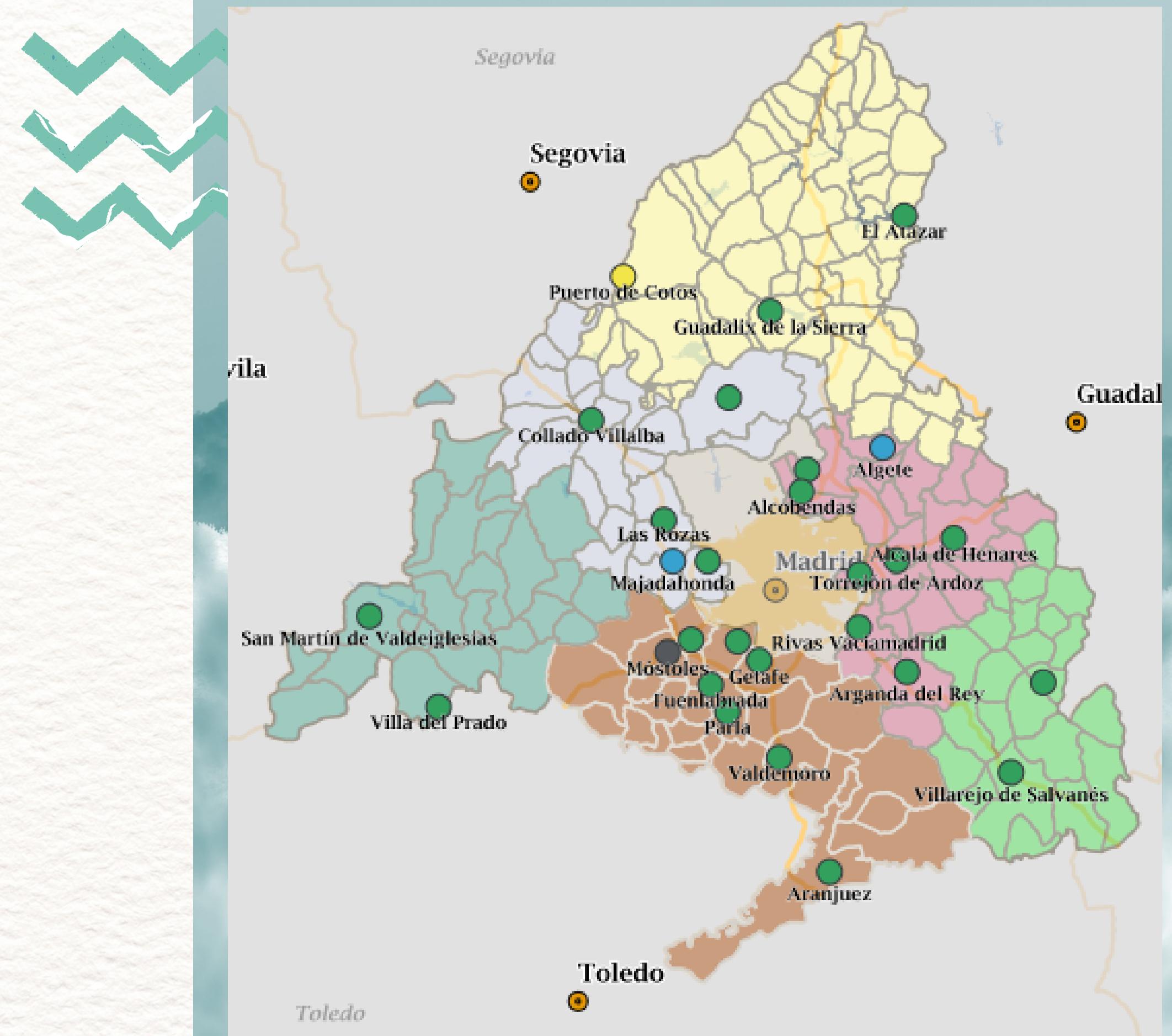


Red de Calidad del Aire Datos Meterologicos. Datos del año en curso

Nombre: Alvaro Steven Mejia Ortiz

¿Que es?

- Conjunto de datos horarios medidos de forma automática en las estaciones de la Red de Calidad del Aire de la Comunidad de Madrid durante el año en curso



Objetivo

El objetivo principal de este proyecto es desarrollar modelos de aprendizaje automático capaces de predecir con precisión las magnitudes meteorológicas futuras en una región específica.



Cantidad de datos

45002 REGISTROS



Variables

56

Tabla 1. ESTRUCTURA DEL ARCHIVO

Provincia	Municipio	Estación en_municipio	Magnitud	Punto_ de_muestreo	año	mes	día	h01	v01	h02..	v02..
28	5	2	1	28005002_1_38	2020	1	1	3	V	7	V

Tabla 1. ESTRUCTURA DEL ARCHIVO

Provincia	Municipio	Estación en_municipio	Magnitud	Punto_de_muestreo	año	mes	día	h01	v01	h02..	v02..
28	5	2	1	28005002_1_38	2020	1	1	3	V	7	V

El código 28 es el que corresponde a la provincia de Madrid.

El código del campo “Municipio” se corresponde con “código municipio” de la “Tabla 2. CÓDIGOS DE ESTACIONES”.

El código del campo “Estación en_municipio” es un correlativo que asigna el ministerio con competencias en medio ambiente y sirve para identificar las posibles ubicaciones de las estaciones en ese municipio.

El código “Magnitud” hace referencia al parámetro medido y se corresponde con el reflejado en la “Tabla 3. MAGNITUDES, UNIDADES y TÉCNICAS DE MEDIDA”.

El campo punto de muestreo incluye el código de la estación completo (provincia, municipio y estación) más la magnitud y la técnica de medida. Esta última se describe en la “Tabla 3. MAGNITUDES, UNIDADES y TÉCNICAS DE MEDIDA”.

Los campos “año”, “mes” y “día” hacen referencia respectivamente al año, mes y día de la medida.

El valor que aparece en el campo “h” indica la concentración medida de la magnitud correspondiente. “h01” hace referencia a la 1 de la mañana, “h02” a las 2 de la mañana y así sucesivamente hasta las 24 horas.

“v01” se refiere a la validación de la hora 1, “v02” a la validación de la hora 2 y así hasta las 24 h.

El código “T” hace referencia a que es un dato Temporal, pendiente de la validación por parte de los técnicos de la Consejería.

Una vez validados, el código “V” hace referencia a que es un dato Válido y el código “N” hace referencia a que es un dato No válido.

CÓDIGO NACIONAL	CÓDIGO MUNICIPIO	NOMBRE ESTACIÓN
28005002	5	ALCALÀ DE HENARES
28006004	6	ALCOBENDAS
28007004	7	ALCORCÒN
28009001	9	ALGETE
28013002	13	ARANJUEZ
28014002	14	ARGANDA DEL REY
28016001	16	EL ATAZAR
28045002	45	COLMENAR VIEJO
28047002	47	COLLADO VILLALBA
28049003	49	COSLADA
28058004	58	FUENLABRADA
28065014	65	GETAFE
28067001	67	GUADALIX DE LA SIERRA
28074007	74	LEGANÈS
28080003	80	MAJADAHONDA
28092005	92	MÓSTOLES
28102001	102	ORUSCO DE TAJUÑA
28120001	120	PUERTO DE COTOS
28123002	123	RIVAS-VACIAMADRID
28133002	133	SAN MARTÍN DE VALDEIGLESIAS
28148004	148	TORREJÓN DE ARDOZ
28161001	161	VALDEMORO
28171001	171	VILLA DEL PRADO
28180001	180	VILLAREJO DE SALVANÉS

CÓDIGO MAGNITUD	DESCRIPCIÓN MAGNITUD	CODIGO DE TÉCNICA DE MEDIDA	UNIDAD	DESCRIPCIÓN UNIDAD
81	Velocidad del viento	89	m/s	metros por segundo
82	Dirección del viento	89	Grd	grados
83	Temperatura	89	°c	grados centígrados
86	Humedad relativa	89	%	porcentaje
87	Presión atmosférica	89	mbar	milibar
88	Radiación solar	89	W/m ²	vatios por metro cuadrado
89	Precipitación	89	l/m ²	litros por metro cuadrado

	provincia	municipio	estacion	magnitud	ano	mes	dia	h01	h02	h03	...	h15	h16	h17	h18	h19	h20
0	28	102	1	81	2023	1	1	1.0	1.0	1.0	...	3.0	2.0	3.0	3.0	3.0	2.0
1	28	102	1	82	2023	1	1	173.0	138.0	136.0	...	171.0	164.0	159.0	97.0	127.0	173.0
2	28	102	1	83	2023	1	1	7.0	7.0	6.0	...	14.0	14.0	13.0	12.0	11.0	10.0
3	28	102	1	86	2023	1	1	50.0	51.0	50.0	...	35.0	33.0	33.0	34.0	38.0	40.0
4	28	102	1	87	2023	1	1	932.0	932.0	932.0	...	928.0	928.0	928.0	928.0	928.0	928.0

5 rows × 31 columns

Random Forest Model

```
Random Forest Model

1 from sklearn.ensemble import RandomForestClassifier
2
3 features = df[['h01', 'h02', 'h03', 'h04', 'h05', 'h06', 'h07', 'h08', 'h09', 'h10', 'h11', 'h12', 'h13', 'h14', 'h15', 'h16', 'h17', 'h18', 'h19', 'h20', 'h21', 'h22', 'h23', 'h24', 'h25', 'h26', 'h27', 'h28', 'h29', 'h30', 'h31', 'h32', 'h33', 'h34', 'h35', 'h36', 'h37', 'h38', 'h39', 'h40', 'h41', 'h42', 'h43', 'h44', 'h45', 'h46', 'h47', 'h48', 'h49', 'h50', 'h51', 'h52', 'h53', 'h54', 'h55', 'h56', 'h57', 'h58', 'h59', 'h60', 'h61', 'h62', 'h63', 'h64', 'h65', 'h66', 'h67', 'h68', 'h69', 'h70', 'h71', 'h72', 'h73', 'h74', 'h75', 'h76', 'h77', 'h78', 'h79', 'h80', 'h81', 'h82', 'h83', 'h84', 'h85', 'h86', 'h87', 'h88', 'h89', 'h90', 'h91', 'h92', 'h93', 'h94', 'h95', 'h96', 'h97', 'h98', 'h99', 'h100']]
4 target = df['magnitud']
5
6 df['municipio'] = df['provincia'].astype(str) + df['municipio'].astype(str)
7
8 features_normalized = preprocessing.scale(features)
9 imputer = SimpleImputer(strategy='mean')
10 features_imputed = imputer.fit_transform(features_normalized)
11
12 group_kfold = GroupKFold(n_splits=5)
13
14 rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
15
16 predictions = cross_val_predict(rf_model, features_imputed, target, cv=group_kfold.split(features_imputed, target, df['municipio']))
17
18 accuracy = accuracy_score(target, predictions)
19 print(f'Precisión del modelo en el conjunto de datos completo: {accuracy}')
20
21 print('\nResultados de la validación cruzada:')
22 for i, score in enumerate(cross_val_score(rf_model, features_imputed, target, cv=group_kfold.split(features_imputed, target, df['municipio']), scoring='accuracy')):
23     print(f'Fold {i + 1}: {score}')
24
25 average_accuracy = cross_val_score(rf_model, features_imputed, target, cv=group_kfold.split(features_imputed, target, df['municipio']))
26 print(f'\nPrecisión media: {average_accuracy}')
27

[  ] Precisión del modelo en el conjunto de datos completo: 0.981778182707051

Resultados de la validación cruzada:
Fold 1: 0.9895732559389444
Fold 2: 0.987419165196943
Fold 3: 0.9861258083480305
Fold 4: 0.973686969283992
Fold 5: 0.9730025139359493

Precisión media: 0.9819615425407718
```

Random Forest Model

```
▶ Predicción para el 1 de septiembre en la Ciudad 5:  
➡ Hora 01: 82.0  
➡ Hora 02: 82.0  
➡ Hora 03: 82.0  
➡ Hora 04: 82.0  
➡ Hora 05: 82.0  
➡ Hora 06: 82.0  
➡ Hora 07: 82.0  
➡ Hora 08: 82.0  
➡ Hora 09: 82.0  
➡ Hora 10: 82.0  
➡ Hora 11: 82.0  
➡ Hora 12: 82.0  
➡ Hora 13: 82.0  
➡ Hora 14: 82.0  
➡ Hora 15: 82.0  
➡ Hora 16: 82.0  
➡ Hora 17: 82.0  
➡ Hora 18: 82.0  
➡ Hora 19: 82.0  
➡ Hora 20: 82.0  
➡ Hora 21: 82.0  
➡ Hora 22: 82.0  
➡ Hora 23: 82.0  
➡ Hora 24: 82.0  
Predicción para el 2 de septiembre en la Ciudad 5:  
Hora 01: 82.0
```

¡Gracias!