

# Informe

Carmen Calvo Olivera

4 de junio de 2020

## Resumen

En este documento se recogen los resultados obtenidos de la aplicación de diferentes técnicas de aprendizaje automático a nuestro conjunto de datos, cuyo objetivo es conseguir un modelo capaz de... calcular en tiempo real el índice de verosimilitud para una predicción meteorológica a partir de la obtención del error cuadrático medio entre la precipitación obtenida por el modelo WRF (Weather Research and Forecasting) y nuestro modelo.

Para ello, se ha llevado a cabo la creación de un conjunto de datos a partir de predicciones obtenidas por el modelo WRF.

## 1. Introducción

A lo largo de la historia, los fenómenos meteorológicos(?) siempre han resultado de gran interés para la humanidad dada su gran influencia en actividades cotidianas de sectores como la agricultura, el transporte o el comercio.

Poner párrafo con el avance del aprendizaje automático y la extensión de su uso...

Ya en 1998, en [1], desarrollan una red neuronal, utilizando como datos de entrada la información del Modelo ETA y los sondeos de aire superior que cuentan con un total de 19 variables meteorológicas, para obtener la probabilidad de precipitación (PoP) y el pronóstico cuantitativo de precipitación (QPF) para el área de Dallas-Fort Worth, Texas. Para ello, implementan dos redes: una red QPF para la cantidad de precipitación y una red PoP para la probabilidad o confianza en el pronóstico.

En 2012, en [2], basan su trabajo en un trabajo previo (Pascual et al. (2012)) y muestran una metodología que pretende calibrar la importancia de las diferentes variables incluidas en un modelo de análogos, mostrando también los resultados de la aplicación del modelo para la predicción de precipitación y nieve en un entorno de orografía compleja como es el área de la Sierra de Guadarrama. Para conseguirlo, hacen uso de modelos de análogos. Los modelos de análogos son un tipo de modelos estadísticos que se basan en el reconocimiento de patrones atmosféricos y búsqueda de situaciones meteorológicas similares en una base de datos histórica. Con ellos, se realiza una comparación entre patrones atmosféricos mediante una función que cuantifica la similitud entre el patrón de análisis y los patrones históricos. Como datos, utilizan datos obtenidos de la AEMET con variables de lluvia (mm) y nieve (dicotómica) y variables de gran

escala (ECMWF), a las que se les aplica un análisis previo de componentes principales para reducir la dimensionalidad espacial. Como resultado, ...

En 2015, en [3] tienen como objetivo predecir la intensidad de las precipitaciones futuras en una región local en un período de tiempo relativamente corto (0-6h) mediante el uso de Machine Learning proponiendo como solución el uso de ConvLSTM (Convolutional LSTM), la cual está implementada en Python con Theano y comparan con FC-LSTM y ROVER (ROVER algoreithm: algoritmo de flujo óptico en tiempo real mediante métodos de variación para ecos de radar). Aunque la capa FC-LSTM ha demostrado ser poderosa para manejar la correlación temporal, contiene demasiada redundancia para los datos espaciales. Para abordar este problema, proponen una extensión del FC-LSTM que tiene estructuras convolucionales tanto en las transiciones de entrada a estado como de estado a estado. Al apilar múltiples capas de ConvLSTM y formar una estructura de codificación y pronóstico, somos capaces de construir un modelo de red no sólo para el problema de la previsión de precipitaciones, sino también para problemas más generales de previsión de secuencias espacio temporales. Basan parte de su trabajo en ([23]). Como datos de entrada utilizan el dataset de MNIST adaptado a sus necesidades mediante algunas modificaciones y para llevar a cabo la evaluación, crean un conjunto de datos de ecos de radar.

En 2016, [4] es Technical tutorial que muestra las ventajas del uso de DL para RS hoy en día. Analizan los últimos desarrollos en DL para el campo RS y ofrecen un tutorial técnico sobre el diseño de métodos basados en DL para datos RS ópticos. También existen varias técnicas avanzadas DL para synthetic aperture radar images [21]-[26] y para la detección de luz y datos de nubes de puntos de alcance (LiDAR)[27], comparten las ideas básicas similares de DL del modelo de análisis de datos. Algunos puntos clave tratados son:

- Analiza ventajas e inconvenientes del uso de DL en RS.
- Marco general: El diagrama de flujo incluye tres componentes principales, los datos de entrada preparados, las redes básicas profundas y los datos de salida previstos. Los datos de entrada varían en función de la aplicación.
- Describen algunos algoritmos básicos y luego revisan las técnicas representativas en DL para el análisis de datos de RS desde cuatro perspectivas: 1) preprocesamiento de imágenes RS, 2) clasificación basada en píxeles, 3) reconocimiento de objetivos y 4) comprensión de escenas.
- Preprocesamiento de img.: dos aplicaciones típicas, la restauración de imágenes RS y el enfoque panorámico. Clasif. Píxeles: tres pasos en el esquema (datos de entrada, modelo y clasificación). En el caso de los datos de entrada, los vectores pueden ser característica espectral, la característica espacial o la característica espectro-espacial. Para las capas de la red existen tanto soluciones supervisadas (CNN) como no supervisadas (AEs, DBNs). En el paso de clasificación, en general, hay dos estilos principales de clasificadores: 1)

los clasificadores duros, como los SVM, que emiten directamente un número entero como etiqueta de clase de cada muestra [76], y 2) los clasificadores blandos, como la regresión logística, que pueden afinar simultáneamente toda la red preformada y predecir la etiqueta de clase de una manera de distribución de probabilidad.

- Reconocimiento de obj.
- Comprensión de escena.

Las redes neuronales parecen ser la opción más popular de modelos de aprendizaje de máquinas para el pronóstico del tiempo debido a la capacidad de capturar las dependencias no lineales de las tendencias meteorológicas pasadas y las condiciones meteorológicas futuras, a diferencia de los modelos de regresión lineal y regresión funcional que utilizan. También en 2016, en [5] tienen por objetivo predecir la temperatura para 7 días a partir de los datos de los últimos 2. Hacen uso de un modelo de regresión lineal y una variación sobre un modelo de regresión funcional, pudiendo este último capturar las tendencias del tiempo. Como datos de entrada, los relativos a los últimos 2 días incluyendo temperaturas (max y min), humedad media, presión atmosférica media y clasificación. Climática para cada uno. Como salida, temperaturas (max y min) de los siguientes 7 días.

En 2017, [6]. Modelo numérico de predicción meteorológica (NWP). La técnica AnEn proporciona una opción alternativa para generar pronósticos probabilísticos sin requerir el gasto computacional de un conjunto de NWP. Es interesante el uso de tec. Basadas en análogos durante los distintos años. Como trabajo previo mencionan ((Lorenz, 1969)). A tener en cuenta diferente conceptos mencionados:

- El AnEn mejora la precisión de la predicción meteorológica a corto plazo, disminuye los costes computacionales en tiempo real y proporciona una estimación de la incertidumbre espacial y temporal.
- ¿Qué es un análogo? Los métodos analógicos son técnicas en las que se compara un estado actual de la atmósfera con un repositorio de otros estados de la atmósfera para determinar el escenario más similar en el pasado (un análogo). Lorenz (1969) declaró que los análogos se refieren a "dos estados de la atmósfera que se parecen bastante estrechamente" continuó diciendo: Cada estado puede entonces considerarse equivalente al otro estado más un 'error' razonablemente pequeño". Tradicionalmente, en meteorología, los análogos se han utilizado principalmente para el procesamiento previo y posterior de las previsiones del NWP.
- ¿Cuál es la técnica del conjunto analógico y cómo funciona? El AnEn estima la distribución de probabilidad de un valor futuro de la predicción y variable, y, dada la predicción de un modelo, f, y el repositorio de la predicción pasada y los valores reales observados x.

- Punto 4. Cómo se desarrolló la técnica AnEn y cómo difiere de la investigación meteorológica anterior relacionada con la tecnología analógica. El Dr. Delle Monache identificó que el problema radicaba en el hecho de que el filtro Kalman da mayor peso a los días más recientes y decidió que necesitaba hacer una corrección usando un día (cualquier día) en el pasado que sea más similar, donde el modelo está cometiendo el mismo error. En lugar de una herramienta de post-procesamiento basada en la media de un conjunto de NWP, el AnEn basado en puntos se genera utilizando predicciones de modelos basados en dinámicas que se buscan de forma independiente en cada ubicación a lo largo de una ventana de tiempo de 3 puntos y medio utilizando una métrica multivariante (Lorenz, 1969; Delle Monache et al., 2013; Vanvyve et al., 2015). Con la técnica AnEn, se genera un conjunto a partir de un pronóstico determinista actual y un repositorio histórico de pronósticos determinista.
- Método Schaake Shuffle: En Clark et al (2004), se clasifican los miembros de un modelo de pronóstico para un día específico, se clasifican los valores de los datos para los parámetros de interés (precipitación y temperatura) de las mismas fechas en el registro histórico, se comparan los miembros del conjunto con los datos históricos clasificados y se reordenan los miembros del conjunto para que coincidan con el orden de los datos históricos.
- How does the AnEn + SS function? El AnEn elige el pronóstico pasado más similar y selecciona la observación de verificación correspondiente. El Schaake Shuffle se aplica al conjunto de análogos donde reordena los miembros resultando en una mejora de la red espacial y temporal. El proceso es estacionario. El proceso genera una realización diferente del mismo proceso estocástico. La siguiente actividad proporciona una descripción pictórica de la combinación de AnEn y Schake Shuffle con un mapa simulado, puntos y valores de datos.

En 2018, en [7] buscan la predicción de precipitaciones mediante la implementación de una API de Python para la lectura de datos meteorológicos, y se han desarrollado modelos ANN utilizando TensorFlow. El modelo resultante se usa para la predicción de la temperatura en Posgrunn. Como datos de entrada hacen uso de datos de frst.met.no y datos recolectado mediante la API de Python, estos datos son postprocesados para obtener los grados en Celsius. Cabe destacar como trabajo previo (Chauhan y Thakur, 2014) en el que se presenta una revisión sobre el uso de métodos de aprendizaje automático para la predicción del tiempo. Como trabajos previos comentan los siguientes:

- Comenta el estado de la cuestión: habla de un trabajo previo de 2007 en que usaron (MLP – multi layer perceptron- NN) entrenada y testada con datos meteo de 10 años. Utilizaban 7 variables para predecir la temperatura del aire seco. Otro trabajo mencionado de 2006 hace uso de ANN para predecir la temperatura del aire con un input linear con 5 variables.
- Otro enfoque más reciente (2015) basa su modelo en un híbrido de ANN, árboles de decisión y gaussian process modeling.

- Ve ventajas en el uso de tf (One possible ML tool which runs in Python is Google's TensorFlow. AR, ARX, and ARMAX models for linear systems are routinely used in system identification, e.g. (Ljung, 2002, 1999; Johansson, 1993).)
- API para datos: hace uso de http get para descargar los datos y genera una lista con 3 items (estación, timestamp y valor medido)

También en 2018, [8], en el que presentan mejoras de un método probabilístico de predicción cuantitativa basado en análogos, desarrollado anteriormente ((Bontron, 2004)) en pequeñas cuencas situadas en el sudeste de Francia, el cual se ha mejorado introduciendo dos variables innovadoras: la temperatura, que permite tener en cuenta los efectos estacionales, y la velocidad vertical, que permite una mejor caracterización del movimiento vertical atmosférico. El nuevo algoritmo se aplica primero en un contexto de pronóstico perfecto (las situaciones de objetivo proceden de un nuevo análisis meteorológico) y luego en un contexto de pronóstico operativo (las situaciones de objetivo proceden de pronósticos meteorológicos) durante un período de tres años. Los resultados muestran que este enfoque proporciona previsiones útiles, con una menor tasa de falsas alarmas y mejores resultados desde el día D actual hasta el día D+2.

Ya en 2019, [9] pretenden comprobar si es posible mejorar los resultados del ECMWF (Manzato et al. (2016)) mediante algunas modificaciones. Primero, los índices de sondeo calculados a partir del perfil vertical del ECMWF previsto en el centro de la llanura de la FVG se añaden a la lista de predictores candidatos. Segundo, se utilizan métodos estadísticos no lineales (ANN), además del exhaustivo enfoque de multi-regresión lineal. Se desarrollaron 32 modelos de multiregresión diferentes: para dos períodos del año (estación convectiva vs. no convectiva), para los cuatro períodos de 6 horas del día y para las cuatro subáreas (costa, llanura, prealpes y montaña) de la FVG. Estos modelos estadísticos se construyeron eligiendo un subconjunto de un conjunto de predictores candidatos derivados del pronóstico de lluvias del ECMWF para las cuatro subzonas, incluyendo sus valores absolutos, anomalías, valores estandarizados, más la media, el máximo y la desviación estándar en el tiempo y/o el espacio.

Igualmente en 2019, [10] da respuesta a preguntas como ¿Cuál es el volumen óptimo para tener una mejor precisión con un coste computacional más bajo? Y ¿qué datos son mas adecuados, antiguos o modernos o ambos? Utilizan 2 escenarios con el fin de dar respuestas a las dos preguntas anteriores. Para ello, describe un sistema de predicción meteo y análisis de un volumen de datos. (Encuentran, además, una relación entre la precisión de la predicción y la actualidad de los datos). Hacen uso de Keras y Pandas. (keras sequential model). (i) The Keras library [23] that is used to implement deep learning architecture for use with the TensorFlow framework; (ii) The Pandas library [24] that is used for data manipulation and preparation for input to the neural network. Realizan una predicción para el siguiente basándose en los datos de los 12 meses anteriores. Como datos para los entrenamientos y el test, usan datos reales de NOAA y NCDC [13]. Archivo de entrada CSV que será procesado por Panda. Tras procesarlos, se crea dos conjuntos (train y test).

Dado el objetivo de este trabajo, en este documento se recogen las diferentes etapas del procesamiento de los conjuntos de datos con los que se va a trabajar así como los resultados obtenidos.

## 2. Dataset

La preparación de los datos supone una parte esencial a la hora de trabajar con aprendizaje automático. En esta sección se describe brevemente tanto la obtención como el tratamiento (o preprocesamiento) de las predicciones a partir de las cuales se obtienen los conjuntos de datos utilizados para el entrenamiento y validación de los distintos clasificadores.

En primer lugar, cabe destacar uso de dos grandes conjuntos de datos, uno para *train* y el otro para *validation*.

- *train\_dataset*: este conjunto de datos abarca desde febrero (a la espera de añadir enero) de 2015 hasta diciembre de 2015.
- *validation\_dataset*: que abarca desde enero de 2016 hasta diciembre de 2016.

### 2.1. Obtención

Los datos de las predicciones se obtienen a partir del modelo WRF, un modelo meteorológico numérico de mesoescala no hidrostático, utilizado para finalidades tanto de predicciones operativas en tiempo real como para investigación atmosférica.

Los datos de inicialización del modelo WRF provienen de análisis y predicciones de otros NWP cuyo formato debe ser GRIB1 o GRIB2. En nuestro caso, todos los ficheros fuente han sido obtenidos de [Research Data Archive](#) del NCAR (Centro Nacional de Investigación Atmosférica).

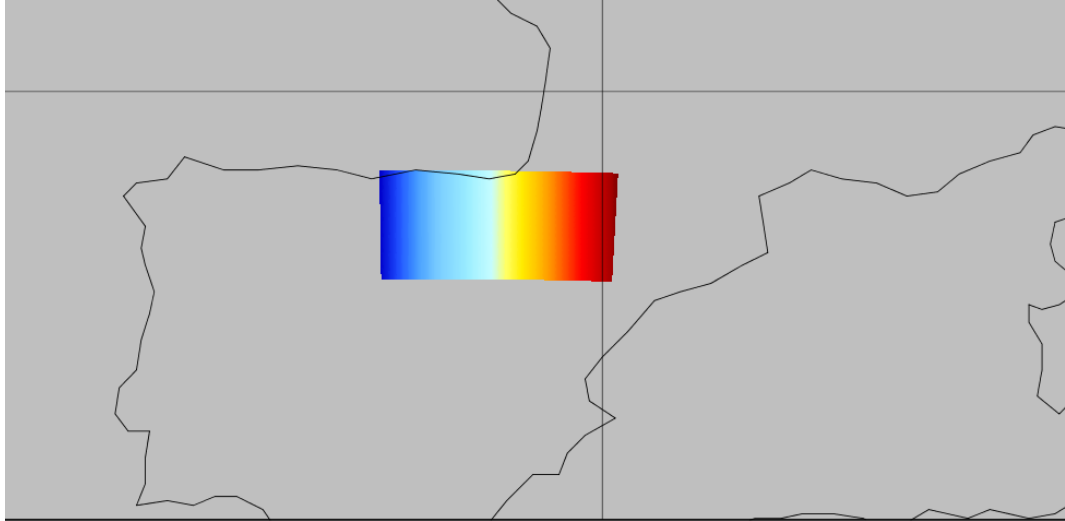
Posteriormente se llevan a cabo todos los pasos incluidos dentro del flujo de procesamiento y, a continuación, comienza la simulación meteorológica. Tras todo esto, obtenemos una serie de ficheros del tipo "*wrfout\_d0X\_yyyy-mm-dd.hh:mm:ss*" donde X es el número del dominio, y la cadena yyyy-mm-dd.hh:mm:ss representa la fecha y hora de la primera salida guardada en el fichero. Todos estos ficheros son obtenidos en formato NetCDF, un formato de archivo destinado a almacenar datos científicos multidimensionales (variables) como la temperatura, la humedad o la presión.

En nuestro caso, se hará uso de los ficheros horarios para el dominio 2 representado en la [Figura 1](#), y definido previamente en los ficheros de configuración, para 24h, es decir, para el día 1 de enero de 2016 se hace uso de 24 ficheros horarios que van desde "*wrfout\_d02\_2016-01-02\_01:00:00*" hasta "*wrfout\_d02\_2016-02-03\_00:00:00*".

### 2.2. Procesamiento

Tras la obtención de las predicciones, se llevan a cabo una serie de pasos para el tratamiento de los datos y finalmente la obtención de nuestro dataset:

**Figura 1.** Representación del dominio de predicción correspondiente a la zona del Ebro



- **Filtración de características:** el primer paso llevado a cabo es la filtración de las variables que obtenemos de las predicciones del WRF y obtener así una selección de variables detalladas en la tabla [Tabla 1](#). Mediante el uso de la librería de python [wrf-python](#) se hace un tratamiento de los datos para, a partir de una predicción horaria (por qué horaria (?)), obtener un único fichero .nc diario con toda la información que necesitamos. A continuación se muestra la cabecera de unos de los ficheros filtrados:

```
[ccalvo@frontend1 nc]$ ncdump -h 2016-04-12.nc
netcdf \2016-04-12 {
dimensions:
    south_north = 78 ;
    west_east = 123 ;
    time = 24 ;
variables:
    float XLAT(time, south_north, west_east) ;
    float XLONG(time, south_north, west_east) ;
    float HGT(time, south_north, west_east) ;
    float RAINC(time, south_north, west_east) ;
    float RAINNC(time, south_north, west_east) ;
    string DATE(time) ;
    float TIMESTAMP(time) ;
    float QVAPOR_500(time, south_north, west_east) ;
    float QVAPOR_700(time, south_north, west_east) ;
    float QVAPOR_850(time, south_north, west_east) ;
    float QCLOUD_500(time, south_north, west_east) ;
```

```

float QCLOUD_700(time, south_north, west_east) ;
float QCLOUD_850(time, south_north, west_east) ;
float QRAIN_500(time, south_north, west_east) ;
float QRAIN_700(time, south_north, west_east) ;
float QRAIN_850(time, south_north, west_east) ;
float QICE_500(time, south_north, west_east) ;
float QICE_700(time, south_north, west_east) ;
float QICE_850(time, south_north, west_east) ;
float QSNOW_500(time, south_north, west_east) ;
float QSNOW_700(time, south_north, west_east) ;
float QSNOW_850(time, south_north, west_east) ;
float QGRAUP_500(time, south_north, west_east) ;
float QGRAUP_700(time, south_north, west_east) ;
float QGRAUP_850(time, south_north, west_east) ;
float T_500(time, south_north, west_east) ;
float T_700(time, south_north, west_east) ;

```

**Tabla 1.** Variables utilizadas para la creación de un dataset

Variable	Descripción
DATE	Fecha de la predicción
TIMESTAMP	
XLAT	
XLONG	
HGT	
RAIN_C	Precipitación convectiva
RAIN_NC	Precipitación no convectiva
T_500hPa	Temperatura a diferentes presiones
T_700hPa	
T_850hPa	
QVAPOR_500	Razón de mezcla
QVAPOR_700	
QVAPOR_850	
QCLOUD_500	
QCLOUD_700	
QCLOUD_850	
QRAIN_500	
QRAIN_700	
QRAIN_850	
QICE_500	
QICE_700	

... continúa en la siguiente página



Variable	Descripción
QICE_850	
QSNOW_500	
QSNOW_700	
QSNOW_850	
QGRAUP_500	
QGRAUP_700	
QGRAUP_850	

- **Creación de los csv:** el siguiente paso es la creación de un dataset, para el cual los ficheros NetCDF se convierten en formato CSV y posteriormente se añaden algunas variables (o etiquetados de los datos). En la tabla [Tabla 2](#) se recogen las variables añadidas a nuestros ficheros.

**Tabla 2.** Variables añadidas a los datasets

Variable	Descripción
PRECIPITACION_WRF	Precipitación acumulada de la predicción del WRF (RAINCL + RAINNC)
PRECIPITACION	Precipitación acumulada real (obtenida a partir pluviómetros de la CHE)
LLUVIA_WRF	Variable binaria para la predicción del WRF (0 → No precipitación y 1 → precipitación)
LLUVIA	Variable binaria para la CHE (0 → No precipitación y 1 → precipitación)
RANGO_WRF	Rango para la predicción del WRF ([0 - 14]*)
RANGO	Rango para la CHE ([0 - 14]*)

\* Rangos (mm): 0.1,1.,1.5,2.5,5.,10.,15.,20.,25.,30.,40.,50.,80. Rangos (representación): [0 - 14]

- **Creación de un único csv:** como último paso, y previo al entrenamiento de todos los modelos, se crean dos ficheros .csv con los días que deseamos incluir en nuestro dataset de *train* o de *validation*.

### 3. Resultados

A continuación, se detallan los resultados obtenidos, los cuales se han organizado de la siguiente manera. Todos ellos cuentan con la tasa de acierto obtenida en el conjunto

de datos destinado a la validación tanto en el train como en el test y posteriormente con el error cuadrático medio.

El error cuadrático medio o mean square error (MSE) corresponde al promedio de los errores al cuadrado, es decir, la diferencia entre el valor real y el valor estimado. Se corresponde con:

$$MSE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{sample}-1} (y_i - \hat{y}_i)^2$$

Donde  $\hat{y}_i$  corresponde al valor predicho para cada  $i$ -muestra e  $y_i$  corresponde al valor real.

- Binaria: en primer lugar se muestran los resultados obtenidos para la predicción de la variable binaria con 3 escenarios distintos.
  - Todas las variables incluidas en el CSV a excepción de aquellas consideradas como etiquetas. Ver [Tabla 2](#). Resultados: [Tabla 3](#)
  - Todas las variables del punto anterior eliminando RAINC y RAINNC. Resultados: [Tabla 5](#)
  - Únicamente con las coordenadas y las variables RAINC y RAINNC. Resultados: [Tabla 6](#)
- Rango: en este caso se realiza la predicción con los rangos definidos con los 3 escenarios anteriores.
  - Todas las variables incluidas en el CSV a excepción de aquellas consideradas como etiquetas. Ver [Tabla 2](#). Resultados: [Tabla 7](#)
  - Todas las variables del punto anterior eliminando RAINC y RAINNC. Resultados: [Tabla 8](#)
  - Únicamente con las coordenadas y las variables RAINC y RAINNC. Resultados: [Tabla 9](#)

El error cuadrático medio correspondiente a la relación entre el valor predicho por el WRF y el valor real obtenido de los pluviómetros sería el siguiente:

**Error cuadrático medio (WRF - CHE)= 0.34860156086592764**

Tras este cálculo, analizamos las predicciones obtenidas de nuestros clasificadores respecto de la verdad terreno y calculamos nuevamente el MSE.

**Tabla 3.** Binaria. Todas características

Clasificador	Train	Test	MSE (model - CHE)
MPL	0.736403	<b>0.756870</b>	0.243130
LogisticRegression	0.728701	<b>0.738289</b>	0.261711
QDA	0.703161	<b>0.731465</b>	0.268535
NeuralNetwork	0.743430	0.731454	0.268546
LDA	0.740250	0.731177	0.268823
OVR	0.688272	0.727229	0.272771
AdaBoost	0.770249	0.721126	0.278874
RandomForest	0.696645	0.712761	0.287239
DecisionTree	0.773973	0.711559	0.288441
NaiveBayes	0.732275	0.705945	0.294055
KNeighbors	0.892685	0.645533	0.354467
SGDClassifier	0.393367	0.358522	0.641478

**Tabla 4.** Binaria. Todas características sin RAIN

Clasificador	Train	Test	Error cuadrático medio
SGDClassifier	0.686392	<b>0.715450</b>	0.284550
LDA	0.734469	<b>0.711946</b>	0.288054
QDA	0.688710	<b>0.704254</b>	0.295746
LogisticRegression	0.709877	0.698761	0.301239
DecisionTree	0.742524	0.698344	0.301656
NeuralNetwork	0.715451	0.694780	0.305220
RandomForest	0.669924	0.690709	0.309291
OVR	0.668298	0.690013	0.309987
MPL	0.718232	0.675204	0.324796
AdaBoost	0.730053	0.672329	0.327671
NaiveBayes	0.709254	0.630211	0.369789
KNeighbors	0.862658	0.617016	0.382984

Tras esta segunda evaluación, en la que se hace uso del dataset completo excluyendo las variables RAINC y RAINNC y las de etiquetado, se obtienen los resultados del MSE para cada día obteniendo los siguientes resultados para los 3 mejores clasificadores:

**Tabla 5.** Binaria. Todas características sin RAIN

Clasificador	Aciertos	Falsos positivos	Falsos negativos	Falsos negativos
SGDClassifier	1.313.498	453.047	69.359	1.835.904
LDA	1.307.064	282.105	246.735	1.835.904
QDA	1.292.942	467.910	75.052	1.835.904
LogisticRegression	1.282.858	336.310	216.736	1.835.904
DecisionTree	1.282.093	339.628	214.183	1.835.904
NeuralNetwork	1.275.550	319589	24076	1.835.904
RandomForest	1.268.075	566376	1453	1.835.904
OVR	1.266.798	568.505	601	1.835.904
MPL	1.239.609	272.675	323.620	1.835.904
AdaBoost	1.234.331	345.846	255.727	1.835.904
NaiveBayes	1.157.006	201.870	477.028	1.835.904
KNeighbors	1.132.782	297.721	405.401	1.835.904

**Tabla 6.** Binaria. Solo RAIN

Clasificador	Train	Test	Error cuadrático medio
MPL	0.741252	<b>0.753171</b>	0.246829
NeuralNetwork	0.744667	<b>0.750569</b>	0.249431
LogisticRegression	0.714578	<b>0.749263</b>	0.250737
NaiveBayes	0.712043	0.748209	0.251791
LDA	0.703919	0.741426	0.258574
DecisionTree	0.764922	0.719928	0.280072
AdaBoost	0.762199	0.717789	0.282211
RandomForest	0.753338	0.715067	0.284933
QDA	0.667406	0.687748	11.073562
KNeighbors	0.813913	0.678051	0.321949
OVR	0.426253	0.407584	0.592416
SGDClassifier	0.333523	0.313059	0.686941

**Tabla 7.** Rango. Todas features

Clasificador	Train	Test	Error cuadrático medio
OVR	0.669560	0.689178	9.145259
MPL	0.669085	0.688222	9.102521

...continúa en la siguiente página

<b>Clasificador</b>	<b>Train</b>	<b>Test</b>	<b>Error cuadrático medio</b>
NeuralNetwork	0.669087	0.688164	9.079000
RandomForest	0.667468	0.687758	11.070078
LogisticRegression	0.668037	0.684587	8.619534
SGDClassifier	0.664522	0.677533	8.401055
LDA	0.664978	0.675906	9.561050
AdaBoost	0.666912	0.655973	9.656310
DecisionTree	0.680278	0.651697	8.835721
NaiveBayes	0.651177	0.620479	8.792356
KNeighbors	0.850602	0.541724	10.216860
QDA	0.414040	0.394458	64.326082

**Tabla 8.** Rango. Todas features sin RAIN

<b>Clasificador</b>	<b>Train</b>	<b>Test</b>	<b>Error cuadrático medio</b>
NeuralNetwork	0.667406	0.687749	11.072704
MPL	0.667406	0.687748	11.073562
RandomForest	0.667415	0.687747	11.073196
LogisticRegression	0.665773	0.685467	10.581673
AdaBoost	0.667699	0.677845	11.002875
LDA	0.662489	0.675684	10.928471
DecisionTree	0.673827	0.662918	10.344793
SGDClassifier	0.654264	0.653064	10.036488
NaiveBayes	0.636050	0.564641	10.302585
KNeighbors	0.809269	0.535653	11.650298
QDA	0.303942	0.293877	83.531754
OVR	0.262102	0.229779	11.982337

**Tabla 9.** Rango. Solo RAIN

<b>Clasificador</b>	<b>Train</b>	<b>Test</b>	<b>Error cuadrático medio</b>
RandomForest	0.667406	0.687748	11.073562
QDA	0.667406	0.687748	11.073562
OVR	0.668867	0.687748	9.380225
NeuralNetwork	0.668598	0.687300	9.158468

... continúa en la siguiente página

Clasificador	Train	Test	Error cuadrático medio
MPL	0.668488	0.686898	9.082758
LogisticRegression	0.667769	0.683597	8.929119
LDA	0.664579	0.681965	10.134818
AdaBoost	0.668623	0.677344	9.771984
DecisionTree	0.670085	0.674760	9.498254
NaiveBayes	0.665110	0.673433	8.316886
KNeighbors	0.786454	0.592601	9.774854
SGDClassifier	0.148744	0.135874	10.248119

Por último, se ha llevado a cabo el entrenamiento de los tres mejores clasificadores para predecir el valor binario con todas las características sin incluir las variables de precipitación RAINC y RAINNC (un total de 24 variables), con diferente número de características para así intentar obtener mejores resultados. La [Tabla 10](#) recoge los resultados obtenidos correspondientes al clasificador LDA, en la [Tabla 11](#) se recogen los resultados para el clasificados QDA y en la [Tabla 12](#) para el clasificador SGD. A mayores, se han incluido los resultados obtenido para el clasificador NeuralNetwork en la [Tabla 13](#).

**Tabla 10.** Resultados para  $k$  características. LDA

$k$	Features	Train	Test	ECM
20	XLAT XLONG HGT	0.734522	0.715925	<b>0.284075</b>
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_500			
	QSNOW_700 QSNOW_850			
	QGRAUP_700			
	QGRAUP_850			

... continúa en la siguiente página

<i>k</i>	Features	Train	Test	ECM
23	XLAT XLONG HGT	0.734466	0.712566	<b>0.287434</b>
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_500			
	QCLOUD_700			
	QCLOUD_850 QRAIN_700			
	QRAIN_850 QICE_500			
	QICE_700 QICE_850			
	QSNOW_500 QSNOW_700			
	QSNOW_850			
	QGRAUP_500			
	QGRAUP_700			
	QGRAUP_850			
21	XLAT XLONG HGT	0.734249	0.712048	<b>0.287952</b>
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_700			
	QRAIN_850 QICE_500			
	QICE_700 QICE_850			
	QSNOW_500 QSNOW_700			
	QSNOW_850			
	QGRAUP_700			
	QGRAUP_850			

... continúa en la siguiente página

$k$	Features	Train	Test	ECM
22	XLAT XLONG HGT	0.734458	0.711965	0.288035
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_700			
	QRAIN_850 QICE_500			
	QICE_700 QICE_850			
	QSNOW_500 QSNOW_700			
	QSNOW_850			
	QGRAUP_500			
	QGRAUP_700			
	QGRAUP_850			
19	XLAT XLONG HGT	0.734515	0.708498	0.291502
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_500			
	QSNOW_700 QSNOW_850			
18	XLAT XLONG HGT	0.734944	0.706961	0.293039
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_500			
	QSNOW_700 QSNOW_850			
	QGRAUP_850			
3	T_500hPa T_700hPa	0.714001	0.705687	0.294313
	T_850hPa			

... continúa en la siguiente página



$k$	Features	Train	Test	ECM
14	XLAT XLONG HGT	0.717245	0.687752	0.312248
	T_500hPa T_700hPa			
	T_850hPa QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QSNOW_700 QSNOW_850			
10	QGRAUP_850	0.716688	0.687748	0.312252
	XLAT XLONG T_500hPa			
	T_700hPa T_850hPa			
	QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QSNOW_700 QSNOW_850			
15	XLAT XLONG HGT	0.717292	0.687748	0.312252
	T_500hPa T_700hPa			
	T_850hPa QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_700			
17	QSNOW_850	0.735030	0.595058	0.404942
	QGRAUP_850			
	XLAT XLONG HGT			
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_850			
16	QCLOUD_700	0.735131	0.535027	0.464973
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_700			
	QSNOW_850			
	QGRAUP_850			
4	XLAT XLONG HGT	0.714624	0.312252	0.687748
	T_500hPa T_700hPa			
	T_850hPa			

... continúa en la siguiente página

<b><i>k</i></b>	<b>Features</b>	<b>Train</b>	<b>Test</b>	<b>ECM</b>
5	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850	0.716302	0.312252	0.687748
6	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850 QSNOW_700	0.716045	0.312252	0.687748
7	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700	0.716373	0.312252	0.687748
8	XLAT XLONG T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700	0.716763	0.312252	0.687748
9	XLAT XLONG T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700 QSNOW_850	0.716719	0.312252	0.687748
11	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QSNOW_700 QSNOW_850	0.716780	0.312252	0.687748
12	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_700 QSNOW_700 QSNOW_850	0.717129	0.312252	0.687748
13	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_700 QSNOW_700 QSNOW_850 QGRAUP_850	0.717213	0.312252	0.687748

**Tabla 11.** Resultados para  $k$  características. QDA

$k$	Features	Train	Test	ECM
21	XLAT XLONG HGT	0.687764	0.704529	0.295471
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_700			
	QRAIN_850 QICE_500			
	QICE_700 QICE_850			
	QSNOW_500 QSNOW_700			
	QSNOW_850			
	QGRAUP_700			
	QGRAUP_850			
	22			
T_500hPa T_700hPa				
T_850hPa QVAPOR_500				
QVAPOR_700				
QVAPOR_850				
QCLOUD_700				
QCLOUD_850 QRAIN_700				
QRAIN_850 QICE_500				
QICE_700 QICE_850				
QSNOW_500 QSNOW_700				
QSNOW_850				
QGRAUP_500				
QGRAUP_700				
QGRAUP_850				
16	XLAT XLONG HGT	0.685077	0.703971	<b>0.296029</b>
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_700			
	QSNOW_850			
	QGRAUP_850			

... continúa en la siguiente página

$k$	Features	Train	Test	ECM
15	XLAT XLONG HGT	0.684032	0.703930	<b>0.296070</b>
	T_500hPa T_700hPa			
	T_850hPa QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_700			
	QSNOW_850			
20	QGRAUP_850	0.686165	0.702544	0.297456
	XLAT XLONG HGT			
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_500			
23	QSNOW_700 QSNOW_850	0.688333	0.700021	0.299979
	QGRAUP_700			
	QGRAUP_850			
	XLAT XLONG HGT			
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_500			
	QCLOUD_700			
	QCLOUD_850 QRAIN_700			
	QRAIN_850 QICE_500			
	QICE_700 QICE_850			
12	QSNOW_500 QSNOW_700	0.687094	0.699652	0.300348
	QSNOW_850			
	QGRAUP_500			
	QGRAUP_700			
	QGRAUP_850			
	XLAT XLONG HGT			
	T_500hPa T_700hPa			
12	T_850hPa QCLOUD_700	0.687094	0.699652	0.300348
	QCLOUD_850 QRAIN_850			
	QICE_700 QSNOW_700			
	QSNOW_850			

... continúa en la siguiente página

<i>k</i>	Features	Train	Test	ECM
17	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QVAPOR_500 QVAPOR_850 QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_500 QICE_700 QICE_850 QSNOW_700 QSNOW_850 QGRAUP_850	0.685408	0.699503	0.300497
3	T_500hPa T_700hPa T_850hPa	0.713155	0.694569	0.305431
14	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_500 QICE_700 QSNOW_700 QSNOW_850 QGRAUP_850	0.683711	0.688878	0.311122
13	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_700 QSNOW_700 QSNOW_850 QGRAUP_850	0.685100	0.688147	0.311853
10	XLAT XLONG T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QSNOW_700 QSNOW_850	0.687190	0.687748	0.312252
11	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QSNOW_700 QSNOW_850	0.687207	0.687748	0.312252

... continúa en la siguiente página

$k$	Features	Train	Test	ECM
18	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QVAPOR_500 QVAPOR_850 QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_500 QICE_700 QICE_850 QSNOW_500 QSNOW_700 QSNOW_850 QGRAUP_850	0.685491	0.681203	0.318797
19	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QVAPOR_500 QVAPOR_700 QVAPOR_850 QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_500 QICE_700 QICE_850 QSNOW_500 QSNOW_700 QSNOW_850 QGRAUP_850	0.685936	0.680043	0.319957
4	XLAT T_500hPa T_700hPa T_850hPa	0.718118	0.312252	0.687748
5	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850	0.699664	0.312252	0.687748
6	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850 QSNOW_700	0.689058	0.312252	0.687748
7	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700	0.690120	0.312252	0.687748
8	XLAT XLONG T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700	0.690843	0.312252	0.687748
9	XLAT XLONG T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700 QSNOW_850	0.687227	0.312252	0.687748

**Tabla 12.** Resultados para  $k$  características. Neural Network

$k$	Features	Train	Test	ECM
17	XLAT XLONG HGT	0.675428	0.688741	<b>0.311259</b>
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_700			
	QSNOW_850			
	QGRAUP_850			
23	XLAT XLONG HGT	0.667474	0.687889	<b>0.312111</b>
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_500			
	QCLOUD_700			
	QCLOUD_850 QRAIN_700			
	QRAIN_850 QICE_500			
	QICE_700 QICE_850			
	QSNOW_500 QSNOW_700			
	QSNOW_850			
14	XLAT XLONG HGT	0.677734	0.687883	<b>0.312117</b>
	T_500hPa T_700hPa			
	T_850hPa QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QSNOW_700 QSNOW_850			
	QGRAUP_850			
	XLAT XLONG T_500hPa			
	T_700hPa T_850hPa			
	QCLOUD_700			
10	QCLOUD_850 QRAIN_850	0.667451	0.687748	0.312252
	QSNOW_700 QSNOW_850			

... continúa en la siguiente página

$k$	Features	Train	Test	ECM
15	XLAT XLONG HGT	0.699792	0.687748	0.312252
	T_500hPa T_700hPa			
	T_850hPa QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_700			
	QSNOW_850			
18	QGRAUP_850	0.667388	0.687748	0.312252
	XLAT XLONG HGT			
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_850			
16	QICE_500 QICE_700	0.332594	0.681350	0.318650
	QICE_850 QSNOW_500			
	QSNOW_700 QSNOW_850			
	QGRAUP_850			
	XLAT XLONG HGT			
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_850			
20	QCLOUD_700	0.702025	0.663193	0.336807
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_500			
	QSNOW_700 QSNOW_850			
	QGRAUP_700			
	QGRAUP_850			

... continúa en la siguiente página



$k$	Features	Train	Test	ECM
22	XLAT XLONG HGT	0.707176	0.659079	0.340921
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_700			
	QRAIN_850 QICE_500			
	QICE_700 QICE_850			
	QSNOW_500 QSNOW_700			
	QSNOW_850			
	QGRAUP_500			
	QGRAUP_700			
	QGRAUP_850			
19	XLAT XLONG HGT	0.682827	0.654114	0.345886
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_500			
	QSNOW_700 QSNOW_850			
3	QGRAUP_850	0.624041	0.546839	0.453161
	T_500hPa T_700hPa			
21	T_850hPa	0.572248	0.508894	0.491106
	XLAT XLONG HGT			
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_700			
	QRAIN_850 QICE_500			
	QICE_700 QICE_850			
	QSNOW_500 QSNOW_700			
	QSNOW_850			
	QGRAUP_700			
	QGRAUP_850			

... continúa en la siguiente página

<b><math>k</math></b>	<b>Features</b>	<b>Train</b>	<b>Test</b>	<b>ECM</b>
4	XLAT T_500hPa T_700hPa T_850hPa	0.667484	0.312252	0.687748
5	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850	0.507912	0.312252	0.687748
6	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850 QSNOW_700	0.604810	0.312252	0.687748
7	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700	0.668499	0.312252	0.687748
8	XLAT XLONG T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700	0.509600	0.312252	0.687748
9	XLAT XLONG T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700 QSNOW_850	0.481599	0.312252	0.687748
11	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QSNOW_700 QSNOW_850	0.676183	0.312252	0.687748
12	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_700 QSNOW_700 QSNOW_850	0.332670	0.312252	0.687748
13	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_700 QSNOW_700 QSNOW_850 QGRAUP_850	0.667747	0.312252	0.687748

**Tabla 13.** Resultados para  $k$  características. Neural Network

$k$	Features	Train	Test	ECM
15	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_500 QICE_700 QICE_850 QSNOW_700 QSNOW_850 QGRAUP_850	0.680742	0.710796	<b>0.289204</b>
14	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_500 QICE_700 QSNOW_700 QSNOW_850 QGRAUP_850	0.709281	0.705374	<b>0.294626</b>
16	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QVAPOR_850 QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_500 QICE_700 QICE_850 QSNOW_700 QSNOW_850 QGRAUP_850	0.715923	0.702667	<b>0.297333</b>
19	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QVAPOR_500 QVAPOR_700 QVAPOR_850 QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_500 QICE_700 QICE_850 QSNOW_500 QSNOW_700 QSNOW_850 QGRAUP_850	0.712400	0.699214	0.300786

... continúa en la siguiente página

<i>k</i>	Features	Train	Test	ECM
23	XLAT XLONG HGT	0.718361	0.694276	0.305724
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_500			
	QCLOUD_700			
	QCLOUD_850 QRAIN_700			
	QRAIN_850 QICE_500			
	QICE_700 QICE_850			
	QSNOW_500 QSNOW_700			
	QSNOW_850			
	QGRAUP_500			
	QGRAUP_700			
	QGRAUP_850			
20	XLAT XLONG HGT	0.717065	0.693742	0.306258
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_700			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_500			
	QSNOW_700 QSNOW_850			
	QGRAUP_700			
	QGRAUP_850			
18	XLAT XLONG HGT	0.717560	0.692513	0.307487
	T_500hPa T_700hPa			
	T_850hPa QVAPOR_500			
	QVAPOR_850			
	QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_500 QICE_700			
	QICE_850 QSNOW_500			
	QSNOW_700 QSNOW_850			
	QGRAUP_850			

... continúa en la siguiente página

$k$	Features	Train	Test	ECM
22	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QVAPOR_500 QVAPOR_700 QVAPOR_850 QCLOUD_700 QCLOUD_850 QRAIN_700 QRAIN_850 QICE_500 QICE_700 QICE_850 QSNOW_500 QSNOW_700 QSNOW_850 QGRAUP_500 QGRAUP_700 QGRAUP_850	0.716623	0.690795	0.309205
8	XLAT XLONG T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700	0.718454	0.687748	0.312252
9	XLAT XLONG T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700 QSNOW_850	0.721118	0.687748	0.312252
10	XLAT XLONG T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QSNOW_700 QSNOW_850	0.723849	0.687748	0.312252
11	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QSNOW_700 QSNOW_850	0.697577	0.687748	0.312252

... continúa en la siguiente página

$k$	Features	Train	Test	ECM
17	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QVAPOR_500 QVAPOR_850 QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_500 QICE_700 QICE_850 QSNOW_700 QSNOW_850 QGRAUP_850	0.720185	0.682017	0.317983
3	T_500hPa T_700hPa T_850hPa	0.717968	0.680073	0.319927
21	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QVAPOR_500 QVAPOR_700 QVAPOR_850 QCLOUD_700 QCLOUD_850 QRAIN_700 QRAIN_850 QICE_500 QICE_700 QICE_850 QSNOW_500 QSNOW_700 QSNOW_850 QGRAUP_700 QGRAUP_850	0.719092	0.666388	0.333612
4	XLAT T_500hPa T_700hPa T_850hPa	0.712840	0.312252	0.687748
5	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850	0.713077	0.312252	0.687748
6	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850 QSNOW_700	0.702289	0.312252	0.687748
7	XLAT T_500hPa T_700hPa T_850hPa QCLOUD_850 QRAIN_850 QSNOW_700	0.719495	0.312252	0.687748
12	XLAT XLONG HGT T_500hPa T_700hPa T_850hPa QCLOUD_700 QCLOUD_850 QRAIN_850 QICE_700 QSNOW_700 QSNOW_850	0.719892	0.312252	0.687748

... continúa en la siguiente página

$k$	Features	Train	Test	ECM
13	XLAT XLONG HGT	0.682436	0.312252	0.687748
	T_500hPa T_700hPa			
	T_850hPa QCLOUD_700			
	QCLOUD_850 QRAIN_850			
	QICE_700 QSNOW_700			
	QSNOW_850			
	QGRAUP_850			

#### 4. Trabajo futuro

Como trabajo inmediato posterior, se pretende complementar los resultados con la siguiente información:

- Incluir enero de 2015 en el *dataset* de entrenamiento.
- Realizar predicciones con lo mm.
- Hacer pruebas para obtener el mejor score para distintos KBest.

#### 5. Referencias

- [1] Precipitation Forecasting Using a Neural Network
- [2] Validación de un modelo de análogos para la predicción de precipitación y nieve en la sierra de Guadarrama
- [3] Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting
- [4] Deep Learning for Remote Sensing Data
- [5] Machine Learning Applied to Weather Forecasting
- [6] Breve análisis introductorio de las técnicas basadas en análogos
- [7] Machine Learning in Python for Weather Forecast based on Freely Available Weather Data
- [8] Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin
- [9] Improving ECMWF-based 6-hours maximum rain using instability indices and neural networks
- [10] A Deep Learning-Based Weather Forecast System for Data Volume and Recency Analysis