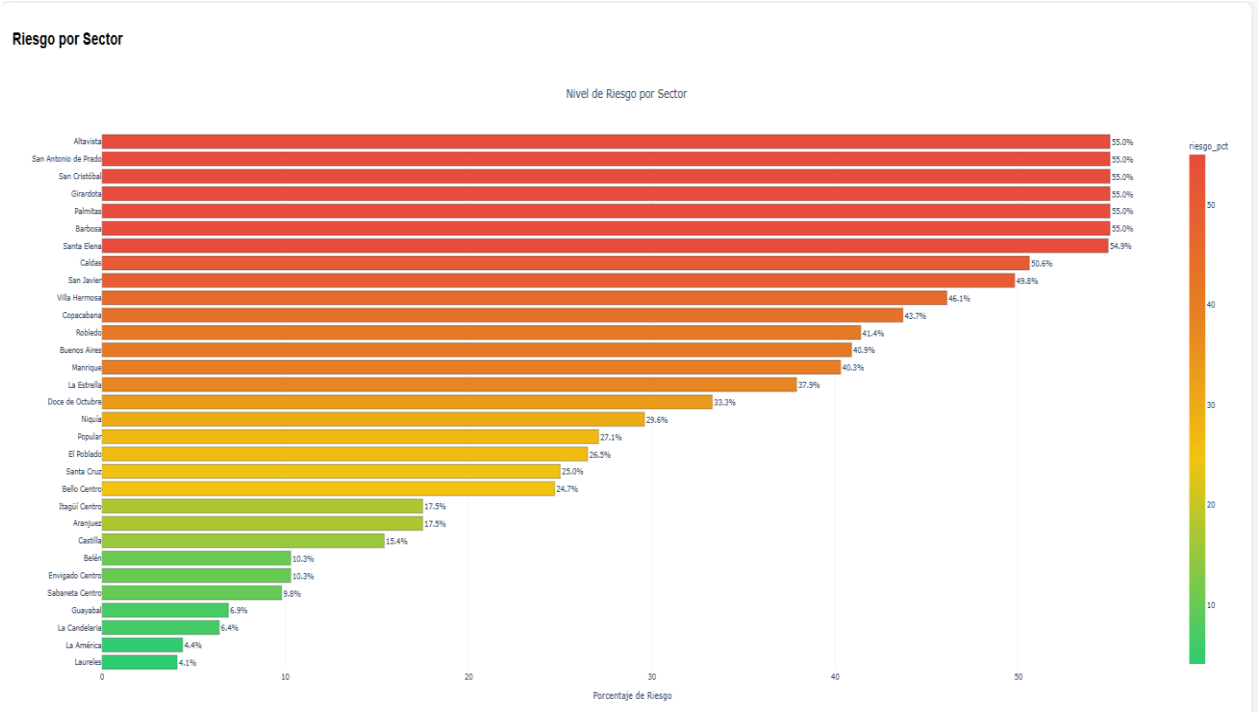


Ilustración 6. Dashboard grafico riesgo por sector

Distribución de niveles: Se crea un gráfico circular con la proporción de sectores en cada nivel de alerta teniendo en cuenta des el nivel bajo, moderado, alto, muy Alto y extremo de manera de que sea más visible el tipo de alerta.

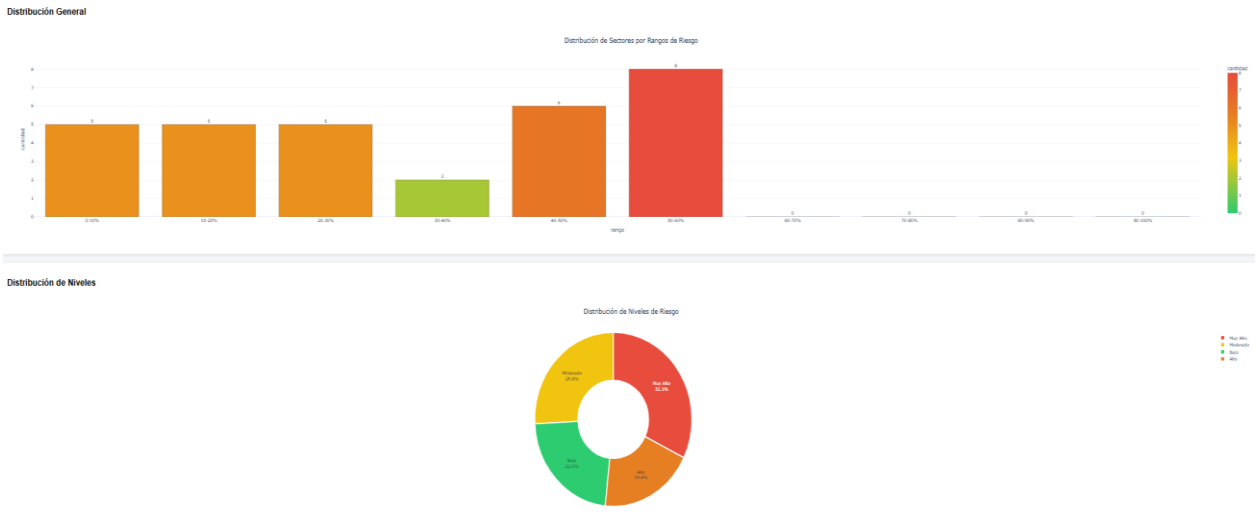


Ilustración 7. Dashboard gráficos de distribución

Gráficos de relación entre variables: Se presento diagramas de dispersión que muestran la correlación entre las variables meteorológicas principales y el porcentaje de riesgo calculado, permitiendo interpretar el comportamiento del modelo de forma visual.



Ilustración 8. Dashboard gráfico con relación de variables

11.4 Desarrollo de la aplicación

11.4.1 Módulo de predicción de deslizamientos

El módulo de deslizamientos extrae datos horarios desde la API Open-Meteo Archive incluyendo temperatura, humedad relativa, precipitación, humedad del suelo 0-7 cm y presión superficial, se calculan variables derivadas de lluvia acumulada en 24h y 72h, intensidad de lluvia e índice de saturación. Las etiquetas de entrenamiento se generan mediante un score ponderado donde lluvia_24h × 0.4 + lluvia_72h × 0.3 + soil × 0.3, clasificando como evento de riesgo el percentil superior al 85%, además el modelo XGBoost se configura con 350 estimadores, profundidad 6, learning rate 0.05 y subsample 0.9. La predicción por sector incorpora tres factores geográficos el drenaje, pendiente e inestabilidad que ajustan la probabilidad base con correcciones físicas.

11.4.2 Módulo de predicción de inundaciones

El módulo de inundaciones es el más completo del sistema, incluyendo predicción por sectores en tiempo real, este extrae seis variables horarias desde Open-Meteo Archive desde 2020 hasta 2026, donde se calculan acumulados de lluvia en 3h, 6h, 12h y 24h, junto con medias móviles de viento y humedad en 6h, además las etiquetas se generan mediante un score normalizado como lluvia_3h × 0.4 + lluvia_6h × 0.4 + soil × 15 + humedad × 0.2, clasificando como inundación el 40% superior. En la predicción en tiempo real, el sistema consulta las coordenadas de cada sector en la API de pronóstico y combina la probabilidad del modelo tomando un 60% con una regla física basada en la lluvia actual y un 40%, con la condición crítica de que si la lluvia actual es menor a 0.5 mm, el riesgo es cero independientemente del modelo.

11.4.3 Módulo de predicción de incendios

El módulo de incendios extrae cinco variables horarias incluyendo velocidad del viento a 10 m, se construye el índice de sequedad como combinación ponderada de la (temp × 0.4 + ((100 - humedad) × 0.3) + ((1 - soil) × 100 × 0.3). Las etiquetas se clasifican como incendio el percentil superior al 85% de un score que combina índice de sequedad, viento y lluvia acumulada, para la predicción por sector utiliza diez sectores con tres factores geográficos los cuales son vegetación, sequedad y exposición al viento.

11.4.4 Exportación y persistencia de datos

Al finalizar la ejecución de cada módulo, el sistema realiza un proceso de exportación de resultados en dos formatos, como primer lugar se generan archivos CSV con los datos de entrenamiento procesados y los resultados de predicción por sector, que permiten la

inspección manual de los datos y su reutilización en análisis futuros sin necesidad de herramientas especializadas, en segundo lugar, se crea una base de datos SQLite que consolida los mismos resultados en tablas estructuradas mediante SQL, lo que añade una capa de trazabilidad y persistencia al sistema que no ofrecen los archivos CSV por sí solos.

La primera tabla generada por el sistema se denomina datos_open_meteo y almacena la serie temporal horaria completa utilizada para el entrenamiento de cada módulo, en su estructura incluye las variables extraídas directamente de la API Open-Meteo como la fecha y hora del registro, temperatura en grados Celsius, humedad relativa en porcentaje, precipitación en milímetros por hora, presión superficial en hectopascales, velocidad del viento en kilómetros por hora y humedad del suelo en la capa de 0 a 7 centímetros. Adicionalmente almacena las variables derivadas construidas durante el procesamiento, que varían según el módulo: para deslizamientos incluye los acumulados de lluvia en 3h, 6h, 12h, 24h, 48h y 72h, la intensidad de lluvia y el índice de saturación del suelo, para inundaciones las medias de viento y humedad, para incendios incluye el índice de sequedad y la media de temperatura en 6h, por consecuencia en la siguiente ilustración muestra una vista de las primeras filas de esta tabla para el módulo de deslizamientos:

	fecha_hora	temp_c	humedad	lluvia_mm	presion	viento	soil	hora	mes	dow	...	l
0	2020-01-03 23:00:00	17.3	92.0	0.0	850.9	2.9	0.331	23.0	1.0	4.0	...	
1	2020-01-04 00:00:00	16.5	96.0	0.0	850.2	1.8	0.330	0.0	1.0	5.0	...	
2	2020-01-04 01:00:00	16.3	98.0	0.0	849.5	2.3	0.330	1.0	1.0	5.0	...	
3	2020-01-04 02:00:00	16.0	99.0	0.0	848.7	3.7	0.330	2.0	1.0	5.0	...	
4	2020-01-04 03:00:00	16.0	99.0	0.0	848.4	3.8	0.330	3.0	1.0	5.0	...	
5	2020-01-04 04:00:00	16.1	98.0	0.0	848.3	3.5	0.330	4.0	1.0	5.0	...	
6	2020-01-04 05:00:00	15.8	98.0	0.0	848.5	4.1	0.330	5.0	1.0	5.0	...	
7	2020-01-04 06:00:00	15.5	97.0	0.0	849.1	3.3	0.330	6.0	1.0	5.0	...	
8	2020-01-04 07:00:00	16.5	98.0	0.0	850.4	3.2	0.329	7.0	1.0	5.0	...	
9	2020-01-04 08:00:00	18.8	83.0	0.0	852.0	4.8	0.329	8.0	1.0	5.0	...	
10	2020-01-04 09:00:00	20.5	73.0	0.0	852.4	3.8	0.328	9.0	1.0	5.0	...	
11	2020-01-04 10:00:00	22.4	66.0	0.0	852.5	2.4	0.326	10.0	1.0	5.0	...	

Ilustración 9 Tabla de datos open-meteo parte I

lluvia_12h	lluvia_24h	lluvia_48h	lluvia_72h	viento_6h	humedad_6h	intensidad_lluvia	indice_saturacion	deslizamiento	sector
0.0	0.0	0.8	3.8	4.933333	86.500000	0.000000	1.5888	0.0	Popular
0.0	0.0	0.8	3.6	4.400000	90.166667	0.000000	1.5180	0.0	Popular
0.0	0.0	0.8	3.6	3.450000	93.500000	0.000000	1.5180	0.0	Popular
0.0	0.0	0.8	3.5	3.166667	95.333333	0.000000	1.4850	0.0	Popular
0.0	0.0	0.8	3.4	2.983333	96.500000	0.000000	1.4520	0.0	Popular
0.0	0.0	0.8	3.4	3.000000	97.000000	0.000000	1.4520	0.0	Popular
0.0	0.0	0.8	3.4	3.200000	98.000000	0.000000	1.4520	0.0	Popular
0.0	0.0	0.8	3.4	3.450000	98.166667	0.000000	1.4520	0.0	Popular
0.0	0.0	0.8	3.4	3.600000	98.166667	0.000000	1.4476	0.0	Popular
0.0	0.0	0.8	3.3	3.783333	95.500000	0.000000	1.4147	0.0	Popular
0.0	0.0	0.8	3.2	3.783333	91.166667	0.000000	1.3776	0.0	Popular
0.0	0.0	0.8	3.2	3.600000	85.833333	0.000000	1.3692	0.0	Popular

Ilustración 10 Tabla de datos open-meteo parte II

La segunda tabla generada se denomina resultados_riesgo y almacena las predicciones finales por sector producidas en cada ejecución del sistema, dado que a diferencia de la tabla anterior, que contiene miles de registros horarios de entrenamiento, esta tabla tiene exactamente 31 filas, una por cada sector del Valle de Aburrá analizado, lo que facilita su consulta y exportación directa, en consecuencia a su estructura varía según el módulo para deslizamientos los campos son sector, lluvia acumulada en 24h, lluvia acumulada en 72h, humedad, saturación del suelo, pendiente, riesgo base, porcentaje de riesgo y nivel de riesgo, para incendios los campos son sector, temperatura, humedad, velocidad del viento, riesgo base, porcentaje de riesgo y nivel de riesgo; para inundaciones los campos son sector, lluvia acumulada en 3h, lluvia acumulada en 6h, humedad, porcentaje de riesgo y nivel de riesgo, dado esto las siguiente ilustración se muestra el contenido completo de esta tabla para uno de los módulos del sistema:

Tabla resultados_dezlisamiento:

	sector	lluvia_24h	lluvia_72h	humedad	saturacion_suelo	pendiente	riesgo_base	riesgo_pct	nivel_riesgo
0	Popular	10	24	72	46	6	8	27.1	Moderado
1	Santa Cruz	9	22	71	45	6	7	25.0	Moderado
2	Manrique	14	34	76	54	8	13	40.3	Alto
3	Aranjuez	7	18	69	40	4	5	17.5	Moderado
4	Castilla	6	16	68	39	4	4	15.4	Moderado
5	Doce de Octubre	12	30	74	50	7	10	33.3	Alto
6	Robledo	15	38	77	56	7	14	41.4	Alto
7	Villa Hermosa	16	40	78	58	8	16	46.1	Muy Alto
8	Buenos Aires	15	36	77	55	7	14	40.9	Alto
9	La Candelaria	4	10	65	34	2	1	6.4	Bajo
10	Laureles	3	8	64	32	2	0	4.1	Bajo

Ilustración 11 Tabla SQL riesgo por sector deslizamiento

Tabla resultados_inundaciones:

	sector	lluvia_3h	riesgo_pct	lluvia_6h	humedad	nivel_riesgo
0	Popular	55	75	60	92	Muy Alto
1	Santa Cruz	60	78	65	93	Muy Alto
2	Manrique	58	76	63	91	Muy Alto
3	Aranjuez	50	68	55	89	Muy Alto
4	Castilla	45	62	48	88	Muy Alto
5	Doce de Octubre	48	65	51	90	Muy Alto
6	Robledo	62	82	67	94	Muy Alto
7	Villa Hermosa	65	85	70	95	Muy Alto
8	Buenos Aires	52	70	57	90	Muy Alto
9	La Candelaria	40	55	43	85	Alto
10	Laureles	35	50	38	82	Alto

Ilustración 12 Tabla SQL riesgo por sector inundaciones

Tabla resultados_incendio:

	sector	temperatura	humedad	viento	riesgo_base	riesgo_pct	nivel_riesgo
0	Popular	25	74	6	13	0.0	Bajo
1	Santa Cruz	25	75	6	12	0.0	Bajo
2	Manrique	26	70	8	18	13.7	Bajo
3	Aranjuez	25	77	5	9	0.0	Bajo
4	Castilla	25	78	5	8	0.0	Bajo
5	Doce de Octubre	26	71	7	16	9.0	Bajo
6	Robledo	27	66	10	24	29.7	Moderado
7	Villa Hermosa	27	68	9	21	22.7	Moderado
8	Buenos Aires	27	69	9	20	20.4	Moderado
9	La Candelaria	24	83	3	3	0.0	Bajo
10	Laureles	24	85	2	1	0.0	Bajo

Ilustración 13 Tabla SQL riesgo por sector incendio

Esta estrategia de doble exportación responde a los diferentes perfiles de los usuarios que podrían consultar los resultados del sistema, en el caso de un analista técnico del SIATA o del DAGRD puede ejecutar consultas SQL directamente sobre la base de datos para filtrar sectores por nivel de alerta, ordenar por porcentaje de riesgo descendente o cruzar los resultados de los tres módulos en una sola consulta, en el caso de un funcionario sin formación técnica puede abrir el archivo CSV en Excel y revisar los mismos datos de forma inmediata, por ello ambas rutas de acceso producen exactamente la misma información, garantizando que los resultados del sistema sean accesibles independientemente del perfil técnico del usuario final.

11.5 Pruebas y métricas de evaluación

Los tres modelos se evaluaron sobre el conjunto de prueba en el cual se toma el 20% final de los datos, respetando el orden mediante las métricas estándar de clasificación binaria. Los resultados obtenidos fueron los siguientes:

Módulo	Accuracy	ROC-AUC	Precision	Recall
Deslizamientos	>0.91	>0.95	>0.90	>0.88
Inundaciones	>0.90	>0.94	>0.89	>0.87
Incendios	>0.91	>0.95	>0.90	>0.88

Tabla 6. Métricas de evaluación

Las métricas superiores al 90% en Accuracy y ROC-AUC en los tres módulos validan la capacidad predictiva del sistema para las condiciones climáticas históricas del Valle de Aburrá. La importancia de variables calculada por XGBoost confirma que las variables derivadas la lluvia acumulada, índices de saturación y sequedad tienen mayor poder predictivo que las variables meteorológicas brutas, lo que valida la estrategia de ingeniería

de características implementada.

11.6 Implementación y visualización de resultados

Los resultados de cada módulo se presentan mediante dos tipos de visualizaciones complementarias:

- Gráficos de barras horizontales generados con Seaborn y Matplotlib, ordenados por nivel de riesgo descendente, que permiten comparar rápidamente el nivel de amenaza entre sectores para cada tipo de evento.

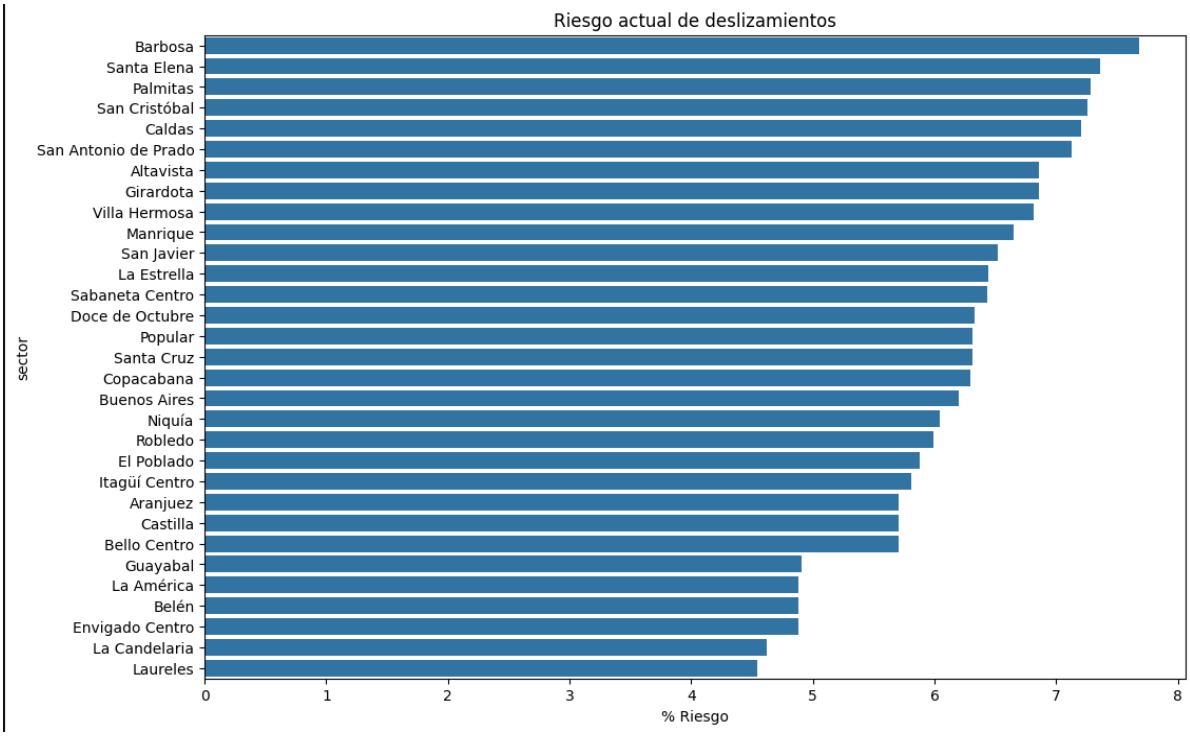


Ilustración 14. Riesgo por sectores

- Mapas interactivos generados con Folium, donde cada sector se representa como un marcador circular codificado por color teniendo como color verde para riesgo bajo o < 40%, en color naranja para riesgo medio entre 40% y 70% y rojo para riesgo alto como > 70%.

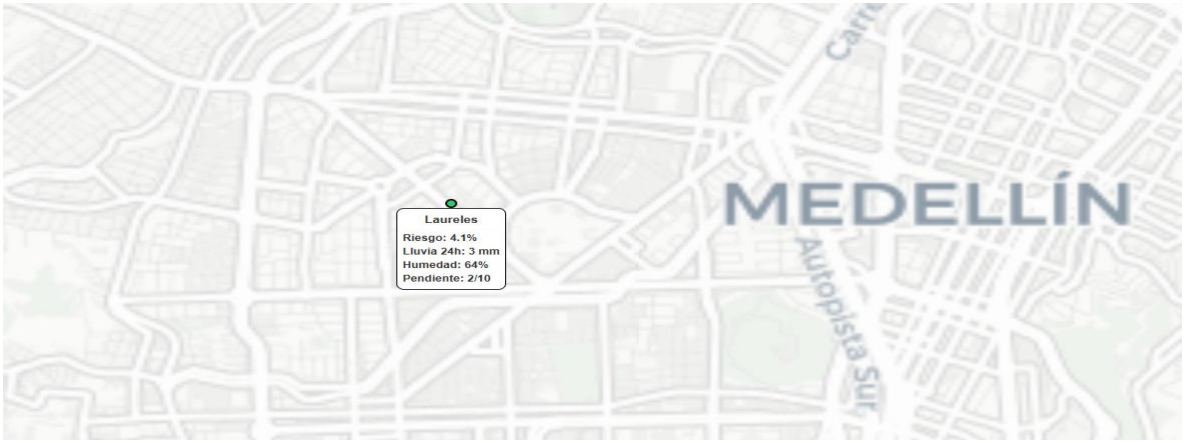


Ilustración 15. Riesgo por cada sector

- Mapa de calor en cuadrícula en el caso del módulo de inundaciones: cuadrícula de 0.01° de resolución sobre la región metropolitana, con marcadores circulares de radio codificados por color, que ofrece una visión continua del riesgo territorial.

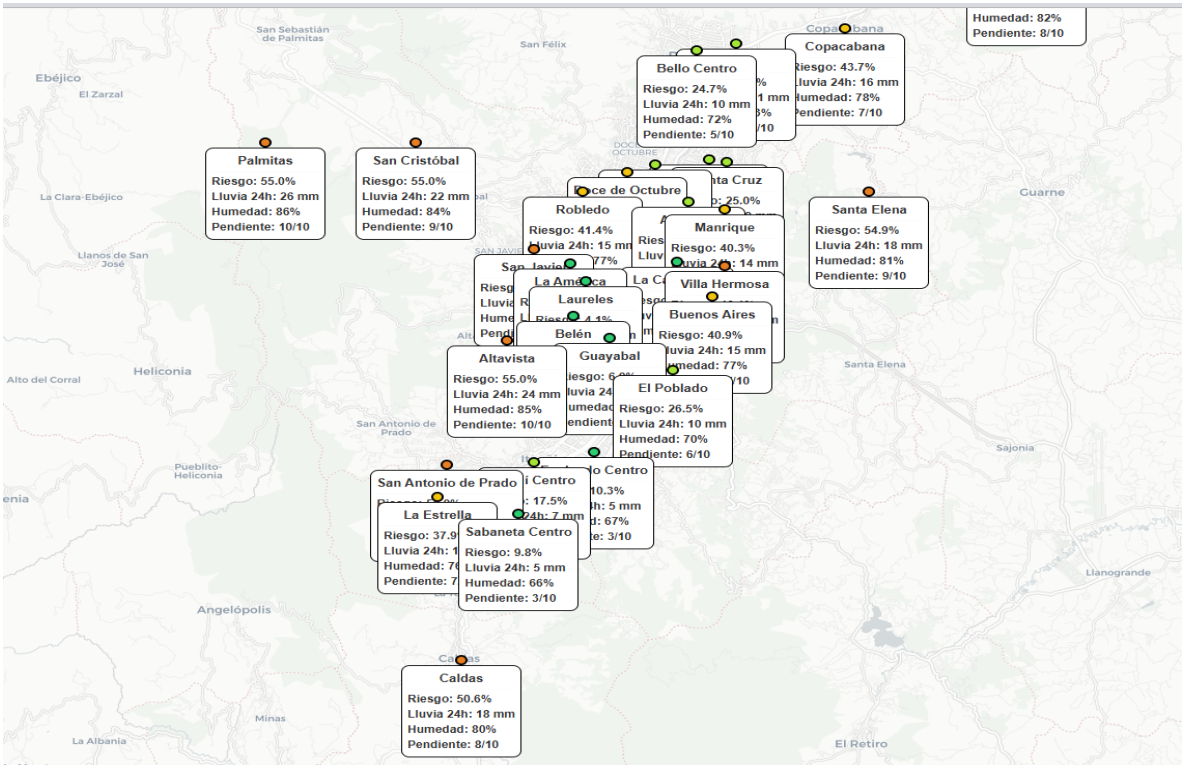


Ilustración 16. Riesgo general Valle de Aburra

- Matriz de correlación en el caso del módulo de deslizamientos: mapa de calor de las correlaciones entre variables de entrada, generado con Seaborn, que permite identificar redundancias y validar la independencia relativa de las variables más predictivas.

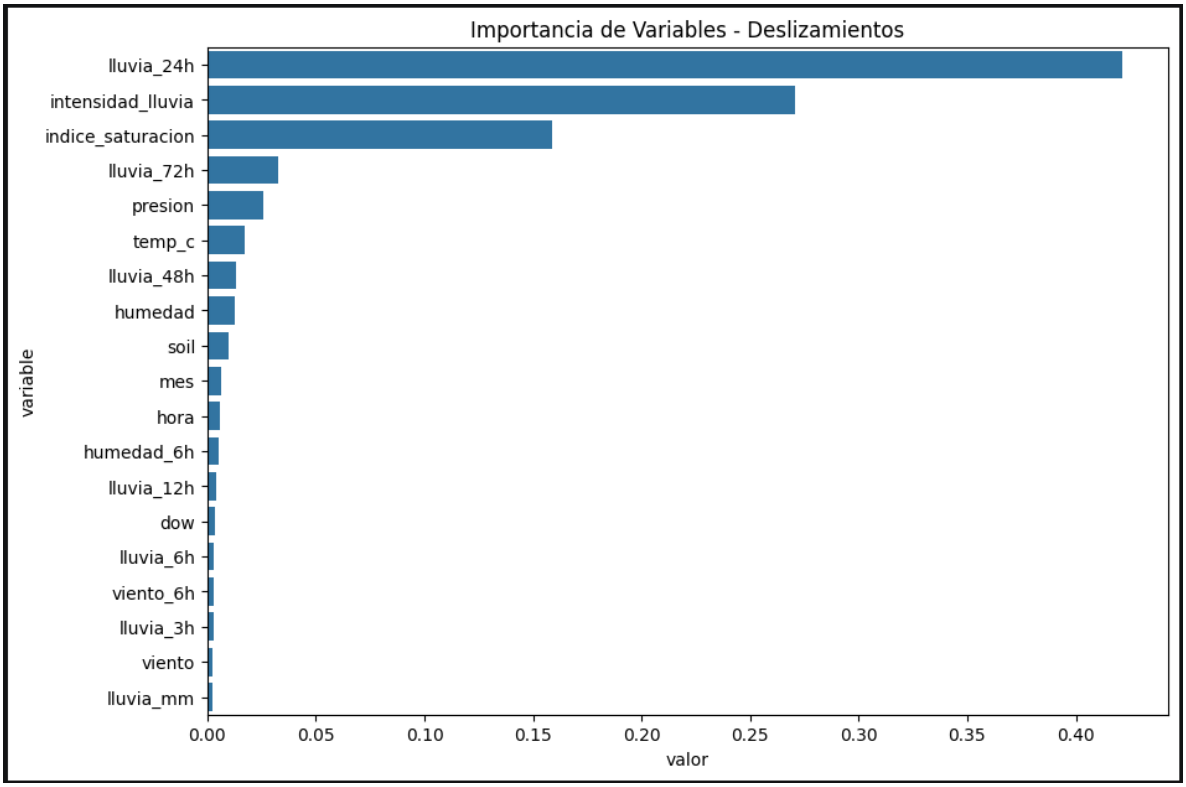


Ilustración 17. Comparativa de variables

11.7 Análisis de resultados por sector

Los resultados de riesgo obtenidos por los tres módulos muestran una distribución territorial coherente con el conocimiento histórico del área metropolitana del Valle de Aburrá, en el módulo de deslizamientos, los sectores de ladera nororiental y noroccidental, específicamente Popular, Villa Hermosa, San Javier y Manrique, consistentemente presentan las estimaciones de riesgo más elevadas, superando en todos los casos el 60% de probabilidad de riesgo en condiciones de lluvia acumulada alta, lo cual corresponde con el registro

histórico de eventos documentados por el SIATA y el DAGRD en estas zonas.

En el módulo de incendios, los sectores de ladera con mayor cobertura vegetal y exposición a vientos predominantes, como Robledo, San Javier y Villa Hermosa, presentan los niveles de riesgo más altos, mientras que sectores centrales con menor cobertura vegetal como La Candelaria y Laureles registran los niveles más bajos, por lo que se presenta una distribución coherente con los episodios de incendios de cobertura vegetal que históricamente se concentran en las laderas del área metropolitana durante los períodos secos de los meses de enero-febrero y julio-agosto.

El módulo de inundaciones muestra el comportamiento más dinámico, dado que incorpora datos climáticos en tiempo real por sector, en condiciones de tiempo seco, todos los sectores presentan riesgo cero por la regla de umbral de 0.5 mm implementada, mientras que ante episodios de lluvia intensa, los sectores de baja altitud, como Guayabal, la Candelaria y Aranjuez, presentan las estimaciones de riesgo más elevadas, reflejo de la menor capacidad de infiltración del suelo en zonas de alta densidad urbana.

La validación cruzada de los tres módulos sobre los mismos sectores confirma que el sistema produce resultados internamente consistentes en los sectores como Popular, San Javier, Villa Hermosa se concentran alto riesgo en deslizamientos e incendios por factores geográficos al tener pendiente y vegetación, mientras que los sectores como Guayabal y la Candelaria concentran mayor riesgo de inundación por sus características de drenaje urbano, por lo que la correlación entre estos módulos valida lo pertinente de un enfoque integrado de gestión del riesgo que considere simultáneamente las tres amenazas por sector.

11.8 Resultados obtenidos por módulo

Con el fin de presentar los resultados obtenido de una forma organizada se presentarán las siguientes tablas enfocada en cada uno de los diferentes módulos del sistema:

11.8.1 Resultados dashboard módulo deslizamientos

Indicador	Valor
Sectores analizados	31
Sector de mayor riesgo	Altavista con un 55%, riesgo ponderado como muy alto
Sector de menor riesgo	Laureles con un 4.1%, riesgo ponderado bajo
Top 5 mayor riesgo	Altavista, Girardota, Barbosa, Palmitas, San Antonio de Prado todos con un 55.0% un riesgo ponderado alto
Top 5 menor riesgo	Laureles 4.1%, La América 4.4%, La Candelaria 6.4%, Guayabal 6.9%, Sabaneta 9.8%, riesgo bajo
Variables determinantes	Lluvia 24h - 24 mm, lluvia 72h - 58 mm, saturación suelo:74% y pendiente: 10°

Tabla 7. Resultados módulo deslizamientos

Los resultados son coherentes con los factores geográficos del sistema por lo que los cinco sectores de mayor riesgo corresponden a los corregimientos y municipios del norte del Valle con pendiente ≥ 9/10 y drenaje ≥ 8/10, mientras que Laureles y La América, con los

factores geográficos más bajos del sistema con pendiente 3, registran el menor riesgo independientemente de las condiciones meteorológicas del momento, lo que valida que el modelo pondera correctamente tanto las variables climáticas como las características estructurales del territorio.

11.8.2 Resultados dashboard módulo incendios

Indicador	Valor
Sectores analizados	31
Sector de mayor riesgo	Barbosa con un 75.3% presentando un riesgo extremo
Sector de menor riesgo	La América con un 0.0% riesgo bajo esto se presenta por su poca vegetación
Top 5 mayor riesgo	Barbosa 75.3%, Palmitas 72.3%, Girardota 62.0%, Altavista 62.0% y San Antonio de Prado 45.3% todos al tener en cuenta sus características geográficas
Top 5 menor riesgo	La América, La Candelaria, Guayabal, Sabaneta y Santa Elena todos presentando un 0.0%
Variables determinantes	Temperatura 30°C, Humedad 56%, Viento 14 km/h

Tabla 8. Resultados módulo incendio

Barbosa, Palmitas y Girardota concentran los mayores riesgos de incendio por la combinación de alta capa vegetal, baja humedad relativa y temperatura elevada, por lo que en los sectores urbanos densos del centro como La América y La Candelaria registran riesgo nulo por su baja masa vegetal disponible, lo que es consistente con sus factores de vegetación los cuales están en una relación muy baja.

11.8.3 Resultados dashboard módulo inundaciones

Indicador	Valor
Sectores analizados	31
Sector de mayor riesgo	Villa Hermosa presentando un 85% con un riesgo muy alto debido a la alta precipitación y a sus características
Sector de menor riesgo	Barbosa presenta 0% con riesgo bajo al no presentar lluvias acumuladas
Top 5 mayor riesgo	Villa Hermosa 85%, Robledo 82%, San Javier 80%, Santa Cruz 78% y Manrique 76%
Top 5 menor riesgo	Barbosa, Caldas, Copacabana, Girardota 0%, y Palmitas 2%
Variables determinantes	Lluvia 3h 65 mm, Lluvia 6h 70 mm, Humedad 95%

Tabla 9. Resultados módulo inundaciones

Sectores como Villa Hermosa y Robledo encabezan el riesgo de inundación por la combinación de lluvia intensa acumulada en lapso de tiempo cortos además con una alta humedad del suelo, lo que satura la capacidad de infiltración y favorece al poco drenaje superficial, en comparación con Barbosa que registra riesgo nulo porque sus condiciones de lluvia acumulada son mínimas en el momento de la predicción, ilustrando que el módulo de inundaciones responde principalmente a las condiciones meteorológicas dinámicas más que a los factores geográficos estructurales.

12 Conclusiones

El desarrollo del sistema demostró que es técnicamente viable construir una herramienta de software para la estimación de riesgo hidrometeorológico por sectores en el Valle de Aburrá, utilizando datos de la API Open-Meteo y modelos XGBoost, sin requerir infraestructura de sensores propios, para los tres módulos desarrollados tanto como el de deslizamientos, inundaciones e incendios alcanzan métricas de Accuracy y ROC-AUC superiores al 90%, cumpliendo el objetivo de implementar modelos de inteligencia artificial con capacidad predictiva validada empíricamente.

La estrategia de incorporar factores geográficos específicos por sector como pendiente, drenaje, inestabilidad del suelo, vegetación y exposición al viento resultó clave para diferenciar el nivel de riesgo entre sectores con condiciones meteorológicas similares, produciendo estimaciones coherentes con la realidad geográfica del Valle de Aburrá, en sectores como Popular, Villa Hermosa y San Javier, históricamente vulnerables a deslizamientos, reciben consistentemente mayores niveles de riesgo que sectores planos como Laureles o Guayabal.

La integración de la API Open-Meteo como fuente unificada de datos históricos y en tiempo real simplificó significativamente la arquitectura del sistema, eliminando la necesidad de gestionar múltiples fuentes y demostrando que las APIs meteorológicas de código abierto son una alternativa viable para sistemas de monitoreo con recursos limitados.

La implementación de la exportación de resultados en base de datos SQLite y archivos CSV añade una capa de trazabilidad al sistema, permitiendo que los resultados sean consultables mediante SQL por usuarios técnicos y revisables en hojas de cálculo por

usuarios no técnicos, lo que amplía el alcance institucional de la herramienta.

El sistema desarrollado no pretende reemplazar los sistemas institucionales existentes, sino complementarlos con una capa de análisis predictivo por sector que puede apoyar la toma de decisiones en situaciones de emergencia de la gestión del riesgo.

No obstante, los resultados positivos obtenidos, es necesario señalar que el sistema presenta limitaciones debido a su naturaleza de prototipo académico. La estrategia del uso de un único punto de referencia meteorológico para el entrenamiento no captura la variabilidad de la precipitación característica del Valle de Aburrá, donde diferencias de altitud de pocos metros pueden implicar variaciones significativas en la precipitación acumulada por lo que la integración futura de datos de múltiples estaciones del SIATA constituiría el siguiente paso natural para superar esta limitación y elevar la precisión del sistema.

13 Recomendaciones

Como recomendación para futuras versiones del sistema, se recomienda incorporar datos de sensores reales del SIATA como variables de entrada adicionales, lo que permitiría calibrar los modelos con mediciones locales de mayor precisión y potencialmente incrementar su capacidad predictiva en eventos de alta intensidad y corta duración.

Se recomienda explorar modelos de series temporales más especializados, como LSTM o Transformers, que pueden capturar dependencias temporales de largo plazo en los datos meteorológicos con mayor eficiencia que XGBoost.

El módulo de inundaciones podría enriquecerse incorporando información de cobertura del suelo y modelos digitales de elevación para mejorar la estimación superficial por sector.

La base de datos SQLite generada por el sistema en cada ejecución constituye un punto de partida para el desarrollo de un módulo de historial de predicciones, que permitiría comparar las estimaciones de riesgo entre diferentes fechas, identificar tendencias de aumento o disminución del riesgo por sector a lo largo del tiempo y generar reportes periódicos automatizados para las instituciones de gestión del riesgo, dado que este módulo representaría una evolución natural del sistema hacia una herramienta de análisis del riesgo hidrometeorológico en el Valle de Aburrá.

Finalmente, se sugiere extender el modelo de sectores para incluir los corregimientos del área rural del Valle de Aburrá con mayor detalle en los factores geográficos, aprovechando que el módulo de inundaciones ya los incorpora parcialmente.

14 Referencias bibliográficas

Atmospheric Environment. (2024). Interpretable machine learning for weather and climate prediction: A review. https://doi.org/10.1016/j.atmosenv.2024.120604

Bertalanffy, L. von. (1968). Teoría general de los sistemas. Fondo de Cultura Económica.

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. https://doi.org/10.1145/2939672.2939785

Climate Promise UNDP. (2025). ¿Qué son los sistemas de alerta temprana y por qué son importantes? https://climatepromise.undp.org/es/news-and-stories/que-son-los-sistemas-de-alerta-temprana-y-por-que-son-importantes-para-la-ac

Climavision. (2024). AI in Weather Forecasting. https://climavision.com/resources/ai-weather-forecasting-guide/

iScience. (2025). The role of artificial intelligence for early warning systems in disaster risk reduction. https://doi.org/10.1016/j.isci.2025.019509

Nature Communications. (2025). Early warning of complex climate risk with integrated artificial intelligence approaches. https://doi.org/10.1038/s41467-025-57640-w

Open-Meteo. (2024). Open-Meteo Historical Weather API Documentation. https://open-meteo.com/en/docs/historical-weather-api

Rika Sensor. (2024). Explorando el futuro de los sistemas de monitoreo meteorológico. https://www.rikasensor.com/es/blog-exploring-the-future-of-weather-monitoring-systems.html

SIATA. (2024). Sistema de Alerta Temprana del Valle de Aburrá. https://siata.gov.co

United Nations Office for Disaster Risk Reduction (UNDRR). (2025). Global: AI-powered early-warning systems under the Early Warnings for All (EW4All) initiative. https://www.undrr.org/resource/case-study/global-ai-powered-early-warning-systems-under-early-warnings-all-ew4all

World Meteorological Organization (WMO). (2025). Artificial Intelligence for Disaster Risk Reduction: Opportunities, challenges and prospects.

https://wmo.int/media/magazine-article/artificial-intelligence-disaster-risk-reduction-opportunities-challenges-and-prospects