

Informe

Carmen Calvo Olivera

16 de julio 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...

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 la necesidad del 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

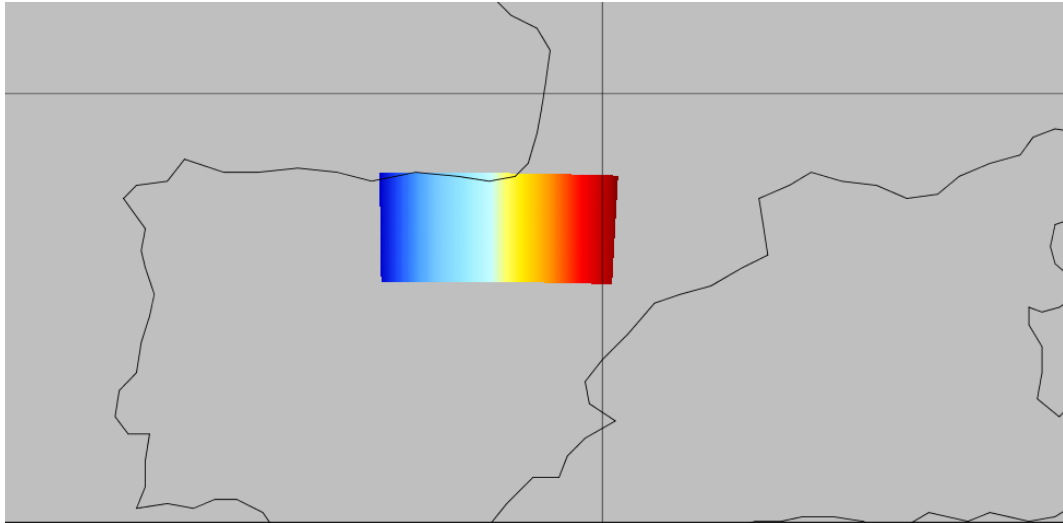
Los datos de las predicciones se obtienen a partir del modelo WRF, un modelo meteorológico numérico de mesoescala no hidrostático, utilizado para finalidades tanto de predicciones operativas en tiempo real como para investigación atmosférica.

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*".

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



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:

- **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 TIMESTAMP	Fecha de la predicción
XLAT XLONG HGT	Coordenadas
RAINNC	Precipitación convectiva
RAINNC	Precipitación no convectiva
T_500hPa T_700hPa T_850hPa	Temperatura a diferentes presiones
QVAPOR_500 QVAPOR_700 QVAPOR_850 QCLOUD_500 QCLOUD_700 QCLOUD_850 QRAIN_500 QRAIN_700 QRAIN_850 QICE_500 QICE_700 QICE_850 QSNOW_500 QSNOW_700 QSNOW_850 QGRAUP_500 QGRAUP_700 QGRAUP_850	Razón de mezcla

- **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 (RAIN + 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. Todos ellos cuentan con la tasa de acierto obtenida sobre el conjunto de datos destinado a la validación de los modelos tanto en el train como en el test y posteriormente con el error cuadrático medio.

En cuanto a la organización de las pruebas llevadas a cabo, encontramos:

- **Binaria:** en primer lugar se ha llevado a cabo en entrenamiento de diferentes clasificadores para la obtención de modelos capaces de predecir si hay precipitación (1) o no la hay (0). Los 3 escenarios planteados, es decir, el número de variables utilizadas para realizar las predicciones son los siguientes.
 - Todas las variables incluidas en el CSV a excepción de aquellas consideradas como etiquetas. Ver [Tabla 2](#).
 - Todas las variables del punto anterior eliminando RAINC y RAINNC. En este caso, podemos consultar los resultados en: [Tabla 4](#)
 - Únicamente con las coordenadas y las variables RAINC y RAINNC.

- **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](#).
 - Todas las variables del punto anterior eliminando RAINC y RAINNC. Resultados: [Tabla 8](#)
 - Únicamente con las coordenadas y las variables RAINC y RAINNC.

3.1. Validación

3.1.1. Binaria

Tabla 3. Binaria. Todas características sin RAIN

Clasificador	Train	Test	Error cuadrático medio
LDA	0.771577	0.715108	0.284892
AdaBoost	0.774947	0.713852	0.286148
DecisionTree	0.792455	0.712291	0.287709
NeuralNetwork	0.734026	0.711560	0.288440
LogisticRegression	0.733779	0.710553	0.289447
QDA	0.728639	0.703577	0.289447
RandomForest	0.712459	0.699477	0.300523
MLP	0.733471	0.697267	0.302733
SGDClassifier	0.709530	0.693649	0.306351
NaiveBayes	0.722458	0.666042	0.333958
KNeighbors	0.784630	0.630177	0.369823

3.1.2. Rango

Tabla 4. Rango. Todas características sin RAIN

Clasificador	Train	Test	Error cuadrático medio
RandomForest	0.715668	0.688289	1.291370
MLP	0.715252	0.685737	1.216564
NeuralNetwork	0.714934	0.683996	0.288440
LogisticRegression	0.715019	0.679012	1.185953
AdaBoost	0.723036	0.676638	1.244854

...continúa en la siguiente página

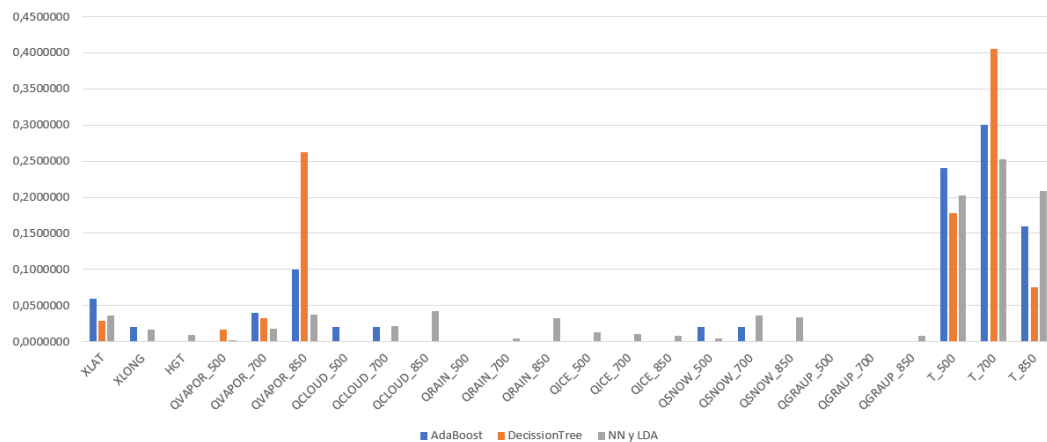
Clasificador	Train	Test	Error cuadrático medio
LDA	0.716019	0.664143	1.177988
DecisionTree	0.724356	0.657308	1.222232
KNeighbors	0.803534	0.580504	1.456317
QDA	0.667436	0.530483	1.104259
NaiveBayes	0.661476	0.516480	1.258967
SGDClassifier	0.375173	0.271367	1.107316

3.2. Features Ranking

3.2.1. Binaria

Teniendo en cuenta los 4 mejores modelos.

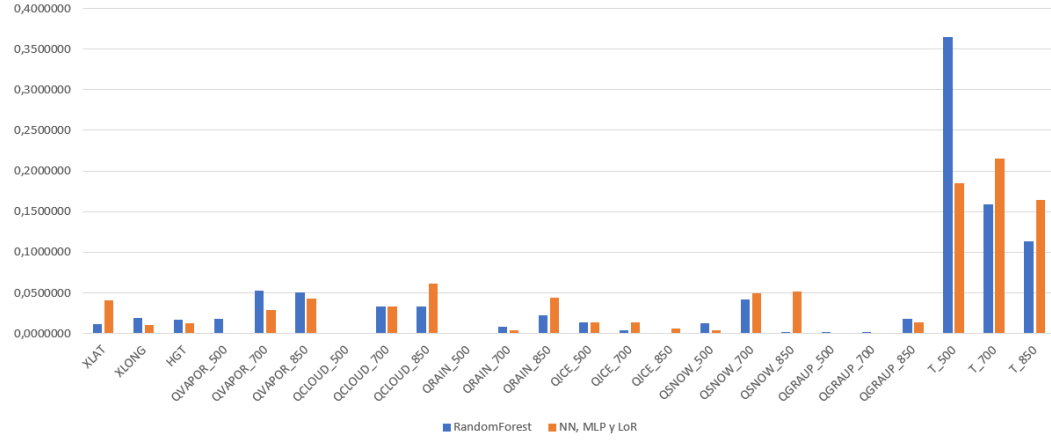
Figura 2. Binaria. Tasa de importacia de cada feature dentro de los 4 mejores clasificadores



3.2.2. Rango

Teniendo en cuenta los 4 mejores modelos.

Figura 3. Rango. Tasa de importacia de cada feature dentro de los 4 mejores clasificadores



3.3. Error Medio

El error medio aplicado, corresponde a la diferencia entre el valor real y el valor estimado de cada muestra. Se corresponde con:

$$ME(y, \hat{y}) = (y_i - \hat{y}_i)$$

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

Para ello, se añadieron a los datasets dos campos más:

- MSE_BINARIA
- MSE_RANGO

Para el entrenamiento de los clasificadores, se han utilizado todas las características sin la precipitación ni las variables de etiquetado. Los resultados obtenido, se recogen en las 2 siguientes tablas:

3.3.1. Binaria

Tabla 5. Binaria. Todas características sin RAIN. Error Medio

Clasificador	Train	Test	Error cuadrático medio
LDA	0.755772	0.654692	0.345795
LogisticRegression	0.754879	0.651824	0.348382

...continúa en la siguiente página

Clasificador	Train	Test	Error cuadrático medio
NeuralNetwork	0.755453	0.651398	0.348602
RandomForest	0.755508	0.651396	0.348604
DecisionTree	0.765977	0.651380	0.353125
AdaBoost	0.740862	0.644787	0.356341
SGDClassifier	0.683076	0.622269	0.411483
NaiveBayes	0.716763	0.617307	0.386409
KNeighbors	0.764198	0.604598	0.436976

3.3.2. Rango

Tabla 6. Rango. Todas características sin RAIN. Error Medio

Clasificador	Train	Test	Error cuadrático medio
RandomForest	0.742000	0.635237	1.077189
NeuralNetwork	0.741769	0.634096	1.078166
LogisticRegression	0.740528	0.632942	1.078572
AdaBoost	0.740355	0.631879	1.093335
DecisionTree	0.742993	0.631253	1.086142
LDA	0.736196	0.627162	1.104678
SGDClassifier	0.716308	0.607397	1.121389
NaiveBayes	0.702831	0.581695	1.186929
KNeighbors	0.706923	0.539344	1.421179

3.4. Error Cuadrático Medio

3.4.1. Binaria

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 y y_i corresponde al valor real.

Analizamos las predicciones obtenidas de nuestros clasificadores respecto de la verdad terreno y calculamos el MSE para los 4 mejores clasificadores:

Tabla 7. MSE diario. Todas features sin RAIN

Día	LDA	DecisionTree	AdaBoost	NeuralNetwork
05/01/2016	0.198039	0.255686	0.298039	0.303725
08/01/2016	0.329020	0.283922	0.371176	0.313333
09/01/2016	0.074510	0.090588	0.244706	0.062157
10/01/2016	0.007451	0.014314	0.003333	0.012745
11/01/2016	0.123137	0.313333	0.270000	0.190588
13/01/2016	0.013922	0.016275	0.015882	0.055098
15/01/2016	0.300980	0.195686	0.157843	0.106471
18/01/2016	0.023529	0.024314	0.034706	0.035686
19/01/2016	0.212941	0.426275	0.206078	0.058039
21/01/2016	0.459020	0.597451	0.918235	0.670588
23/01/2016	0.902157	0.901765	0.972549	0.887255
24/01/2016	0.880980	0.896471	0.620196	0.717059
26/01/2016	0.135686	0.162549	0.125294	0.466275
27/01/2016	0.721373	0.816078	0.704510	0.932157
28/01/2016	0.021961	0.019804	0.096863	0.117451
30/01/2016	0.512941	0.446078	0.057451	0.272353
06/02/2016	0.411569	0.411569	0.431961	0.767843
11/02/2016	0.250980	0.401373	0.429412	0.476471
13/02/2016	0.012157	0.226275	0.247451	0.373333
16/02/2016	0.137059	0.078235	0.050784	0.071961
19/02/2016	0.160000	0.095686	0.100392	0.171765
20/02/2016	0.056471	0.073922	0.088235	0.184510
22/02/2016	0.920196	0.373922	0.170196	0.228235
25/02/2016	0.416275	0.320784	0.310196	0.435098
27/02/2016	0.096863	0.198039	0.128235	0.000392
29/02/2016	0.214314	0.172941	0.189804	0.194314
01/03/2016	0.412745	0.242549	0.223333	0.185882
02/03/2016	0.129020	0.050980	0.050000	0.349608
03/03/2016	0.124118	0.380392	0.441569	0.363333
04/03/2016	0.115882	0.131373	0.167647	0.203529
06/03/2016	0.045098	0.035098	0.041569	0.043725
09/03/2016	0.001961	0.001765	0.001961	0.002157
10/03/2016	0.023137	0.022549	0.021569	0.167843
12/03/2016	0.316471	0.078039	0.101765	0.230588
14/03/2016	0.149608	0.155098	0.147647	0.156667
17/03/2016	0.055490	0.012353	0.288235	0.008431
18/03/2016	0.158824	0.233529	0.154510	0.284118
23/03/2016	0.310196	0.317843	0.371373	0.705882
28/03/2016	0.209608	0.036078	0.384706	0.504314

... continúa en la siguiente página

Día	LDA	DecisionTree	AdaBoost	NeuralNetwork
29/03/2016	0.635882	0.639216	0.738235	0.620588
30/03/2016	0.429608	0.170588	0.291373	0.119020
03/04/2016	0.065686	0.080392	0.077255	0.092353
06/04/2016	0.644902	0.636667	0.605294	0.205882
08/04/2016	0.415294	0.427255	0.429804	0.443922
10/04/2016	0.424314	0.841961	0.912941	0.643137
11/04/2016	0.001765	0.164902	0.003725	0.230000
12/04/2016	0.178431	0.180196	0.179412	0.396471
13/04/2016	0.590784	0.619020	0.590392	0.593529
15/04/2016	0.370784	0.465686	0.109216	0.336667
17/04/2016	0.885686	0.792353	0.924314	0.934314
19/04/2016	0.391373	0.439804	0.391176	0.539412
21/04/2016	0.029804	0.286078	0.264902	0.194510
23/04/2016	0.253137	0.012353	0.001569	0.431176
25/04/2016	0.571569	0.119020	0.574902	0.382353
27/04/2016	0.492549	0.141176	0.204118	0.559412
28/04/2016	0.033725	0.012941	0.000392	0.000196
30/04/2016	0.014902	0.022157	0.070196	0.005098
01/05/2016	0.230784	0.575294	0.676471	0.810980
03/05/2016	0.000000	0.000000	0.000000	0.005490
07/05/2016	0.949216	0.458431	0.107647	0.296275
08/05/2016	0.361569	0.380392	0.758627	0.538627
09/05/2016	0.550000	0.598824	0.749608	0.577647
10/05/2016	0.104706	0.005882	0.272745	0.043529
12/05/2016	0.486471	0.250784	0.305098	0.415490
13/05/2016	0.489412	0.471176	0.425686	0.263922
17/05/2016	0.010980	0.194118	0.162745	0.131176
18/05/2016	0.938431	0.620196	0.931569	0.722745
19/05/2016	0.078431	0.089804	0.068824	0.061373
20/05/2016	0.958824	1.000000	0.902353	0.990588
21/05/2016	0.063137	0.063137	0.063137	0.066078
23/05/2016	0.024902	0.197059	0.499804	0.521176
24/05/2016	0.000196	0.030588	0.061961	0.060392
26/05/2016	0.026471	0.061961	0.005490	0.079804
29/05/2016	0.709020	0.719804	0.639020	0.543137
31/05/2016	0.623137	0.624118	0.187647	0.713137
01/06/2016	0.001176	0.000000	0.005098	0.040000
06/06/2016	0.055686	0.047255	0.065098	0.000392
08/06/2016	0.085686	0.015490	0.179804	0.107451
14/06/2016	0.268824	0.176863	0.001176	0.342745
18/06/2016	0.089216	0.089216	0.089216	0.088824

... continúa en la siguiente página

Día	LDA	DecisionTree	AdaBoost	NeuralNetwork
19/06/2016	0.009608	0.022941	0.025098	0.154706
20/06/2016	0.000000	0.000000	0.001961	0.000000
22/06/2016	0.000196	0.000000	0.007451	0.000000
23/06/2016	0.193922	0.335490	0.131961	0.417059
24/06/2016	0.169804	0.008627	0.182157	0.000000
26/06/2016	0.922745	0.653333	0.686471	0.752549
28/06/2016	0.000000	0.000000	0.000000	0.000000
29/06/2016	0.224314	0.495098	0.921765	0.666275
02/07/2016	0.098039	0.000000	0.083137	0.000000
03/07/2016	0.001961	0.000000	0.000000	0.000000
06/07/2016	0.075294	0.000000	0.111569	0.000000
07/07/2016	0.623922	0.661569	0.454118	0.661569
08/07/2016	0.170000	0.196078	0.377059	0.004706
09/07/2016	0.970784	0.970784	0.970784	0.970784
10/07/2016	0.019216	0.539020	0.699412	0.493137
11/07/2016	0.107255	0.000000	0.027255	0.000000
13/07/2016	0.405294	0.405294	0.407255	0.405882
15/07/2016	0.013725	0.000000	0.038235	0.000000
16/07/2016	0.000000	0.003725	0.089804	0.015882
17/07/2016	0.033922	0.049804	0.044118	0.000000
19/07/2016	0.337843	0.337843	0.337059	0.337843
21/07/2016	0.246275	0.672941	0.197843	0.425294
23/07/2016	0.094118	0.023333	0.186275	0.000392
26/07/2016	0.161176	0.188431	0.397647	0.252941
29/07/2016	0.080784	0.080784	0.080784	0.080784
30/07/2016	1.000000	1.000000	0.976471	1.000000
31/07/2016	0.238039	0.131373	0.306863	0.002549
01/08/2016	0.078235	0.067843	0.124706	0.066471
02/08/2016	0.627647	0.548627	0.493137	0.013922
03/08/2016	0.446471	0.326667	0.213725	0.283725
08/08/2016	0.010980	0.000000	0.045882	0.000000
09/08/2016	0.338824	0.041569	0.034510	0.303529
11/08/2016	0.033922	0.045294	0.167647	0.065098
12/08/2016	0.500000	0.308235	0.361373	0.665294
13/08/2016	0.000000	0.000000	0.000000	0.000000
14/08/2016	0.300392	0.165294	0.187451	0.146078
21/08/2016	0.000000	0.000000	0.000000	0.000000
21/08/2016	0.945098	0.997059	0.660392	0.997059
22/08/2016	0.000000	0.000000	0.000000	0.000000
23/08/2016	0.000000	0.000000	0.000000	0.000000
24/08/2016	0.003529	0.000000	0.002549	0.000000

... continúa en la siguiente página

Día	LDA	DecisionTree	AdaBoost	NeuralNetwork
27/08/2016	0.030196	0.033529	0.032745	0.006667
28/08/2016	0.006667	0.000000	0.014118	0.000000
02/09/2016	0.919216	0.919216	0.918431	0.919216
05/09/2016	0.640980	0.679608	0.615882	0.431569
06/09/2016	0.000000	0.000000	0.000000	0.000000
08/09/2016	0.179804	0.369412	0.000000	0.158431
09/09/2016	0.367451	0.083333	0.270000	0.083333
10/09/2016	0.519412	0.592157	0.356078	0.592157
12/09/2016	0.032157	0.032157	0.036275	0.032157
13/09/2016	0.019608	0.000000	0.083333	0.000000
15/09/2016	0.266471	0.145490	0.836078	0.331961
16/09/2016	0.444118	0.460196	0.592157	0.603333
18/09/2016	0.197647	0.299608	0.247059	0.324314
19/09/2016	0.014314	0.004902	0.025098	0.004902
21/09/2016	0.924510	0.971373	0.997843	0.921373
22/09/2016	0.000000	0.000000	0.000000	0.000000
24/09/2016	0.916078	0.740980	0.741373	0.920000
27/09/2016	0.000980	0.000000	0.132549	0.319216
28/09/2016	0.071569	0.071176	0.072157	0.071176
30/09/2016	0.002745	0.000000	0.100588	0.019412
01/10/2016	0.286471	0.310588	0.265490	0.367255
03/10/2016	0.104314	0.005490	0.002157	0.392941
04/10/2016	0.450588	0.450588	0.391765	0.450588
05/10/2016	0.001176	0.041373	0.051176	0.000000
06/10/2016	0.520784	0.171765	0.194902	0.476471
08/10/2016	0.000000	0.006471	0.050784	0.007647
09/10/2016	0.566275	0.540980	0.686471	0.706667
11/10/2016	0.000000	0.000000	0.000000	0.006471
14/10/2016	0.000000	0.000000	0.138235	0.001373
16/10/2016	0.506667	0.318235	0.497647	0.556078
17/10/2016	0.321765	0.431961	0.253137	0.570196
18/10/2016	0.068627	0.497255	0.260392	0.470980
19/10/2016	0.464118	0.464118	0.464118	0.460196
20/10/2016	0.303333	0.303333	0.303333	0.283529
21/10/2016	0.250588	0.212745	0.249216	0.275098
23/10/2016	0.153529	0.197059	0.195490	0.311373
25/10/2016	0.275686	0.217647	0.087059	0.146667
27/10/2016	0.783922	0.271176	0.811765	0.750196
03/11/2016	0.002353	0.009216	0.002353	0.106863
05/11/2016	0.402353	0.400784	0.115686	0.282549
06/11/2016	0.114118	0.243725	0.353137	0.020980

... continúa en la siguiente página

Día	LDA	DecisionTree	AdaBoost	NeuralNetwork
07/11/2016	0.209804	0.215490	0.220196	0.610980
11/11/2016	0.213529	0.360392	0.050980	0.194510
13/11/2016	0.948824	0.805294	0.213137	0.488235
20/11/2016	0.038627	0.159804	0.472353	0.501765
23/11/2016	0.141569	0.507843	0.251373	0.458627
24/11/2016	0.084118	0.025882	0.003922	0.324118
27/11/2016	0.432353	0.524510	0.502353	0.506863
28/11/2016	0.891176	0.882549	0.882549	0.877059
30/11/2016	0.945490	0.865686	0.787647	0.868431
01/12/2016	0.003725	0.066275	0.004706	0.429412
04/12/2016	0.846078	0.720196	0.849804	0.541373
05/12/2016	0.224706	0.092353	0.457451	0.292745
06/12/2016	0.177843	0.215294	0.459216	0.495098
07/12/2016	0.168235	0.733725	0.788235	0.474902
09/12/2016	0.999020	0.500392	0.514314	0.414706
12/12/2016	0.698431	0.886667	0.714314	0.910784
14/12/2016	0.000000	0.009412	0.000000	0.210784
16/12/2016	0.309020	0.362941	0.303922	0.316667
17/12/2016	0.144118	0.200000	0.005686	0.020588
18/12/2016	0.685490	0.442353	0.750392	0.676863
22/12/2016	0.661569	0.529804	0.957059	0.894510
23/12/2016	0.617647	0.905294	0.820196	0.616667
24/12/2016	0.771765	0.875490	0.092745	0.667647
25/12/2016	0.602941	0.602941	0.627255	0.610392
27/12/2016	0.012549	0.001569	0.000000	0.058039
28/12/2016	0.000000	0.000000	0.000000	0.016471
29/12/2016	0.540196	0.332745	0.194510	0.577451
31/12/2016	0.313333	0.206863	0.518627	0.476667
01/01/2017	0.029804	0.001373	0.001961	0.388431

3.4.2. Rango

Tabla 8. MSE diario. Todas features sin RAIN

Día	RandomForest	MLP	NeuralNetwork	LoR
05/01/2016	8.881961	8.803333	8.838627	8.202353
08/01/2016	1.247255	1.247255	1.247255	1.247255
09/01/2016	2.109412	2.109412	2.109412	2.099020
10/01/2016	0.456275	0.456275	0.456275	0.393333
11/01/2016	3.407843	3.407843	3.407843	3.369608
13/01/2016	0.128824	0.118824	0.108627	0.045294

... continúa en la siguiente página

Día	LDA	DecisionTree	AdaBoost	NeuralNetwork
15/01/2016	2.812745	1.710196	1.391961	1.258824
18/01/2016	0.137451	0.137451	0.137451	0.107843
19/01/2016	1.727647	1.255686	1.235490	1.040784
21/01/2016	0.013725	0.013725	0.013725	0.017451
23/01/2016	1.187843	1.187843	1.187843	1.187843
24/01/2016	4.915098	4.915098	4.915098	4.915098
26/01/2016	0.140588	0.140588	0.140588	0.140588
27/01/2016	0.772157	0.772157	0.772157	0.772157
28/01/2016	0.019608	0.019608	0.019608	0.019608
30/01/2016	0.150196	0.150196	0.150196	0.100784
06/02/2016	0.411569	0.411569	0.411569	0.416667
11/02/2016	5.859216	5.859216	5.859216	5.740392
13/02/2016	5.201569	5.201569	5.201569	5.201569
16/02/2016	2.776078	1.289412	1.287255	1.348431
19/02/2016	1.207451	0.436471	0.414902	0.731961
20/02/2016	0.056471	0.056667	0.062157	0.103725
22/02/2016	0.000392	0.000392	0.000392	0.000392
25/02/2016	3.368431	3.137451	3.258431	3.196078
27/02/2016	2.104902	1.016275	1.073529	1.008039
29/02/2016	3.470588	1.597647	1.639608	1.644902
01/03/2016	0.749412	0.425098	0.366275	0.327059
02/03/2016	0.034902	0.034902	0.034902	0.099412
03/03/2016	1.132353	0.965294	0.775686	0.525294
04/03/2016	1.195294	1.195294	1.195294	0.854706
06/03/2016	1.523137	0.570392	0.569412	0.774902
09/03/2016	0.640784	0.401765	0.410000	0.216667
10/03/2016	4.617843	2.596275	2.580588	2.451373
12/03/2016	0.822549	0.438235	0.325490	0.327647
14/03/2016	0.206275	0.210000	0.207059	0.241176
17/03/2016	0.582353	0.228824	0.207451	0.195490
18/03/2016	0.120784	0.100784	0.126863	0.188627
23/03/2016	4.068039	4.046275	4.100980	4.152941
28/03/2016	0.000000	0.000000	0.000000	0.011765
29/03/2016	1.363529	1.363529	1.363529	1.363529
30/03/2016	0.869804	0.869804	0.869804	0.869804
03/04/2016	0.027059	0.027059	0.027059	0.027059
06/04/2016	0.652745	0.652549	0.636863	0.477451
08/04/2016	1.054314	1.183529	1.184314	1.337255
10/04/2016	0.000000	0.000000	0.000000	0.036667
11/04/2016	8.015294	7.139608	7.411176	6.939412
12/04/2016	0.309020	0.292549	0.308627	0.259804

... continúa en la siguiente página

Día	LDA	DecisionTree	AdaBoost	NeuralNetwork
13/04/2016	4.276863	4.252353	4.276863	4.026275
15/04/2016	0.000000	0.000000	0.000000	0.000000
17/04/2016	0.001765	0.300588	0.333725	0.476863
19/04/2016	0.391373	0.391373	0.391373	0.391373
21/04/2016	3.223529	3.223529	3.228039	3.221569
23/04/2016	8.724706	8.724706	8.724706	8.723725
25/04/2016	0.000000	0.056471	0.057843	0.084510
27/04/2016	0.130000	0.094510	0.133922	0.321765
28/04/2016	0.075882	0.037059	0.030784	0.003137
30/04/2016	0.331765	0.330784	0.331765	0.288039
01/05/2016	0.771176	0.999020	1.290196	1.319020
03/05/2016	0.000000	0.000000	0.000000	0.000980
07/05/2016	0.000000	0.000000	0.000000	0.000000
08/05/2016	7.340588	7.340588	7.340588	7.340588
09/05/2016	7.081176	7.081176	7.081176	7.081176
10/05/2016	2.539412	2.539412	2.539412	2.539412
12/05/2016	0.871176	0.871176	0.871176	0.871176
13/05/2016	6.250392	6.250392	6.250392	6.238627
17/05/2016	0.000000	0.000000	0.000000	0.000784
18/05/2016	0.930980	0.930980	0.930980	0.930980
19/05/2016	0.090392	0.090392	0.090392	0.077451
20/05/2016	1.000000	1.000000	1.000000	1.000000
21/05/2016	0.063137	0.063137	0.063137	0.063137
23/05/2016	0.000000	0.000000	0.000000	0.027451
24/05/2016	0.000000	0.000000	0.000000	0.000000
26/05/2016	0.000000	0.000000	0.000000	0.000000
29/05/2016	0.000000	0.000000	0.000000	0.000000
31/05/2016	1.316863	1.316863	1.316863	1.316863
01/06/2016	0.000000	0.000000	0.000000	0.003333
06/06/2016	0.000000	0.000000	0.000000	0.000000
08/06/2016	0.107451	0.107451	0.107451	0.107451
14/06/2016	4.378235	4.378235	4.378235	4.378235
18/06/2016	0.000000	0.000000	0.000000	0.007647
19/06/2016	1.882549	0.583922	0.581569	0.904902
20/06/2016	0.000000	0.000000	0.000000	0.000000
22/06/2016	0.000000	0.000000	0.000000	0.000000
23/06/2016	0.662549	0.662549	0.662549	0.662549
24/06/2016	0.000000	0.000000	0.000000	0.000000
26/06/2016	1.000000	1.000000	1.000000	1.000000
28/06/2016	0.000000	0.000000	0.000000	0.000000
29/06/2016	0.800784	0.800784	0.800784	0.800784

... continúa en la siguiente página

Día	LDA	DecisionTree	AdaBoost	NeuralNetwork
02/07/2016	0.000000	0.000000	0.000000	0.000000
03/07/2016	0.000000	0.000000	0.000000	0.000000
06/07/2016	0.000000	0.000000	0.000000	0.000000
07/07/2016	5.954118	5.954118	5.954118	5.954118
08/07/2016	0.000000	0.000000	0.000000	0.000000
09/07/2016	8.375294	8.375294	8.375294	8.375294
10/07/2016	7.524706	7.524706	7.524706	7.524706
11/07/2016	0.000000	0.000000	0.000000	0.000000
13/07/2016	0.405294	0.405294	0.405294	0.405294
15/07/2016	0.000000	0.000000	0.000000	0.000000
16/07/2016	0.000000	0.000000	0.000000	0.000000
17/07/2016	0.000000	0.000000	0.000000	0.000000
19/07/2016	0.337843	0.337843	0.337843	0.337843
21/07/2016	0.910392	0.900000	0.882941	0.800784
23/07/2016	0.000000	0.000000	0.000000	0.000000
26/07/2016	0.422157	0.422157	0.422157	0.422157
29/07/2016	0.727059	0.727059	0.727059	0.727059
30/07/2016	8.344314	8.344314	8.344314	8.344314
31/07/2016	0.000000	0.000000	0.000000	0.000000
01/08/2016	0.121765	0.121765	0.121765	0.121765
02/08/2016	0.000000	0.000000	0.000000	0.000000
03/08/2016	1.091961	1.091961	0.995294	1.007255
08/08/2016	0.000000	0.000000	0.000000	0.000000
09/08/2016	0.851765	0.851765	0.851765	0.850980
11/08/2016	0.038431	0.033725	0.035882	0.036471
12/08/2016	5.987647	5.987647	5.987647	5.987647
13/08/2016	0.000000	0.000000	0.000000	0.000000
14/08/2016	0.341765	0.341765	0.341765	0.341765
21/08/2016	0.000000	0.000000	0.000000	0.000000
21/08/2016	5.580784	5.580784	5.580784	5.580784
22/08/2016	0.000000	0.000000	0.000000	0.000000
23/08/2016	0.000000	0.000000	0.000000	0.000000
24/08/2016	0.000000	0.000000	0.000000	0.000000
27/08/2016	0.000000	0.000000	0.000000	0.000000
28/08/2016	0.000000	0.000000	0.000000	0.000000
02/09/2016	0.919216	0.919216	0.919216	0.919216
05/09/2016	0.000000	0.000000	0.000000	0.000000
06/09/2016	0.000000	0.000000	0.000000	0.000000
08/09/2016	0.000000	0.000000	0.000000	0.000000
09/09/2016	0.083333	0.083333	0.083333	0.083333
10/09/2016	0.592157	0.592157	0.592157	0.592157

... continúa en la siguiente página

Día	LDA	DecisionTree	AdaBoost	NeuralNetwork
12/09/2016	0.250196	0.250196	0.250196	0.250196
13/09/2016	0.000000	0.000000	0.000000	0.000000
15/09/2016	0.270980	0.270980	0.270980	0.270980
16/09/2016	0.911961	1.065294	1.055294	0.990784
18/09/2016	1.496471	1.496471	1.496471	1.496471
19/09/2016	0.004902	0.004902	0.004902	0.004902
21/09/2016	0.924510	0.924510	0.924510	0.924510
22/09/2016	0.000000	0.000000	0.000000	0.000000
24/09/2016	0.920000	0.920000	0.920000	0.920000
27/09/2016	0.000000	0.199216	0.377647	0.444314
28/09/2016	0.071176	0.071176	0.071176	0.071176
30/09/2016	0.000000	0.000000	0.000000	0.000000
01/10/2016	1.099216	1.099216	1.099216	1.099216
03/10/2016	7.814314	7.814314	7.814314	7.814314
04/10/2016	0.450588	0.450588	0.450588	0.450588
05/10/2016	0.000000	0.000000	0.000000	0.000000
06/10/2016	1.836078	1.836078	1.836078	1.835882
08/10/2016	0.000000	0.000000	0.000000	0.000000
09/10/2016	6.294706	6.294706	6.294706	6.294706
11/10/2016	0.000000	0.000000	0.000000	0.000000
14/10/2016	0.178627	0.178627	0.178627	0.178627
16/10/2016	0.192745	0.348627	0.403529	0.628824
17/10/2016	4.946471	4.946471	4.942549	4.908235
18/10/2016	2.401765	2.401765	2.401765	2.401765
19/10/2016	4.177059	4.177059	4.177059	4.177059
20/10/2016	2.730000	2.730000	2.730000	2.716275
21/10/2016	0.199020	0.246667	0.268627	0.322157
23/10/2016	2.842941	2.842941	2.842941	2.842941
25/10/2016	0.000000	0.000000	0.000000	0.000000
27/10/2016	0.000000	0.494902	0.503529	0.789216
03/11/2016	0.002353	0.002353	0.002353	0.002353
05/11/2016	0.000000	0.000000	0.000000	0.000000
06/11/2016	5.917059	4.936667	5.149412	4.639608
07/11/2016	0.045098	0.209412	0.214510	0.517647
11/11/2016	0.000000	0.000000	0.000000	0.095294
13/11/2016	8.657647	8.008627	7.549804	7.026275
20/11/2016	0.001765	0.001765	0.001765	0.023333
23/11/2016	8.968235	8.989216	8.995098	8.307843
24/11/2016	9.000000	8.865686	8.632353	7.479412
27/11/2016	4.979608	4.979608	4.979608	4.977647
28/11/2016	1.939608	1.939608	1.939608	1.939608

... continúa en la siguiente página

Día	LDA	DecisionTree	AdaBoost	NeuralNetwork
30/11/2016	1.714118	1.714118	1.714118	1.751373
01/12/2016	0.000000	0.000000	0.000000	0.008039
04/12/2016	0.923725	0.923725	0.923725	0.919020
05/12/2016	1.629608	1.629608	1.629608	1.641765
06/12/2016	0.363922	0.363922	0.349216	0.272549
07/12/2016	0.126275	0.126275	0.126275	0.126275
09/12/2016	0.414706	0.414706	0.414706	0.414706
12/12/2016	6.225882	6.225882	6.225882	6.225882
14/12/2016	0.000000	0.000000	0.000000	0.018431
16/12/2016	0.000000	0.002353	0.000000	0.070392
17/12/2016	5.175882	5.175882	5.175882	5.079804
18/12/2016	2.127647	2.291961	2.353922	1.748627
22/12/2016	3.049412	3.049412	3.049412	3.049608
23/12/2016	0.082157	0.082157	0.090392	0.082157
24/12/2016	7.044706	7.044706	6.985882	6.702353
25/12/2016	0.602941	0.602941	0.602941	0.600784
27/12/2016	0.220588	0.220588	0.220588	0.220588
28/12/2016	0.000000	0.000000	0.000000	0.000000
29/12/2016	0.124314	0.121961	0.124510	0.143137
31/12/2016	0.343529	0.396863	0.447059	0.385882
01/01/2017	0.000000	0.000000	0.000000	0.022745

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.

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