

Modelo predictivo para la acción anticipatoria y alerta temprana en eventos climáticos asociados a inundaciones (MONITOR)

Jeremías Pabón López¹, Jesús Alfredo Sanabria Mejía²

¹⁻²Facultad de Ingeniería y Ciencias Básicas (FICB), ³

Universidad Central de Colombia

Maestría en Analítica de Datos

Automatización e Integración de Datos IA

Bogotá, Colombia

¹jpabonl@ucentral.edu.co, ²jsanabriam4@ucentral.edu.co

May 10, 2024

Contents

1	Glosario	3
2	Introducción (Max 250 Palabras)	3
3	Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA (Max 500 Palabras)	4
3.1	Titulo del proyecto de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	4
3.2	Objetivo general (Max 100 Palabras))	4
3.2.1	Objetivos específicos (Max 100 Palabras)	4
3.3	Alcance (Max 200 Palabras)	4
3.4	Pregunta de investigación (Max 100 Palabras)	5
3.5	Hipotesis (Max 100 Palabras)	5
4	Reflexiones sobre el origen de datos e información	6
4.1	Origen de los datos	6
4.2	Consideraciones legales o éticas del uso de la información	6
4.3	Retos de la información y los datos en Integración y Automatización de Datos para IA	6
4.4	Resultado esperado de la Integración y Automatización de Datos para IA	7

5	Diseño de integración y Automatización de Datos para IA (Diagrama)	8
6	integración de Datos	10
7	Automatización de Datos	10
8	Inteligencia Artificial	11
9	Proximos pasos (<i>Tercera entrega</i>)	13
10	Lecciones aprendidas (<i>Tercera entrega</i>)	14
11	Bibliografía	15

1 Glosario

IDEAM: Instituto de Hidrología, Meteorología y Estudios Ambientales

DIMAR: Dirección General Marítima

OSSO: Corporación OSSO es una organización sin ánimo de lucro y de beneficio público dedicada a investigar actividades afines a la dinámica de la tierra, la hidrósfera y la atmósfera.

SGC: Servicio Geográfico Colombiano

CAR: Corporación Autónoma Regional

IDIGER: Instituto Distrital de Gestión de Riesgos y Cambio Climático

SIATA: El Sistema de Alerta Temprana de Medellín y el Valle de Aburrá.

2 Introducción (Max 250 Palabras)

Colombia cuenta con diversos sistemas de alerta temprana para fenómenos de gran escala, gestionados por entidades nacionales como el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), la Dirección General Marítima (DIMAR), Corporación OSSO y el Servicio Geográfico Colombiano (SGC). A nivel regional, las Corporaciones Ambientales, como la CAR, también han implementado sistemas, al igual que algunos municipios, como el Área Metropolitana del Valle de Aburrá, o a nivel distrital las experiencias relacionadas con IDIGER en Bogotá; o el El Sistema de Alerta Temprana de Medellín y el Valle de Aburrá (SIATA)

Los sistemas existentes no están diseñados para generar alertas a nivel local que le permita a las comunidades prepararse de manera anticipada frente a los efectos del fenómeno de la niñez, otros eventos climáticos repentinos o alteraciones en los sistemas hídricos como taponamiento de ríos. Los datos y pronósticos existentes pueden estar dispersos en varias fuentes, lo que dificulta que las comunidades y autoridades accedan a información completa y actualizada.

En la actualidad, las comunidades no son alertadas oportunamente sobre eventos de inundaciones, estos eventos ocurren e impactan la vida de las comunidades desde el punto de vista de la salud; causando infecciones y otras enfermedades relacionadas con picaduras de mosquitos (dengue); en la educación, las inundaciones interrumpen el aprendizaje, dificultando la permanencia en la escuela; también muchas familias pierden sus pertenencias, reciben daños en sus viviendas e incluso sus medios de vida se ven severamente afectados, lo que impacta directamente en la economía de los hogares y en su seguridad alimentaria. Como consecuencia de lo anterior, las familias se ven afectadas emocionalmente, lo que requiere una atención adecuada para que las familias puedan aplicar estrategias de afrontamiento positivas, contribuir a la resiliencia y mitigar los riesgos de protección para los niños.

La propuesta de este proyecto busca incorporar fuentes de información oficiales disponibles e información generada por las comunidades para ofrecer una solución que permita generar información y alertas oportunamente a las comunidades, autoridades locales y nacionales, organizaciones de la sociedad civil y organismo de socorro en la preparación anticipada y respuesta efectiva sobre los eventos asociados a las inundaciones.

3 Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA (Max 500 Palabras)

3.1 Título del proyecto de investigación (Max 100 Palabras) - *(Primera entrega)*

Modelo predictivo para la acción anticipatoria y alerta temprana en eventos climáticos asociados a inundaciones (MONITOR)

3.2 Objetivo general (Max 100 Palabras))

Desarrollar y pilotar un modelo predictivo basado en redes neuronales para la acción anticipatoria y alerta temprana en eventos climáticos asociados a inundaciones en Bogotá.

3.2.1 Objetivos específicos (Max 100 Palabras)

- Identificar e integrar datos de fuentes oficiales y comunitarias pertinentes para la predicción de inundaciones.
- Desarrollar un modelo basado en redes neuronales convolucionales (CNN) y redes neuronales recurrentes (RNN) para predecir las inundaciones en la zona focalizada.
- Diseñar una visualización de monitoreo sobre los eventos climáticos y no climáticos para informar a los actores involucrados en la zona focalizada.

3.3 Alcance (Max 200 Palabras)

Se propone desarrollar un sistema de alerta temprana que se basará en la recopilación de datos provenientes de plataformas institucionales de monitoreo del clima y de información comunitaria accesible. Utilizando esta información, se desarrollará un modelo predictivo que identificará variables clave para la predicción de fenómenos climáticos. Este modelo se implementará en un tablero de información público, disponible para ser insertado en plataformas web de instituciones y organizaciones relevantes.

Además, se establecerá un mecanismo de reporte accesible para la comunidad, permitiendo el envío de información que alimente el modelo a través de dispositivos tecnológicos y la red de internet. El sistema generará alertas de acuerdo con los protocolos y umbrales establecidos por las autoridades competentes, las cuales serán comunicadas mediante correos electrónicos a instituciones, organismos de socorro y organizaciones de la sociedad civil. Es importante destacar que este modelo no pretende reemplazar los protocolos y alertas existentes, sino complementarlos para mejorar la toma de decisiones por parte de las partes interesadas.

3.4 Pregunta de investigación (Max 100 Palabras)

¿Cómo puede la aplicación de un modelo predictivo basado en redes neuronales mejorar la capacidad de anticipación y respuesta ante eventos climáticos y no climáticos relacionados con inundaciones en Bogotá?

3.5 Hipotesis (Max 100 Palabras)

La implementación de redes neuronales convolucionales y recurrentes permitirá hacer una predicción de la ocurrencia de inundaciones en Bogotá basada en datos climáticos y no climáticos .

La predicción de ocurrencia de inundaciones en Bogotá podría revolucionar la capacidad de anticipación y alerta temprana de las autoridades, comunidades y organizaciones de la sociedad civil, minimizando el impacto negativo en la vida de las personas en la zona focalizada

4 Reflexiones sobre el origen de datos e información

4.1 Origen de los datos

Para el desarrollo del proyecto, se considerarán dos fuentes de información clave. En primer lugar, se aprovecharán los sistemas de información meteorológica proporcionados por entidades nacionales, regionales e internacionales, los cuales suministran datos y variables determinantes relacionados con las inundaciones. Por otra parte, se contará con un canal de chatbox integrado con WhatsApp que permitirá recopilar datos directamente desde las comunidades, incluyendo conocimientos ancestrales y reportes sobre eventos climáticos y no climáticos en tiempo real.

Informes satelitales de la Aerocivil: Este sitio web ofrece acceso a imágenes satelitales relevantes para el análisis y pronóstico meteorológico en Colombia. Las imágenes de satélite proporcionadas pueden ser valiosas para monitorear los patrones climáticos, la nubosidad y otras condiciones atmosféricas que influyen en la aviación y los pronósticos meteorológicos generales.

4.2 Consideraciones legales o éticas del uso de la información

El uso de la información en un proyecto como este conlleva una serie de consideraciones legales y éticas importantes que deben ser tenidas en cuenta. En términos generales, es fundamental destacar el compromiso con la precisión y la transparencia en la comunicación de pronósticos y alertas. Esto es crucial para evitar consecuencias negativas. Además, un mal uso de esta información podría generar pánico, lo que a su vez podría resultar en accidentes o una falta de preparación real para atender emergencias. Por lo tanto, también es esencial garantizar un acceso equitativo a esta información para asegurar la seguridad y el bienestar de toda la sociedad,

4.3 Retos de la información y los datos en Integración y Automatización de Datos para IA

Los desafíos en la Integración y Automatización de Datos para IA implican superar barreras como el acceso a datos actualizados y fiables, esenciales para construir modelos predictivos efectivos. Además, se requiere capacidad de procesamiento de datos en tiempo real para manejar el volumen y la variedad de datos de múltiples fuentes. Es crucial evitar costos excesivos al integrar estas fuentes y asegurar eficiencia y rentabilidad en el proceso. Los resultados deben ser seguros y confiables, proporcionando información clara y útil para las autoridades y comunidades. Esto permitirá una toma de decisiones optimizada y una mejor preparación y respuesta frente a los fenómenos climáticos, específicamente las inundaciones.

4.4 Resultado esperado de la Integración y Automatización de Datos para IA

Se espera que la utilización de diversas fuentes de información, junto con el conocimiento comunitario, alimente un modelo computacional autónomo para gestionar datos según parámetros técnicos establecidos. Este enfoque busca generar resultados eficientes que fortalezcan la preparación ante inundaciones, reduciendo sus impactos adversos en áreas como la salud, la educación, la economía y la salud mental. Además, se pretende abordar riesgos específicos de género, asegurando una gestión inclusiva y equitativa de desastres. Integrar esta perspectiva en la herramienta puede proporcionar información clave a instituciones para desarrollar planes de prevención y respuesta adaptados a cada comunidad, fortaleciendo la capacidad de anticipación y respuesta. Este enfoque proactivo busca no solo mitigar los efectos negativos de las inundaciones, sino también promover una mayor resiliencia comunitaria.

5 Diseño de integración y Automatización de Datos para IA (Diagrama)

En el siguiente gráfico podrán identificar el flujo de información para la integración y automatización.

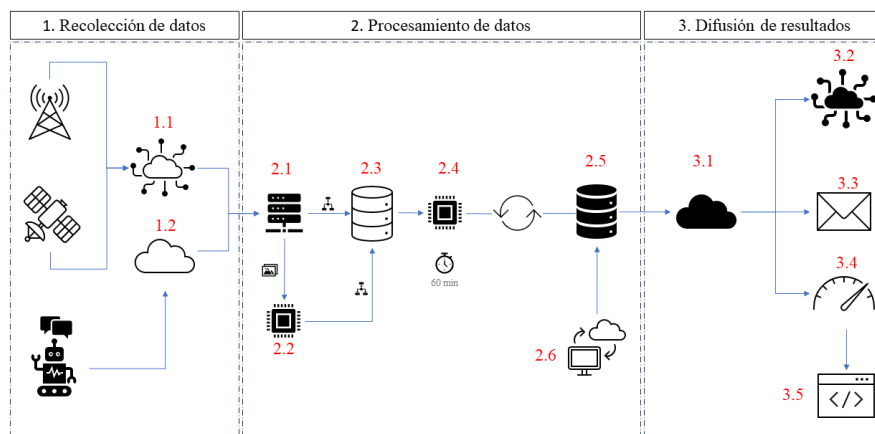


Figure 1: Diseño de integración y Automatización de Datos para IA - Fuente: elaboración propia

A continuación se describe a detalle cada uno de los componentes del diagrama, en la que reaccionan los recursos que se utilizarán en la implementación del proyecto.

1. Fuente de datos

1.1. Datos institucionales API REST

- i. Datos de Estaciones de IDEAM y de Terceros, disponible en este enlace: https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Datos-de-Estaciones-de-IDEAM-y-de-Terceros/57sv-p2fu/about_data
- ii. Imágenes satelitales de la Aeronautica Civil disponibles en el siguiente enlace: https://meteorologia.aerocivil.gov.co/satellite_image
- iii. Datos Hidrometeorológicos Crudos - Red de Estaciones IDEAM : Temperatura, disponible en este enlace: <https://www.datos.gov.co>

`gov.co/Ambiente-y-Desarrollo-Sostenible/Datos-Hidrometeorol-gicos-Crudos-Red-d
sbwg-7ju4/data_preview`

- 1.2. Chatbot PLN con Simla que integra Whatsapp una de las plataformas más usadas, simla utiliza el asistente de AI basado en ChatGPT
2. Procesamiento de datos
 - 2.1. Nodo para separar el tipo de datos, los datos estructurados pasan a la base de datos, el los datos de imágenes se envían a un procesamiento de imágenes.
 - 2.2. Se usará un modelo de procesamiento de imágenes de Python que generan como resultado datos estructurados que se almacenarán en la base de datos.
 - 2.3. Se usará el servicio de big data Azure Synapse Analytics, que integra almacenamiento, análisis y unificación de datos de diferentes fuentes en una única plataforma.
 - 2.4. Para el desarrollo del modelo predictivo con redes neuronales, se utilizarán las bibliotecas y marcos de trabajo de Python, como TensorFlow, Keras y PyTorch, los cuales simplifican la implementación de redes neuronales convolucionales y recurrentes.
 - 2.5. Se usará Azure Data Lake Storage para el almacenamiento de los datos ya procesados por el modelo predictivo y que será accedido para el desarrollo de informes y que retroalimenta el modelo para mejorar su predicción.
 - 2.6. Administración de parámetros e información de notificaciones para las alertas / Análisis y generación de indicadores.
3. Difusión de resultados
 - 3.1. Se usará Azure Data Lake Storage para el almacenamiento de los datos ya procesados por el modelo predictivo y que será accedido para el desarrollo de informes y que retroalimenta el modelo para mejorar su predicción.
 - 3.2. Se podrá disponible API para que sea accedida de forma externa, para esto se hará uso de Swagger Editor que es un editor de código abierto basado en navegador para diseñar y documentar API utilizando la especificación OpenAPI.
 - 3.3. Para la activación del sistema de alerta para los actores interesados, se hará uso de Azure Communication Services que permite la comunicación multicanal para agregar voz, vídeo, chat, mensajería de texto/SMS y correo electrónico.
 - 3.4. Se hará el desarrollo de un tablero de datos o dashboard en herramientas BI como Power BI y Tableau.
 - 3.5. Código HTML para incrustar tablero de datos en sitios web externos con analizar automatizados claros y concisos que puedan ser usados y difundidos por diferentes canales (radios, redes sociales, etc)

6 integración de Datos

Para abordar la integración de datos en este proyecto, hemos seleccionado un enfoque que aprovecha las fortalezas de varias tecnologías clave: logos de datos y datos federados. A continuación, detallamos cómo integraremos datos abiertos, MongoDB y servicios de Data Lake y Data Federation: Datos Abiertos Colombia: Utilizaremos la amplia gama de datos climáticos disponibles en Datos Abiertos Colombia, incluidas las mediciones de precipitación, niveles de ríos, y otros datos relevantes. Accederemos a estos datos a través de la API REST proporcionada por Datos Abiertos Colombia, utilizando solicitudes HTTP para obtener datos actualizados de manera dinámica. MongoDB: Implementaremos una base de datos MongoDB para almacenar y gestionar los datos climáticos recopilados de Datos Abiertos Colombia, así como otros datos relevantes generados por nuestro sistema. Utilizaremos MongoDB como una solución flexible y escalable para almacenar datos no estructurados y semiestructurados, lo que nos permitirá adaptarnos a diferentes tipos de datos y requisitos de almacenamiento. Servicios de Data Lake y Data Federation: Estableceremos un Data Lake para almacenar datos de diferentes formatos y estructuras, incluidos los datos de Datos Abiertos Colombia, así como datos generados por nuestro sistema y datos de otras fuentes. Emplearemos servicios de Data Federation para integrar datos distribuidos y heterogéneos sin necesidad de moverlos físicamente. Esto nos permitirá acceder y combinar datos de diferentes fuentes de manera virtual, simplificando así el proceso de integración y análisis de datos.

7 Automatización de Datos

A continuación, detallamos cómo automatizaremos el proceso, teniendo en cuenta la ejecución de código en Python y el uso de bibliotecas como TensorFlow, Keras y PyTorch para el desarrollo del modelo predictivo con redes neuronales:

1. Desarrollo del Modelo Predictivo: - Utilizaremos Python como lenguaje principal para el desarrollo del modelo predictivo, aprovechando su flexibilidad y las bibliotecas especializadas disponibles. - Emplearemos las bibliotecas TensorFlow, Keras y PyTorch para implementar redes neuronales convolucionales y recurrentes, simplificando así el proceso de desarrollo y entrenamiento del modelo. - Automatizaremos el proceso de entrenamiento del modelo utilizando scripts de Python, que nos permitirán ejecutar de manera eficiente múltiples experimentos con diferentes configuraciones de red y conjuntos de datos.
2. Gestión de Datos: - Automatizaremos la descarga y preprocesamiento de los datos de precipitaciones y otros datos relevantes desde fuentes externas, como Datos Abiertos Colombia, utilizando scripts de Python. - Implementaremos flujos de trabajo automatizados para la integración de datos en nuestra base de datos MongoDB y en el Data Lake, asegurando la

disponibilidad y la actualización constante de los datos para el entrenamiento y evaluación del modelo.

3. Evaluación y Optimización del Modelo: - Estableceremos procedimientos automatizados para la evaluación del rendimiento del modelo predictivo, utilizando métricas específicas como la precisión, el recall y el F1-score.
- Emplearemos técnicas de optimización automática de hiperparámetros para mejorar el rendimiento del modelo, ajustando automáticamente los valores de los parámetros del modelo para maximizar su precisión y generalización.
4. Despliegue y Monitoreo: - Automatizaremos el proceso de despliegue del modelo en un entorno de producción, asegurando su disponibilidad para la generación de predicciones en tiempo real. - Implementaremos sistemas de monitoreo automatizado para supervisar el rendimiento del modelo en producción, detectar posibles problemas y realizar ajustes automáticos si es necesario.

8 Inteligencia Artificial

La inteligencia artificial (IA) desempeñará un papel crucial en nuestro proyecto, permitiéndonos aprovechar el poder del procesamiento del lenguaje natural (NLP) para entender y analizar el lenguaje humano en WhatsApp y otras plataformas de mensajería. A continuación, detallamos cómo implementaremos la IA en nuestro proyecto:

1. API de ChatGPT: - Utilizaremos la API de ChatGPT para integrar capacidades de procesamiento del lenguaje natural en nuestro sistema. - ChatGPT es un modelo de lenguaje desarrollado por OpenAI que puede generar respuestas coherentes y relevantes a partir de mensajes de texto en lenguaje natural. - Utilizaremos la API de ChatGPT para responder a preguntas, proporcionar información y mantener conversaciones naturales con los usuarios en WhatsApp y otras plataformas de mensajería.
2. Integración con WhatsApp: - Implementaremos una solución de integración con WhatsApp que permita a los usuarios enviar mensajes de texto a nuestro sistema y recibir respuestas automáticas generadas por ChatGPT. - Utilizaremos APIs y servicios de mensajería para interactuar con la plataforma de WhatsApp y enviar y recibir mensajes de manera programática.
3. Recopilación de Información en Lenguaje Natural: - Aprovecharemos la capacidad de ChatGPT para comprender el lenguaje humano y extraer información relevante de los mensajes enviados por los usuarios en WhatsApp. - Utilizaremos técnicas de procesamiento del lenguaje natural para analizar los mensajes de los usuarios y identificar información relacionada con eventos climáticos y condiciones de inundaciones en Bogotá.

4. **Análisis y Extracción de Datos:** - Analizaremos los mensajes de los usuarios en tiempo real para identificar patrones, tendencias y solicitudes comunes relacionadas con eventos climáticos y condiciones de inundaciones. - Utilizaremos algoritmos de IA para extraer datos relevantes de los mensajes de los usuarios y alimentar estos datos en nuestro sistema de predicción de inundaciones y alerta temprana.
5. **Interacción y Retroalimentación:** - Mantendremos conversaciones interactivas y naturales con los usuarios a través de la integración con ChatGPT, proporcionando respuestas útiles y relevantes a sus preguntas y consultas. - Utilizaremos la retroalimentación de los usuarios para mejorar continuamente la capacidad de nuestro sistema para comprender y responder de manera efectiva a las solicitudes de información relacionadas con eventos climáticos y condiciones de inundaciones en Bogotá.

9 Proximos pasos (*Tercera entrega*)

10 Lecciones aprendidas (*Tercera entrega*)

11 Bibliografía