Propuesta de Proyecto

Seminario de Analítica y Ciencia de Datos

# Presentación del Proyecto

Los radares meteorológicos son esenciales para prever la lluvia en grandes áreas. En Medellín y el Valle de Aburrá, el Sistema de Alertas Tempranas de Medellín y el Valle de Aburrá (SIATA) opera uno de estos radares que permite monitorear el 97% de Antioquia. Sin embargo, la estimación de la precipitación por parte de este sensor aún tiene un margen de incertidumbre importante de alrededor del 30%, asociado a su principio de medición (Sepúlveda Berrío, 2015).

Al integrar modelos avanzados de Machine Learning, no solo se mejorará la estimación de la lluvia, sino que también permitirían al SIATA realizar investigaciones con información más precisa lo que se traduciría en una gestión del riesgo y desastres más eficaz, reduciendo potencialmente el impacto social, económico y ambiental en la comunidad por parte de fenómenos hidrometeorológicos.

Con esto, no solo se optimiza la tecnología actual, sino que también se contribuye a una estimación más efectiva y precisa de la precipitación en la región.

# Objetivo del Proyecto

Implementar modelos de Machine Learning para mejorar la estimación de la precipitación proporcionada por el radar meteorológico del SIATA, con el fin de reducir la incertidumbre en la cuantificación de la lluvia y optimizar la precisión de la información meteorológica que permita mejorar los procesos de gestión del riesgo a desastres.

# Contexto del Problema

El entendimiento de la variabilidad espacial y temporal de la precipitación ha sido suscitado a lo largo del tiempo, ya que esta tiene implicaciones directas en actividades productivas en diferentes sectores económicos, salud y en las condiciones de riesgo por eventos extremos que pueden afectar los asentamientos humanos (Sloat et al., 2018).

El SIATA ha hecho un gran esfuerzo por construir una amplia red de sensores pluviómetros) que permita el monitoreo de la precipitación in-situ, cuantificando detalladamente la lluvia que llega a superficie, pero careciendo del entendimiento de la variabilidad espacial de la lluvia. En este sentido, el radar meteorológico, por el contrario, es capaz de brindar información espacial de la precipitación, haciendo de este una herramienta de alto valor en el estudio de la dinámica de los sistemas que generan precipitación en una región. Sin embargo, si bien esta información tiene un sin número de aplicaciones, existen limitaciones asociadas al principio de medición de la precipitación, pues esta no es medida de manera directa por el radar meteorológico, sino que corresponden a un proxy de la cantidad de gotas que tiene una nube, lo cual se convierte en un indicador de la intensidad de lluvia que llega a superficie. Así pues, se hace necesario desarrollar metodologías que permitan una correcta estimación de la intensidad de la precipitación y la cantidad de esta usando de forma conjunta las mediciones del radar meteorológico y de estaciones en tierra como lo son los pluviómetros (Sepúlveda Berrío, 2015).

# Propuesta de Solución y Alcance

La metodología propuesta para el desarrollo de este proyecto se basa en tres pilares fundamentales: Obtención de datos, preprocesamiento de datos, desarrollo y evaluación del modelo.

1. **Obtención de datos**:

En primer lugar, se obtendrá la información del radar meteorológico del AMVA, ubicado en el municipio de Santa Elena, el cual cuenta con información desde el año 2013 a una resolución temporal de 5 minutos y cubre un radio de 120 km a la redonda. Por otro lado, se recopilará información de las estaciones pluviométricas, tanto al interior del Valle de Aburrá como por fuera de este, que también son operadas por el proyecto SIATA. Esta red cuenta con alrededor de 160 instrumentos que registran información de precipitación a una escala temporal de 1 minuto y será la información que permita entrenar el modelo para mejorar la estimación de la cuantificación de la precipitación.

1. **Preprocesamiento de datos:**

Dado que la resolución temporal del radar es más amplia en comparación con la de los pluviómetros, se ajustará la información de estos últimos para que coincida con la escala temporal del radar. Específicamente, los datos de acumulados de lluvia de 1 minuto de los pluviómetros se llevarán a acumulados de 5 minutos. Este ajuste permitirá una comparación coherente y efectiva entre ambas bases de datos.

Por otro lado, es importante considerar que las 160 estaciones pluviométricas no fueron instaladas en el mismo año. Por lo tanto, es necesario seleccionar las estaciones que proporcionen datos en un rango de tiempo coincidente para asegurar la coherencia en el análisis.

Otro aspecto crucial por considerar es que, con frecuencia, tanto las estaciones pluviométricas como el radar meteorológico no registran eventos de precipitación o solo registran acumulados muy bajos, que pueden no reflejar condiciones reales de lluvia y, en cambio, estar relacionados con ruido en las mediciones. Por ello, es necesario enfocarse únicamente en eventos de precipitación que superen un umbral mínimo, asegurando así que se trate de lluvia real en la superficie y garantizando un análisis eficaz. Una vez identificados estos eventos significativos, se procederá a dividir el conjunto de datos en datos de prueba y datos de entrenamiento, para posteriormente evaluar el modelo de manera adecuada.

Finalmente, se debe tener en cuenta metodologías aplicables a datos faltantes sin perder la naturaleza del fenómeno de la lluvia, así como la selección de variables polarimétricas entregadas por el radar y que sean de relevancia para el modelo, entre otros.

1. **Desarrollo y evaluación del modelo:**

Para este proyecto se propone el uso de dos modelos de machine learning: Random Forest (RF) y Gradient Boosting (GB), los cuales han demostrado ser herramientas efectivas en tareas tanto de clasificación como de regresión aplicadas a la estimación precisa de la precipitación (Verdelho, 2024). Por su parte, la tarea de clasificación agrupa la información entre “Lluvia” y “No lluvia” a partir de umbrales de precipitación; mientras que la tarea de regresión tiene por objetivo estimar la intensidad de la precipitación a partir de los datos de “lluvia”, De esta forma se aprovecha de manera efectiva las fortalezas de ambos modelos en el tratamiento de datos meteorológicos.

Finalmente, para evaluar el desempeño del modelo se deberá comparar las tasas de precipitación estimadas con los valores de los pluviómetros del conjunto de datos de prueba. Para determinar la precisión del modelo de regresión, se utilizarán métricas estadísticas como el error cuadrático medio (RMSE), el error absoluto medio (MAE) y la eficiencia de Kling-Gupta (KGE), además de utilizar matrices de confusión y métricas de exactitud que permitan una evaluación precisa de este.

# Definición del Alcance

Este proyecto pretende desarrollar un modelo que fusione diferentes fuentes de datos de precipitación y emplee un enfoque combinado de técnicas tanto de clasificación, para determinar “Lluvia” y “No lluvia”, así como de regresión para estimar la intensidad de la precipitación en los casos clasificados como “Lluvia”. Se espera que el uso de este modelo reduzca la incertidumbre generada por las técnicas convencionales actuales que son usadas para estimar dicha intensidad a partir de información de reflectividad entregada por el radar meteorológico y que, de esta manera, el proyecto SIATA y otras entidades y personas en general que se beneficien de una u otra forma de esta información puedan tener mayor certeza a la hora de tomar decisiones más precisas en términos de la información de lluvia.

Se debe aclarar que, no se considerarán parámetros meteorológicos adicionales, además de la entregada por el radar meteorológico y por las estaciones pluviométricas. Si bien esta información adicional, tal como temperatura, humedad, velocidad y dirección del viento, entre otras, podrían mejorar las estimaciones de la intensidad de la precipitación, no se tendrán en cuenta ya que un mayor conjunto de datos implicaría un preprocesamiento y evaluación adicional antes de ser utilizados por el modelo, lo que puede obstaculizar los tiempos de entrega de este proyecto.

# Riesgos e Impacto del Negocio

Actualmente, el uso de radares meteorológicos para el monitoreo de la precipitación es ampliamente utilizado en diferentes lugares del mundo. Uno de los problemas intrínsecos del uso de estos instrumentos, es que estos no miden directamente la precipitación, sino que se basan en una señal electromagnética (conocida como reflectividad) recibida desde las gotas que hay al interior de una nube. Para procesar esta reflectividad y traducirla a información de intensidad de precipitación se usan métodos matemáticos conocidos como relación Z-R que relacionan la reflectividad con la intensidad de la precipitación registrada por pluviómetros (Sepúlveda Berrío, 2015). Si bien esta estrategia ha mostrado ser efectiva en el caso del radar del AMVA está sujeta a incertidumbres relacionadas con diferentes factores tales como la complejidad de los fenómenos físicos en el desarrollo de eventos de precipitación en nuestra región. Es por esto que el uso de técnicas de machine Learning pueden capturar estos comportamientos complejos y reducir la incertidumbre mencionada anteriormente en regiones tropicales, como lo mostrado por Verdelho et al., (2024) y Rollenbeck1 et al. (2021), quienes utilizaron técnicas de machine learning para mejorar la estimación cuantitativa de la precipitación en una región de Brasil y en el norte de Perú, respectivamente y encontraron resultados satisfactorios en sus investigaciones.

En términos de los riesgos potenciales en la implementación de este modelo tiene que ver con un alto consumo de recursos computacionales, dada la amplia cobertura espacial del radar, la cantidad de estaciones pluviométricas y la fina resolución temporal de ambos conjuntos de datos. Además, la evaluación periódica y el mantenimiento del modelo desarrollado puede ser otro factor de riesgo a tener en cuenta.

Referencias

Sloat, L. L., Gerber, J. S., Samberg, L. H., Smith, W. K., Herrero, M., Ferreira, L. G., ... & West, P. C. (2018). Increasing importance of precipitation variability on global livestock grazing lands. Nature Climate Change, 8(3), 214-218. https://www.nature.com/articles/s41558-018-0081-5

Sepúlveda Berrío, J. (2015). Estimación cuantitativa de precipitación a partir de la información de Radar Meteorológico del Área Metropolitana del Valle de Aburrá (Doctoral dissertation). https://repositorio.unal.edu.co/bitstream/handle/unal/58046/1020423608.2015.pdf

Rollenbeck, R., Orellana-Alvear, J., Rodriguez, R., Macalupu, S., & Nolasco, P. (2021). Calibration of X-band radar for extreme events in a spatially complex precipitation region in north Peru: Machine learning vs. empirical approach. Atmosphere, 12(12), 1561. https://www.mdpi.com/2073-4433/12/12/1561

Verdelho, F. F., Beneti, C., Pavam Jr, L. G., Calvetti, L., Oliveira, L. E., & Zanata Alves, M. A. (2024). Quantitative Precipitation Estimation Using Weather Radar Data and Machine Learning Algorithms for the Southern Region of Brazil. Remote Sensing, 16(11), 1971. https://www.mdpi.com/2072-4292/16/11/1971