

PRIMERA ENTREGA DE PROYECTO

JOSÉ DAVID HERNÁNDEZ BUELVAS  
SANTIAGO ROJAS POSADA

INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

RAÚL RAMOS POLLÁN



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE INGENIERÍA  
MEDELLÍN  
2022

## 1. Planteamiento del problema

Para algunos sectores como la agricultura es de suma importancia conocer cuánto llovió en un determinado campo. No obstante, es muy difícil medir debido a que estas son variables en el espacio y el tiempo, y resulta casi imposible cuantificarlas con pluviómetros en todas partes. Es así entonces, por lo que se emplean radares polarimétricos para la estimación de las precipitaciones, sin embargo, actualmente las observaciones de radar se "corrigen" utilizando medidores cercanos y se proporciona una sola estimación de la lluvia a los usuarios que necesitan saber cuánto llovió. Es por esta razón que se pretende desarrollar un modelo probabilístico de la precipitación horaria con una cobertura más amplia, la cual podría impulsar modelos hidrológicos y agronómicos.

## 2. Dataset

El dataset que se va a emplear proviene de una competencia de Kaggle, este se divide en dos, un archivo que será de entrenamiento de algoritmos y otro para pruebas llamados train y test respectivamente. El primero proporciona datos recopilados en 20 días entre abril y agosto de 2014 en los estados productores de maíz del medio oeste de estados unidos, mientras que datos de prueba consisten en datos de los mismos radares y medidores durante los días restantes de ese mes.

Ambos archivos contienen columnas de información tales como:

- Id: El identificador único para cada una de las observaciones durante una hora en un indicador.
- minutes\_past: Para cada conjunto de observaciones de radar, los minutos pasados de la parte superior de la hora en que se llevaron a cabo las observaciones de radar. Las observaciones de radar son instantáneas en ese momento.
- radardist\_km: Distancia de ancho de vía del radar cuyas observaciones se están reportando.
- Ref: Reflectividad del radar en km.
- Ref\_5x5\_10th: Percentil 10 de los valores de reflectividad en el vecindario 5x5 alrededor del medidor.
- Ref\_5x5\_50th: El percentil 50.
- Ref\_5x5\_90th: El percentil 90.
- RefComposite: Reflectividad máxima en la columna vertical por encima del medidor. En dBZ.
- RhoHV: Coeficiente de correlación (sin unidad).
- RhoHV\_5x5\_10th
- RhoHV\_5x5\_50th
- RhoHV\_5x5\_90th
- Zdr: Reflectividad diferencial en dB.
- Zdr\_5x5\_10th

- Zdr\_5x5\_50th
- Zdr\_5x5\_90th
- Kdp: Fase diferencial específica (deg/km).
- Kdp\_5x5\_10th
- Kdp\_5x5\_50th
- Kdp\_5x5\_90th
- Expected: Observación real del indicador en mm al final de la hora.

Las 6 columnas restantes para tener un total de 30, se van a crear con los datos conocidos de las otras variables, sea con porcentajes, desviaciones estándar, promedios, percentiles diferentes a los ya existentes, cuartiles y caracterización por rangos de valores o categorías.

### 3. Métricas

La métrica de evaluación principal para este problema es el Error Absoluto Medio (MAE), el cual se puede estimar como se muestra a continuación:

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}.$$

Esta métrica se calcula como la suma de los errores absolutos entre los valores predichos y observados, dividida por el tamaño de la muestra. Donde,  $y_i$  es la predicción,  $x_i$  es el valor real de la variable objetivo y  $n$  el total de observaciones.

Ahora bien, la métrica de negocio que se espera en este proyecto es poder predecir, con una mayor exactitud, las mediciones de lluvia por hora a partir de los datos arrojados por radares polarimétricos y de este modo conocer cómo se dan las precipitaciones en lugares donde la lluvia tiene alto grado de influencia en diferentes ámbitos.

### 4. Desempeño

Con el desarrollo de un modelo como este se espera que se tengan predicciones en el total de precipitaciones que se tienen en una zona determinada durante periodos de tiempo, con intervalos de una hora. Dicho esto, se espera tener una mejor aproximación a los modelos climatológicos que se tienen para predecir el total de lluvias, que son de vital importancia en diferentes áreas, como es el caso del sector agricultor, el cual en muchas ocasiones para el cuidado de sus cultivos se ve dependiente de las condiciones climatológicas de la zona de interés, no obstante, a pesar de que el modelo a implementar se hace sobre un área específica, este puede ser replicado en otros territorios con las mediciones adecuadas.

### 5. Bibliografía

- "How Much Did It Rain? II | Kaggle". Kaggle: Your Machine Learning and Data Science Community.

<https://www.kaggle.com/competitions/how-much-did-it-rain-ii/overview>  
(accedido el 2 de julio de 2022).