Estimación cuantitativa de precipitación mediante datos del radar meteorológico y algoritmos de aprendizaje automático.

Seminario de Analítica y Ciencia de Datos

# Resumen Descriptivo del Proyecto

[El proyecto tiene como objetivo mejorar la estimación cuantitativa de la precipitación del Valle de Aburrá mediante la integración de algoritmos, combinando Random Forest y Gradient Boosting. Al integrar datos de radares y pluviómetros, se busca reducir la incertidumbre en las mediciones, optimizando la gestión del riesgo ante desastres.]

Marco Teórico:

1. Revisión de Literatura.

[La estimación cuantitativa de la precipitación (QPE) no es una tarea trivial, en particular en regiones con topografías complejas y en sistemas climáticos altamente variables como lo son las regiones tropicales (Sepúlveda, 2015). En este sentido, diferentes estudios han intentado aprovechar las capacidades de los algoritmos de aprendizaje automático para entregar una mejor QPE, y que además pueda ser operativo.

Wolfensberger et al (2021), utilizan el algoritmo de Random Forest (RF) con el fin de mejorar estas estimaciones para una región ubicada en Suiza. Para esto cuentan con una base de datos que comprende información de un radar meteorológico (del cual se obtiene el QPE) y de estaciones pluviométricas en tierra, las cuales permiten que el modelo identifica la intensidad real de la precipitación que llega a superficie. Una vez ajustan los hiperparámetros de su modelo, Wolfensberger et al (2021) comparan los resultados del modelo contra el QPE estimado por métodos tradicionales. Los resultados sugirieron que el modelo de RF reduce considerablemente bien el sesgo de las intensidades calculadas con los métodos tradicionales, especialmente en eventos de precipitación con intensidades altas, sin embargo, este modelo mostró tener dificultades para corregir los sesgos en precipitaciones con intensidades leves, donde tiene a sobreestimar la intensidad de dichos eventos.

En el mismo sentido, Rollenbeck et al (2021) utilizan algoritmos de aprendizaje automático para calibrar el QPE de un radar meteorológico ubicado en Piura, Perú. Ellos, al igual que Wolfensberger et al (2021) utilizan el algoritmo de RF para mejorar las estimaciones de la intensidad de la precipitación, y además, comparan los resultados del modelo contra un enfoque empírico modificado de estimación del QPE. Los resultados de Rollenbeck et al (2021) muestran cómo el algoritmo de RF adquiere conocimiento sobre las características de la lluvia y crea árboles de decisión para generar estimaciones cuantitativas de precipitación correspondientes a cada rango de reflectividad observado, logrando así capturar la variabilidad espacial de la precipitación. Aunque estos resultados muestran además que el enfoque empírico tuvo un mejor desempeño en estimar la precipitación, el RF se ajusta más a la distribución de reflectividad y es independiente de los datos de pluviómetros en tiempo real.

Resultados similares fueron obtenidos por Verdelho et al (2024), quienes realizaron el mismo análisis, pero para una región al sur de Brasil. Estos, a diferencia de Wolfensberger et al (2021) y Rollenbeck et al (2021) no sólo utilizan el algoritmo de RF sino que lo combinan con un algoritmo de Gradient Boosting (GB) cuyos resultados sugieren que la combinación de estos dos algoritmos demuestran mejoras significativas en la precisión de la estimación de la precipitación. Asimismo, sus resultados muestran que el rendimiento de los modelos estuvo fuertemente influenciado por la calidad y la cantidad de los datos utilizados. El algoritmo de GB mostró mayor solidez en comparación con el RF al ser evaluado en distintos escenarios de precipitación. Así pues, los modelos de aprendizaje automático superaron los métodos convencionales para la estimación del QPE, lo que resalta su potencial en el campo de la meteorología.Realice una investigación exhaustiva sobre los trabajos previos y teorías relacionadas con su proyecto. Se espera que identifique los estudios más relevantes, métodos utilizados y cómo estos contribuyen a resolver problemas similares al suyo. Se sugiere incluir un análisis crítico de al menos 3-5 fuentes clave.]

1. Modelos y Métodos ya existentes.

[Los diferentes estudios analizados convergen en el uso del algoritmo de RF para mejorar la estimación cuantitativa de la precipitación, sin embargo, Verdelho et al (2024) muestran que el método de GB combinado con RF puede ayudar a mejorar las estimaciones de la precipitación a partir de información de radares meteorológicos. Estos algoritmos presentan ventajas significativas, como la capacidad de manejar grandes volúmenes de datos y detectar patrones que se integran en modelos de clasificación y estimación de precipitaciones.

Breiman (2001) define los bosques aleatorios como una técnica de aprendizaje en conjunto que combina múltiples árboles de decisión, los cuales son considerados como aprendices débiles, para generar una predicción más precisa a través de un esquema de votación. El algoritmo utiliza el muestreo bootstrap y la selección aleatoria de características en cada árbol de decisión. En un conjunto de datos con N muestras y M características, se crea un nuevo conjunto de entrenamiento para cada árbol, tomando muestras aleatorias con reemplazo. Luego, se generan árboles de decisión donde en cada división del árbol solo se consideran m características seleccionadas al azar (con m < M). Este proceso se repite hasta formar el bosque aleatorio. En el caso de la regresión, la predicción final es el promedio de los resultados obtenidos por todos los árboles.

Como lo indican Wolfensberger et al (2021), dado que la regresión de RF es un promedio de los árboles, hay una subestimación de los valores extremos fuertes y una sobreestimación de los valores extremos débiles, sin embargo, estos indican que estos problemas pueden ser corregidos ajustando, adecuadamente, algunos hiperparámetros del modelo. Esta misma conclusión fue obtenida por Rollenbeck et al (2021)

Por su parte, Verdelho et al (2024) combinan el método de RF con el de GB dado que su excelente rendimiento en tareas tanto de regresión como de clasificación, que son importantes para una predicción más precisa de la precipitación. El método GB construye una serie de modelos ordenados secuencialmente, en el que cada modelo posterior tiene la tarea de corregir los errores del modelo antecesor.

Este enfoque combinado, que integró el método de conjunto de RF con la capacidad de reducción de errores del GB, les permitió desarrollar un modelo de predicción cuantitativa más preciso y robusto, aprovechando eficientemente las ventajas de ambas técnicas para trabajar con datos meteorológicos complejos.

Basándose en los resultados obtenidos por Verdelho et al. (2024), el presente proyecto no se limitará a un solo algoritmo. Se propone una combinación de RF y GB, aprovechando las fortalezas de ambos modelos, dado que se tiene como objetivo mejorar la precisión en la estimación de la precipitación y superar las limitaciones observadas en estudios anteriores, donde el RF mostró dificultades para estimar intensidades leves de precipitación. Al integrar estos métodos, se busca optimizar las predicciones y ofrecer un enfoque más completo para el análisis climático.]

1. Diferenciación del Proyecto

[En un contexto donde la gestión del riesgo y la prevención de desastres son cada vez más críticas, el proyecto que busca integrar modelos avanzados de Machine Learning en el radar meteorológico del Sistema de Alertas Tempranas de Medellín y el Valle de Aburrá (SIATA) se distingue por su enfoque innovador y multidimensional, dado que a diferencia de los métodos convencionales que dependen únicamente de los datos del radar meteorológico o de pluviómetros por separado, se propone una integración sinérgica de ambas herramientas. Al combinar datos de medición espacial (radar) con datos in-situ (pluviómetros), se logrará una representación más precisa de la variabilidad espacial y temporal de la precipitación. Esta metodología no solo mejora la estimación de la lluvia, sino que también proporciona un marco más robusto para el análisis de fenómenos hidrometeorológicos.

Posteriormente, al utilizar la combinación de Random Forest (RF) y Gradient Boosting (GB), se optimiza la estimación de la aceleración al aprovechar las fortalezas de ambos algoritmos, lo que permitirá abordar el margen de incertidumbre del 30% que actualmente afecta las estimaciones del SIATA. A través de estos algoritmos que aprenden patrones complejos en los datos, se podrá generar un modelo predictivo que ajuste dinámicamente las estimaciones de precipitación, adaptándose a diferentes condiciones atmosféricas y geográficas.

A diferencia de los enfoques empíricos, se enfatiza la calibración continua del modelo mediante el ajuste de hiperparámetros, algo que se ha demostrado efectivo en estudios como los de Wolfensberger et al (2021) y Rollenbeck et al (2021). Este enfoque proactivo busca minimizar sesgos en las estimaciones, especialmente en eventos de precipitación de diferentes intensidades.

La propuesta se basa en una metodología bien definida que incluye la obtención y preprocesamiento de datos, así como el desarrollo y evaluación del modelo. Este enfoque sistemático no solo busca obtener resultados confiables, sino que también facilita la replicabilidad del proyecto en otras regiones que enfrentan desafíos similares, aumentando su impacto potencial.]

1. Referencias.

Breiman, L. (2021). Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324.

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

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

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

Wolfensberger, D., Gabella, M., Boscacci, M., Germann, U., & Berne, A. (2021). RainForest: a random forest algorithm for quantitative precipitation estimation over Switzerland. Atmospheric Measurement Techniques, 14(4), 3169-3193. https://doi.org/10.5194/amt-14-3169-2021.