



Calidad del aire en Valencia



Mikel Navarro - Tono Villarejo - Santiago Cámara
Proyecto Inteligencia Artificial y Big Data
9 de junio de 2022





Presentación y objetivos

Proyecto Peaky Blinders



Los datos de calidad del aire promueven estudios que predicen el bienestar en las ciudades.

¿Que habría que cambiar? ¿Que ha de mejorar?



Presentación y objetivos

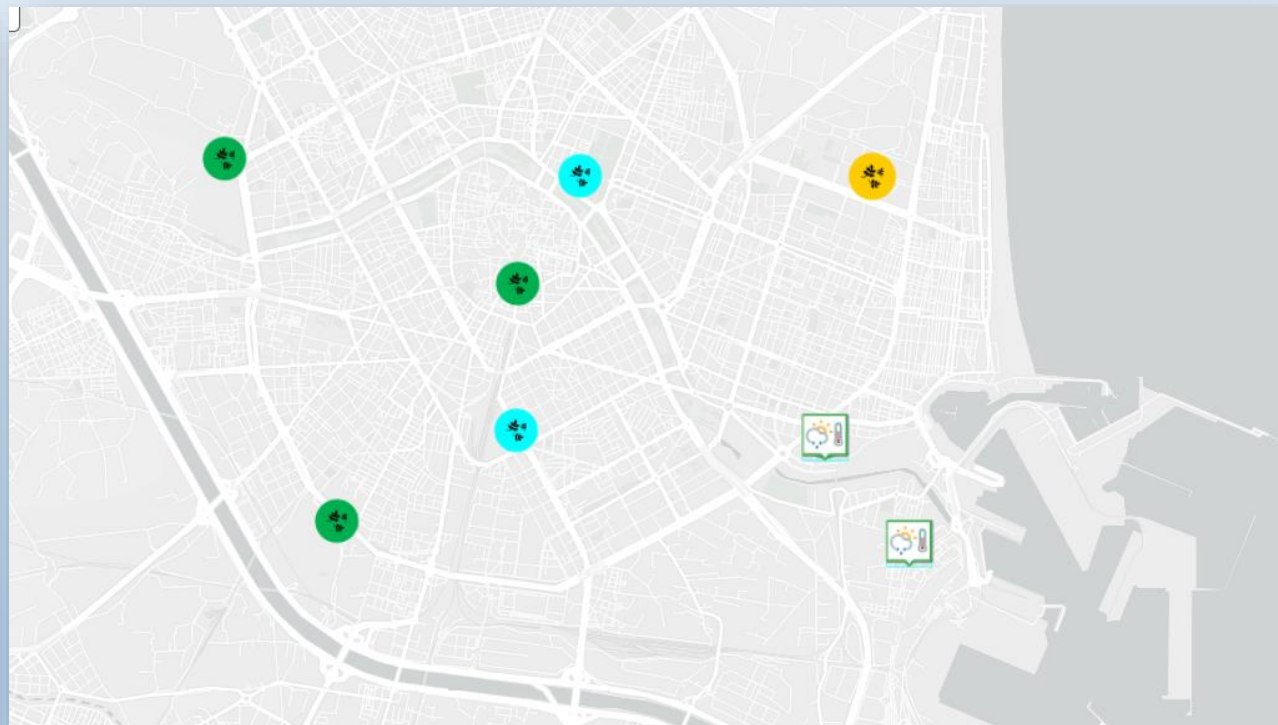
Objetivos:

- Base de datos histórica con indicadores de calidad del aire e información meteorológica.
- Estudio especial de los años 2020 y 2021 (pandemia).
- Módulo para predecir indicadores de calidad del aire a futuro.
- Visualización de datos en una página web creada a tal efecto.
- Hacer una web de datos abiertos para poner a disposición del público los dataset elaborados en el proyecto.



Presentación y objetivos

Estaciones de medición y control de la calidad del aire





Presentación y objetivos

Los 5 indicadores principales de calidad del aire^(*)

- Partículas en suspensión (**PM10**) menores de 10 micrómetros (μm)
 - Partículas en suspensión (**PM2.5**) menores de 2.5 micrómetros (μm)
 - Ozono troposférico (**O₃**)
 - Dióxido de nitrógeno (**NO₂**)
 - Dióxido de azufre (**SO₂**)
- } microgramos/metro cúbico ($\mu\text{g}/\text{m}^3$)

() Fuente: Ministerio para la Transición Ecológica*



Presentación y objetivos

Otros indicadores de calidad del aire y meteorológicos disponibles

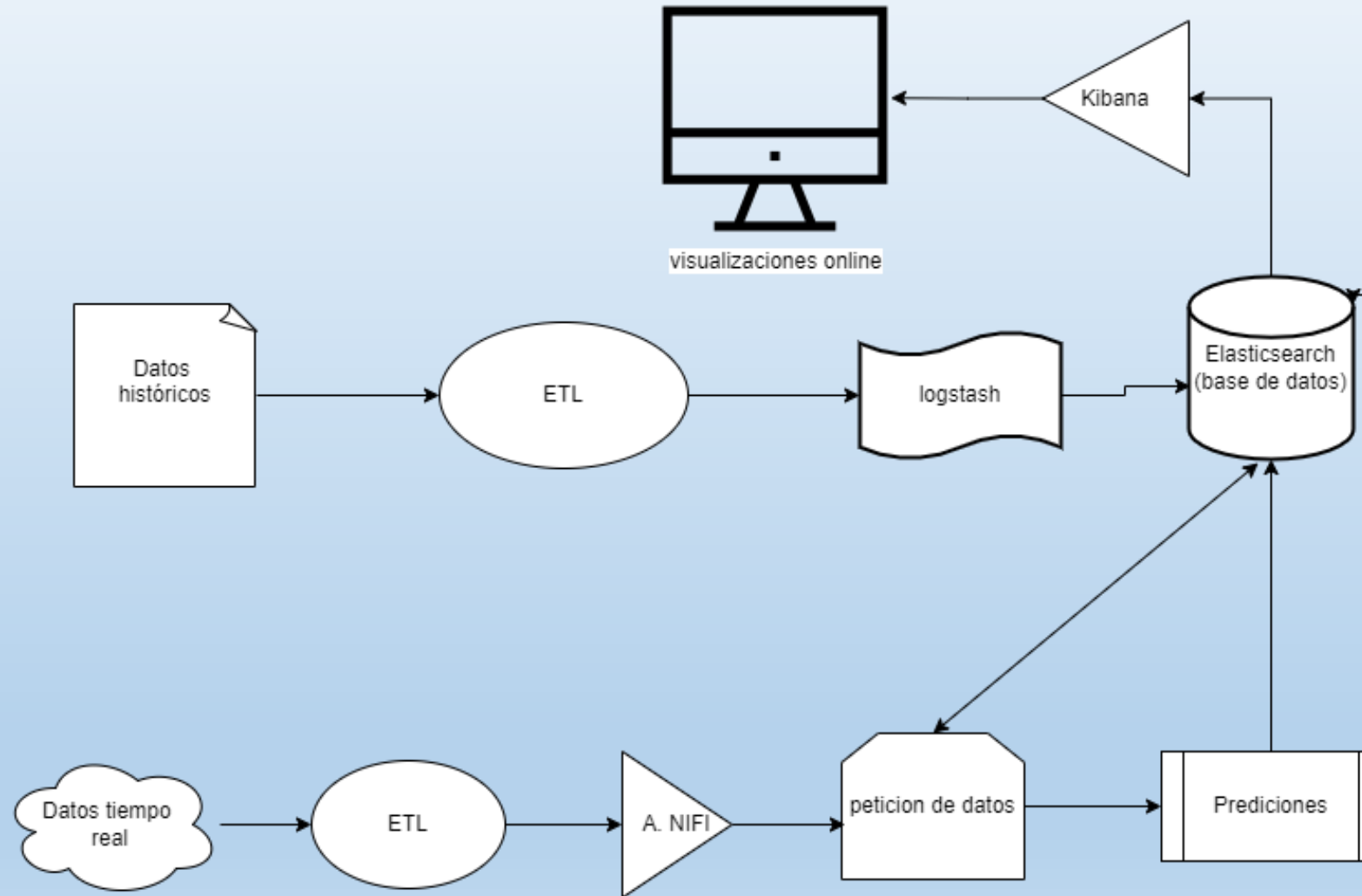
- PM1: Partículas menores de 1 micrómetro (μm)
- NO: Óxido de nitrógeno
- NOx: Óxidos de nitrógeno
- CO: monóxido de carbono
- NH3: amoníaco
- C7H8: tolueno
- C6H6: Benzol o benceno
- C8H10: xileno
- Ruido
- Velocidad viento
- Dirección del viento
- Temperatura
- Humedad relativa
- Presión
- Radiación Solar
- Precipitación
- Velocidad Máxima del viento



DISEÑO DEL PROYECTO



Diseño del proyecto





Herramientas y soluciones tecnológicas

- **Big Data**
 - Elasticsearch
 - Logstash
 - Kibana
- **Modelos Aprendizaje Automático**
 - Scripts en lenguaje Python
 - Biblioteca Tensorflow
 - Diversas bibliotecas complementarias para python: pandas, numpy, matplotlib, dataprep, datetime, etc.



Requerimientos de hardware

- **Servidor IES Abastos**
 - Instalación de Elasticsearch, Logstash, Kibana
- **2 Servidores VPS**
 - Hospedaje del portal de acceso a los conjuntos de datos procesados en el proyecto y para las visualizaciones



DESARROLLO DEL PROYECTO



Desarrollo del proyecto

Resumen del reparto de tareas

| Tono | Mikel | Santi |
|--|--|---|
| <ul style="list-style-type: none">• Instalación y configuración de aplicaciones en servidor.• Carga de datos en B.D.• Visualizaciones Kibana | <ul style="list-style-type: none">• Ingesta de datos - EDA y ETL• Captura datos tiempo real• Framework CKAN• Web del proyecto | <ul style="list-style-type: none">• Colaboración con EDA y ETL• Selección datos• Algoritmos de Aprendizaje automático |

Tareas compartidas: Investigar y analizar fuentes de datos



Desarrollo del proyecto

Ingesta de Datos. Fuentes de datos Organismos Oficiales

CALIDAD DEL AIRE

Portal de datos abiertos
de la Generalitat
Valenciana

<https://portaldadesobertes.gva.es/es>



Conselleria de
agricultura, desarrollo
rural, emergencia
climática y transición
ecológica

<https://agroambient.gva.es/>





Desarrollo del proyecto

Ingesta de Datos. Fuentes de datos. Portales especializados

CALIDAD DEL AIRE

| | | |
|---|---|--|
| Portal medición calidad del aire en el mundo (fork) | https://waqi.info/es/ |  |
| Portal de calidad del aire en el mundo | https://aqicn.org/map/world/es/ |  |



Desarrollo del proyecto

Ingesta de Datos. Fuentes de datos Organismos Oficiales

METEOROLÓGICOS

Agencia Estatal de
Meteorología

<http://www.aemet.es/>





Desarrollo del proyecto

Ingesta de Datos. Fuentes de datos Portales especializados

METEOROLÓGICOS

Base de datos
Meteorológica

<https://datosclima.es/>



Associació
Valenciana de
meteorologia
Josep Peinado

<https://www.avamet.org/>





EDA (Exploratory Data Analysis)

- Inicialmente se trataron archivos de tipo txt, filtrando la ciudad de Valencia.
- Se pusieron de manifiesto múltiples problemas de formato al tener datos diferentes según la estación y el año a considerar.
- Se detectaron inconsistencias que forzaron empezar de nuevo el tratamiento de datos con archivos de tipo csv (datos abiertos GVA)



Desarrollo del proyecto

EDA (Exploratory Data Analysis)

- Se disponía de datos desde 1994 hasta la actualidad, según de que estación se trataba.
- Muchas estaciones carecían de información meteorológica o era incompleta, se debía conseguir de otra fuente.
- La elección de la fuente de datos meteorológicos se basó en que era gratuita y se disponía de datos desde 2013.



Desarrollo del proyecto

ETL (Extract Transform and Load)

- Se crearon scripts en Python para fusionar los datos de clima con los de calidad del aire.
- La prioridad ha sido siempre los datos facilitados por las estaciones de calidad de aire, complementando nulos con datos de clima.
- Después de la fusión de datos se comprueba que hay altos porcentajes de datos faltantes en algunas estaciones y para algunos indicadores.



Desarrollo del proyecto

ETL (Extract Transform and Load)

- Para poder rellenar huecos por datos faltantes se opta por realizar un proceso de interpolación, que estima los datos ausentes considerando los valores conocidos en el contexto de las fechas anteriores y posteriores de los huecos.
- Finalmente se obtuvo un dataset con la información disponible de todas las estaciones de Valencia en cuanto a calidad del aire y clima.



Desarrollo del proyecto

Big Data. Almacenamiento de datos.

- Se utiliza Elasticsearch como base de datos para almacenar los dataset necesarios y poder realizar visualizaciones con Kibana
- La carga de datos en Elasticsearch, desde los archivos csv que se han creado, se efectúa con Logstash



Desarrollo del proyecto

Big Data. Almacenamiento de datos.

- Se crean 2 grupos de índices en la Base de datos:
 - Datos históricos de todas las estaciones^(*)
 - Predicciones por estación e indicador^(*)

^() Cada estación tiene su propio índice*



Desarrollo del proyecto

Big Data. Visualización de la información.

- Debido a la imposibilidad de conectar Tableau con Elasticsearch y la gran dificultad de conectar con Power BI, se optó por usar Kibana.
- Kibana está perfectamente integrado con Elasticsearch



Desarrollo del proyecto

Big Data. Visualización de la información.

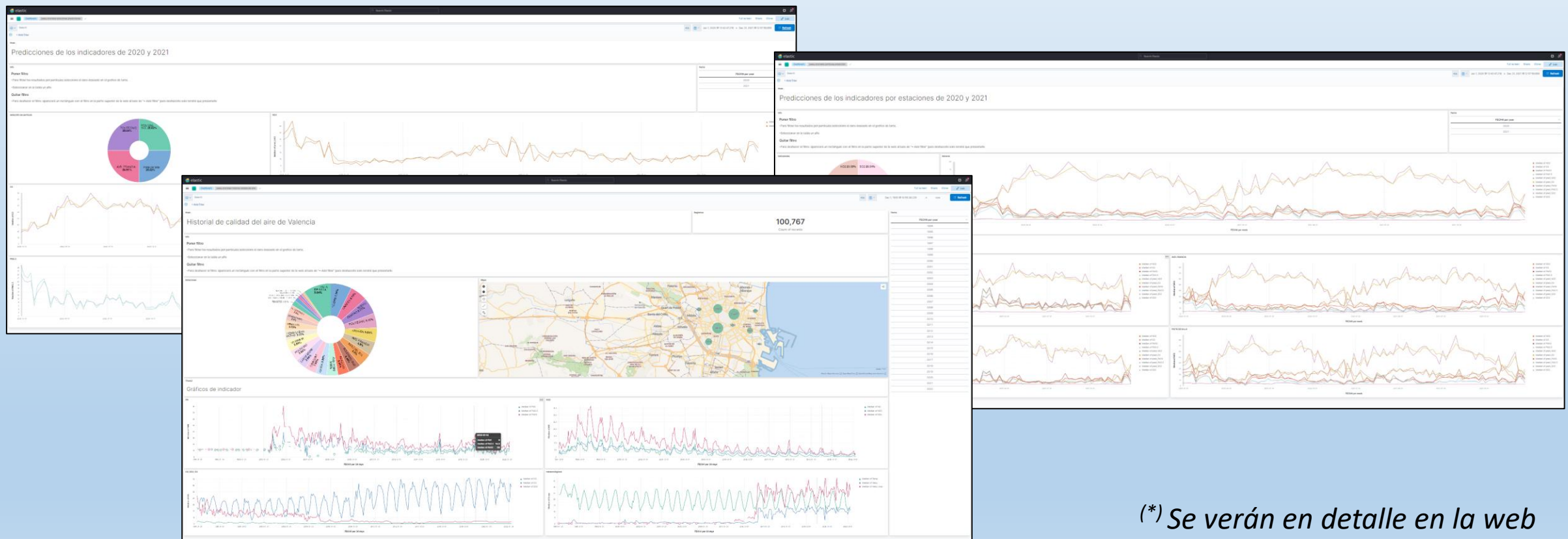
- Se crean 6 index patterns (enlaces Kibana-Elastic):
 - Datos históricos de todas las estaciones.
 - Predicciones de cada estación principal (4).
 - Predicciones de todas las estaciones.



Desarrollo del proyecto

Big Data. Visualización de la información.

Ejemplos de dashboard(*)



(*) Se verán en detalle en la web



Desarrollo del proyecto

Machine Learning. Selección de datos.

- Recopilados datos de 27 estaciones de Valencia, pero con una calidad y cantidad de información muy desigual.
- Cada estación necesitará su propio modelo de aprendizaje automático.
- Para afrontar los entrenamientos de los modelos se requiere hacer una selección de las estaciones con la mayor calidad/cantidad de datos posible.



Desarrollo del proyecto

Machine Learning. Selección de datos.

| Nº de registros totales y PORCENTAJE DE atributos nulos por estación: | | | | | | | | | | | | | | | | |
|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--|
| ===NOM_ESTACION=== | Nº reg | =SO2== | =CO== | =NO== | =NO2== | =NOx== | =O3== | =PM10= | PM2.5= | =PM1== | Preci= | =Temp= | Veloc= | Vemax= | TOTAL= | |
| CEMENTERIO | 2922 | 12.66 | 97.09 | 92.98 | 92.98 | 92.98 | 94.73 | 94.83 | 97.13 | 99.59 | 93.57 | 93.63 | 93.63 | 95.21 | 88.54 | |
| LABORATORI | 3984 | 10.22 | 98.74 | 97.72 | 97.72 | 97.72 | 98.37 | 98.87 | 98.95 | 99.90 | 98.49 | 98.57 | 98.49 | 98.80 | 91.73 | |
| VIVERS-M | 3984 | 12.07 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 93.24 | |
| POLIGONO | 3984 | 9.69 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 93.05 | |
| ARAGÓ-M | 3984 | 9.89 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 93.07 | |
| C/SAGUNT | 3984 | 6.50 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 92.81 | |
| AVD. PORT | 3984 | 8.58 | 87.93 | 88.10 | 88.10 | 88.10 | 88.25 | 100.00 | 100.00 | 100.00 | 100.00 | 87.85 | 87.85 | 100.00 | 86.52 | |
| PISTA DE SILLA | 9919 | 4.00 | 4.48 | 3.40 | 3.40 | 3.40 | 4.74 | 58.12 | 71.94 | 85.44 | 69.74 | 2.02 | 2.02 | 66.56 | 29.17 | |
| NUEVO CENTRO | 5810 | 2.05 | 2.01 | 2.58 | 2.58 | 2.58 | 4.54 | 99.52 | 99.52 | 99.52 | 100.00 | 87.61 | 87.61 | 100.00 | 53.09 | |
| GRAN VIA | 3253 | 7.41 | 7.10 | 10.05 | 12.88 | 10.05 | 19.40 | 99.97 | 99.97 | 99.97 | 99.97 | 99.97 | 99.97 | 99.97 | 58.98 | |
| ARAGÓN | 5079 | 2.60 | 25.20 | 8.49 | 8.51 | 7.95 | 2.48 | 95.31 | 95.31 | 95.31 | 99.98 | 99.98 | 99.98 | 99.98 | 57.01 | |
| LINARES | 6175 | 3.42 | 4.71 | 14.38 | 14.20 | 14.36 | 3.69 | 95.34 | 97.80 | 99.53 | 99.98 | 99.21 | 99.21 | 99.98 | 57.37 | |
| AYORA | 3984 | 16.82 | 85.54 | 86.90 | 86.90 | 86.90 | 85.34 | 98.29 | 98.67 | 99.67 | 98.67 | 98.09 | 98.09 | 98.57 | 87.57 | |
| N. CENTRO-2 | 2922 | 11.26 | 95.38 | 95.31 | 95.31 | 95.31 | 95.24 | 96.89 | 97.57 | 98.39 | 98.25 | 96.89 | 96.89 | 97.98 | 90.05 | |
| NAZARET | 731 | 24.76 | 35.70 | 31.19 | 31.19 | 31.19 | 32.56 | 67.99 | 67.99 | 67.99 | 100.00 | 100.00 | 100.00 | 100.00 | 60.81 | |
| TENDETES | 1827 | 4.98 | 91.79 | 91.95 | 91.95 | 91.95 | 91.13 | 94.03 | 97.81 | 98.47 | 97.26 | 97.10 | 97.10 | 100.00 | 88.12 | |
| NAZARET-MET. | 731 | 78.93 | 94.66 | 93.57 | 93.57 | 93.57 | 93.43 | 95.08 | 96.17 | 97.26 | 41.04 | 40.22 | 40.22 | 96.17 | 81.07 | |
| VIVERS | 7396 | 2.33 | 39.82 | 6.79 | 6.79 | 6.80 | 3.66 | 10.57 | 34.28 | 90.89 | 54.83 | 52.43 | 52.43 | 54.92 | 32.04 | |
| POLITÈCNIC | 5205 | 3.88 | 83.86 | 3.55 | 3.55 | 3.55 | 2.19 | 2.29 | 2.63 | 33.39 | 37.79 | 37.56 | 33.49 | 36.79 | 21.89 | |
| AVD. FRANCIA | 4839 | 1.74 | 2.52 | 1.84 | 1.84 | 1.84 | 0.06 | 23.52 | 23.52 | 47.37 | 33.33 | 33.33 | 0.02 | 25.87 | 15.14 | |
| MOLÍ DEL SOL | 4839 | 3.64 | 10.35 | 5.56 | 5.56 | 5.56 | 3.64 | 10.60 | 10.85 | 10.27 | 32.01 | 31.43 | 27.13 | 31.02 | 14.43 | |
| BULEVARD SUD | 4474 | 4.83 | 78.07 | 6.64 | 6.64 | 6.64 | 2.77 | 31.40 | 92.27 | 96.96 | 25.86 | 24.41 | 24.41 | 25.77 | 32.82 | |
| CONSELLERIA METEO. | 3744 | 74.76 | 87.87 | 75.93 | 75.93 | 75.93 | 74.12 | 78.18 | 87.58 | 87.58 | 4.51 | 4.54 | 4.54 | 4.54 | 56.62 | |
| VALÈNCIA CENTRE | 1552 | 75.26 | 94.97 | 14.82 | 14.82 | 14.82 | 81.44 | 16.49 | 17.98 | 98.45 | 1.80 | 1.80 | 1.80 | 1.80 | 33.56 | |
| ALBUFERA2 (MÒBIL) | 731 | 41.59 | 35.29 | 36.80 | 36.80 | 36.80 | 39.26 | 76.74 | 84.13 | 88.10 | 40.90 | 40.63 | 40.63 | 95.08 | 53.29 | |
| VALENCIA-ALBUFERA | 365 | 33.70 | 40.00 | 36.44 | 36.44 | 36.44 | 33.42 | 44.66 | 70.96 | 98.90 | 0.00 | 0.00 | 0.00 | 0.00 | 33.15 | |
| VALENCIA | 365 | 76.99 | 92.05 | 89.86 | 89.86 | 89.86 | 90.68 | 95.07 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 94.18 | |
| TOTAL ESTACIONES: | 100767 | 11.03 | 53.91 | 41.85 | 41.93 | 41.83 | 41.77 | 69.49 | 76.19 | 86.26 | 74.05 | 65.80 | 63.78 | 74.18 | 57.08 | |



Desarrollo del proyecto

Machine Learning. Selección de datos.

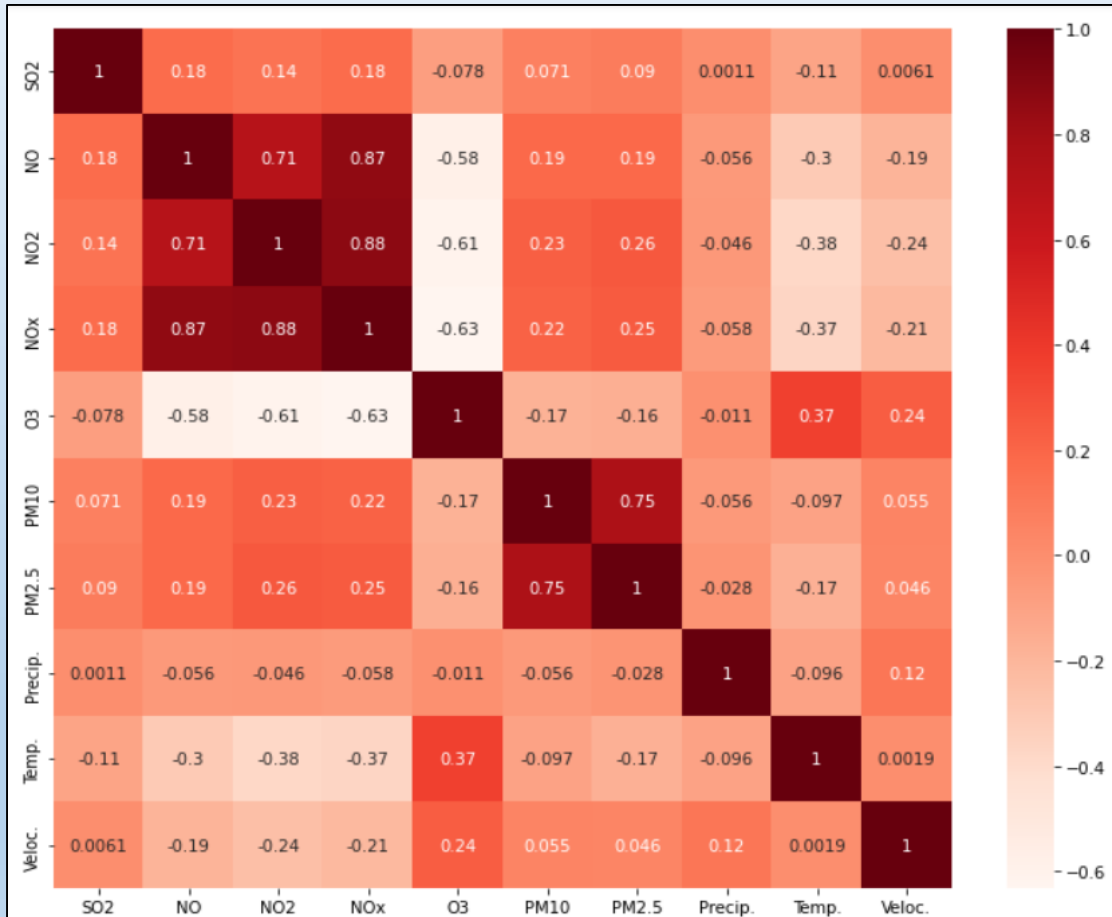
Estaciones principales seleccionadas por su bajo ratio de datos faltantes

```
Nº de registros totales y rango de fechas por estación:  
==CODIGO== ===NOM_ESTACION=== ==Desde=== ==Hasta=== =Nº reg=  
[46250030] PISTA DE SILLA      1994-01-01 2022-03-31    9919  
[46250046] POLITÈCNIC          2008-01-01 2022-03-31    5205  
[46250047] AVD. FRANCIA         2009-01-01 2022-03-31    4839  
[46250048] MOLÍ DEL SOL        2009-01-01 2022-03-31    4839
```



Desarrollo del proyecto

Machine Learning. Correlación.



- No hay correlación entre los distintos contaminantes y las partículas. Solo influyen entre sí las partículas PM10 con PM2.5 y por otra parte los contaminantes óxidos de nitrógeno con el ozono.
- No existe correlación entre las variables meteorológicas y las partículas o contaminantes.



Desarrollo del proyecto

Machine Learning. Modelos estudiados.

Hay que utilizar un modelo de series temporales
Se hacen pruebas y se estudian los siguientes:

Basados en **tensorflow**

- Base line
- Linear
- Dense
- Multi step Dense
- Convolutional
- LSTM

Basados en **statsmodels**

- Sarima



Desarrollo del proyecto

Machine Learning. Modelos estudiados.

Modelos con mejores resultados

| MODELO | ERROR CUADRÁTICO MEDIO ⁽¹⁾ | TIEMPO DE ENTRENAMIENTO ⁽²⁾ |
|---|---------------------------------------|--|
| LSTM – Long Short Term Memory | 9.99 | 1 hora 26 minutos |
| SARIMA - Seasonal Autoregressive Integrated Moving Average | 15.95 | 11 horas 12 minutos |

Se elige modelo LSTM

⁽¹⁾ Estación Av. de Francia – Indicador O3

⁽²⁾ Entrenamiento de 20 modelos (4 estaciones x 5 indicadores)



Desarrollo del proyecto

Machine Learning. Modelos estudiados.

- Se hacen distintas pruebas determinando que, con una ventana de entrada de 30 días y una de salida de 1 día, se consiguen buenos resultados y velocidad razonable.
- En cuanto a fechas, se toma toda la serie temporal histórica hasta final de 2019, para evitar los dos años de pandemia.
- Se crea un script que automáticamente genera un modelo predictivo LSTM por cada combinación de estación e indicador de calidad del aire principal



Machine Learning. Predicciones.

- Para poner a prueba los 20 modelos, se hacen test para predecir la evolución de indicadores en los años de pandemia.
- Las predicciones se hacen con datos nunca vistos por el modelo.

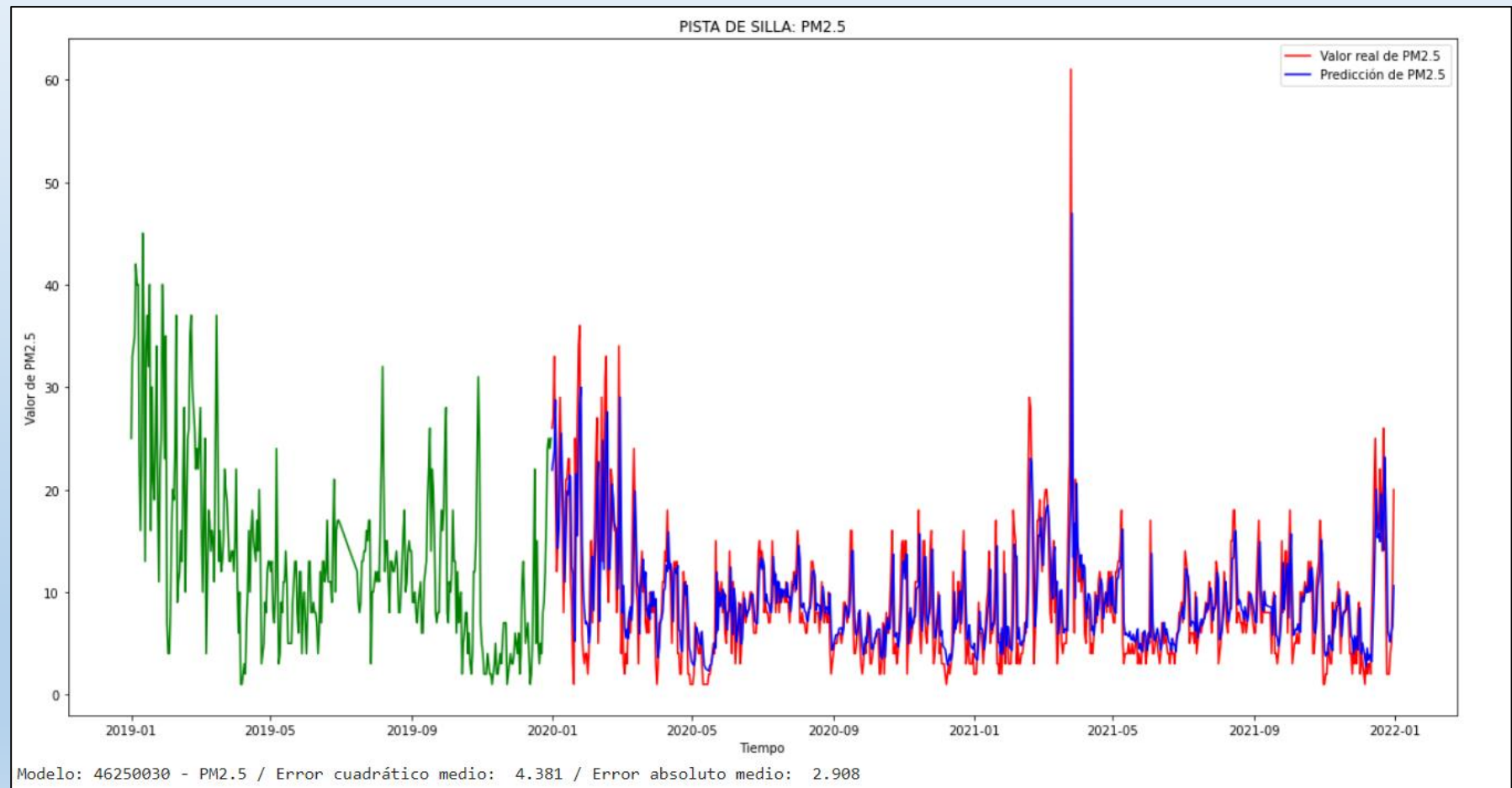


Desarrollo del proyecto

Machine Learning. Predicciones.

PISTA DE SILLA

PM2.5



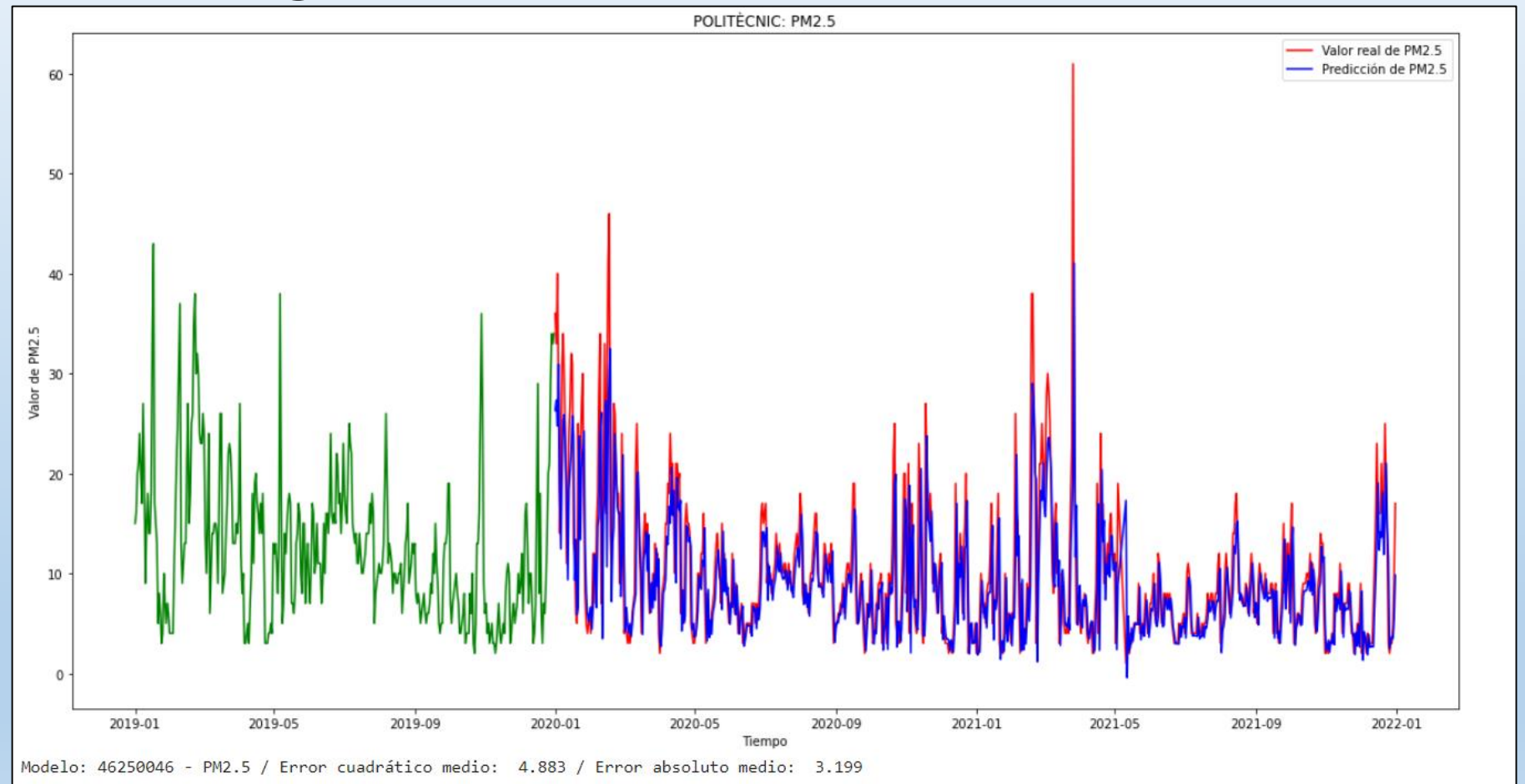


Desarrollo del proyecto

Machine Learning. Predicciones.

POLITÉCNICO

PM2.5



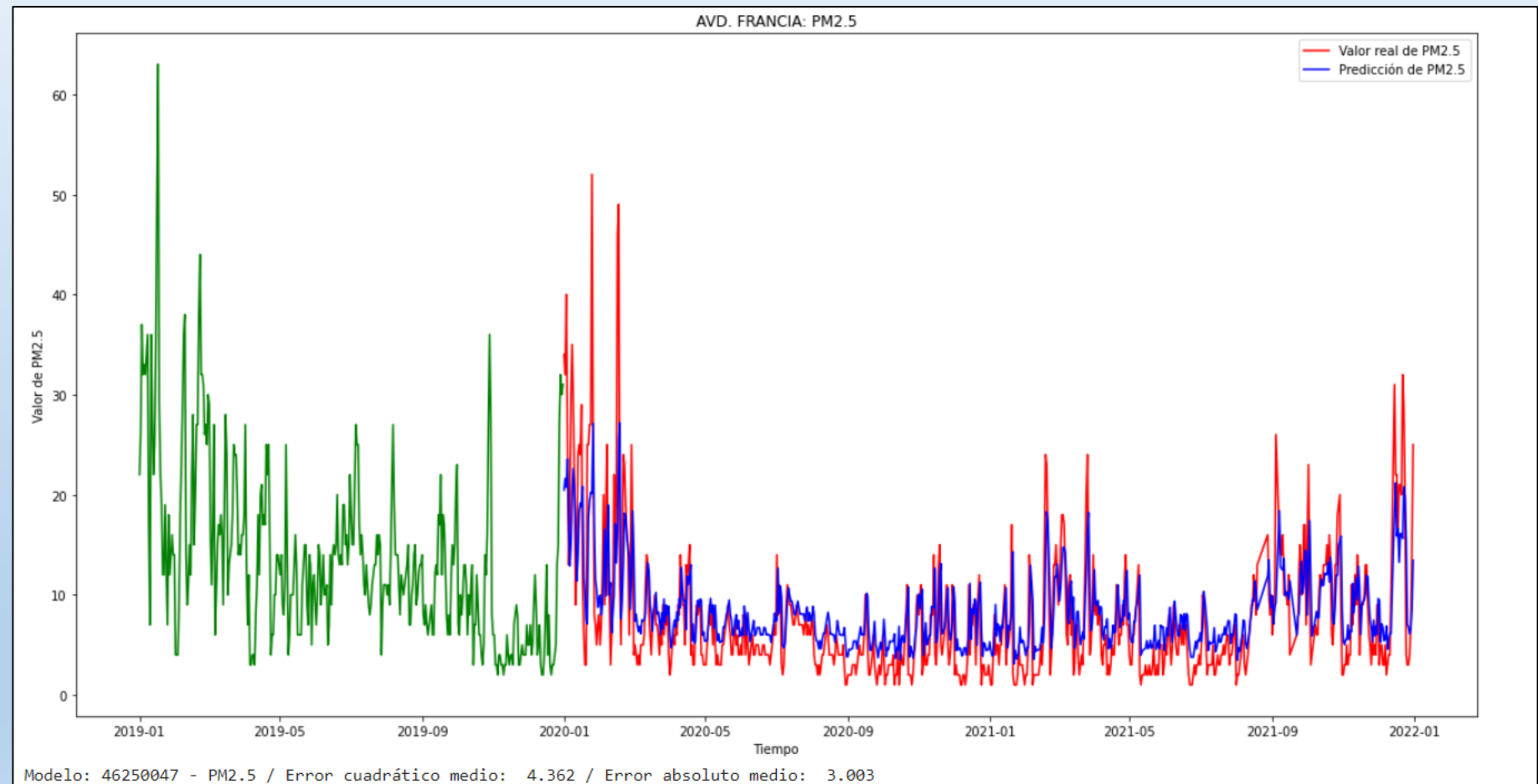


Desarrollo del proyecto

Machine Learning. Predicciones.

AV. FRANCIA

PM2.5



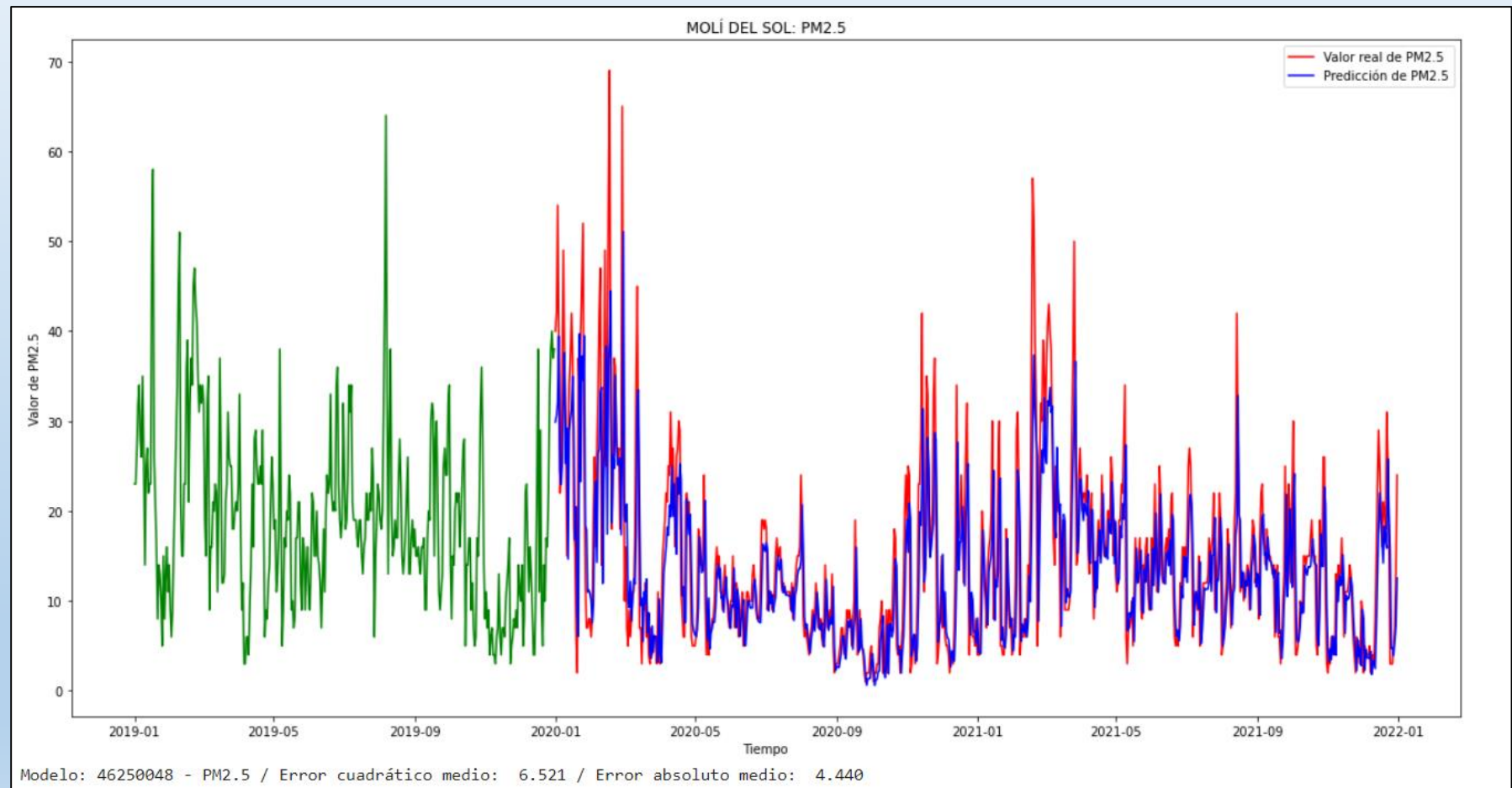


Desarrollo del proyecto

Machine Learning. Predicciones.

MOLÍ DEL SOL

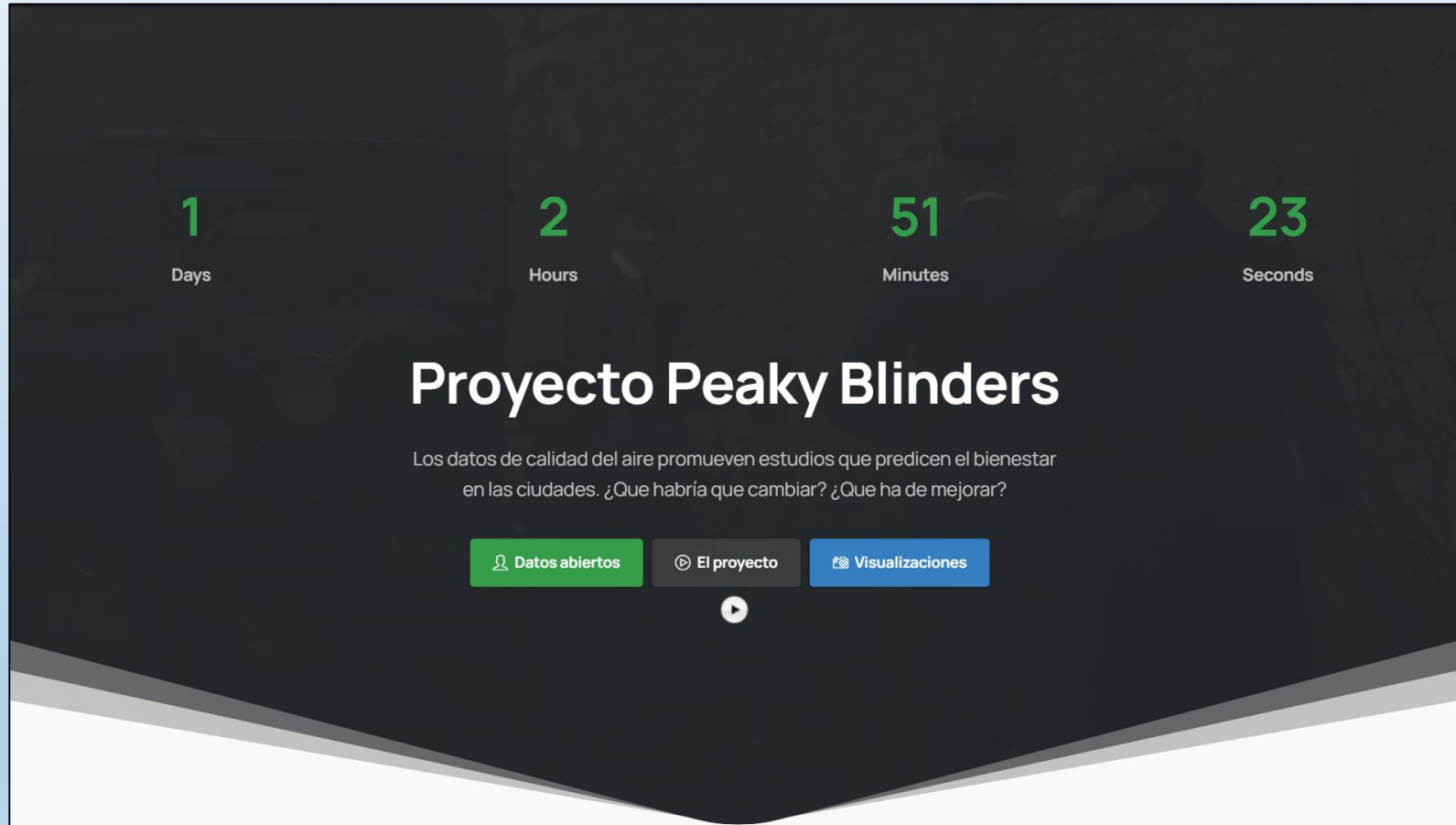
PM2.5





Desarrollo del proyecto

Portal web. Acceso: <http://humanobasico.net>






Desarrollo del proyecto

Portal web: Datos abiertos

[Iniciar Sesión](#) [Registro](#)

 [Conjuntos de datos](#) [Organizaciones](#) [Grupos](#) [Acerca de](#)

[/](#) [Organizaciones](#) / [Peakylinders Advance](#) / [AirQualityValencia](#) / **Calidad del aire valencia**

Calidad del aire valencia

[Descargar](#)

URL: http://humanobasico.com/dataset/d71fcb3-d49c-4c86-9314-39b7116ac275/resource/e7976639-bec5-464d-bdc2-07a1a3eb2505/download/airweather_2013-2019_v3_co...

[Explorador de Datos](#)

[Pantalla completa](#) [Incrustar](#)

[Tabla](#) [Gráfico](#) [Mapa](#)

 10224 records

« 1 - 100 »

[Go »](#) [Filtros](#)

| FECHA | COD_E... | NOM_E... | latitud | longitud | SO2 | NO | NO2 | NOx | O3 | PM10 | PM2.5 | lluvias | Temp |
|-------------|----------|----------|---------|-----------|-----|------|------|-------|------|------|-------|---------|------|
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 1.0 | 5.0 | 30.0 | 39.0 | 27.0 | 6.0 | 5.0 | 0.0 | 13.7 |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 6.0 | 83.0 | 85.0 | 211.0 | 16.0 | 18.0 | 6.0 | 0.0 | 11.4 |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 3.0 | 32.0 | 69.0 | 117.0 | 16.0 | 18.0 | 12.0 | | |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 4.0 | 50.0 | 79.0 | 156.0 | 12.0 | 19.0 | 13.0 | | |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 3.0 | 47.0 | 75.0 | 147.0 | 9.0 | 22.0 | 16.0 | | |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 2.0 | 17.0 | 57.0 | 83.0 | 14.0 | 19.0 | 15.0 | | |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 4.0 | 56.0 | 70.0 | 155.0 | 8.0 | 25.0 | 23.0 | | |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 3.0 | 51.0 | 61.0 | 139.0 | 10.0 | 27.0 | 21.0 | | |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 1.0 | 22.0 | 50.0 | 84.0 | 20.0 | 25.0 | 19.0 | | |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 1.0 | 18.0 | 57.0 | 84.0 | 22.0 | 14.0 | 10.0 | | |
| 2013-01-11 | 46250043 | VIVERS | 39.4793 | -0.368225 | 3.0 | 47.0 | 75.0 | 147.0 | 12.0 | 20.0 | 15.0 | | |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 1.0 | 16.0 | 52.0 | 77.0 | 19.0 | 14.0 | 10.0 | | |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 0.0 | 4.0 | 19.0 | 25.0 | 54.0 | 11.0 | 9.0 | | |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 2.0 | 10.0 | 40.0 | 55.0 | 49.0 | 11.0 | 10.0 | | |
| 2013-01-... | 46250043 | VIVERS | 39.4793 | -0.368225 | 3.0 | 7.0 | 35.0 | 45.0 | 42.0 | 14.0 | 16.0 | | |



Desarrollo del proyecto

Portal web: El proyecto

The screenshot displays the 'El Proyecto' web portal. At the top, a navigation bar includes links for 'Inicio', 'Proyecto', 'Visualizaciones', 'Datos Abiertos', and 'Blog'. Below this, a green header bar contains the title 'El Proyecto' and the breadcrumb 'Home > El Proyecto'. The main content area is divided into two sections: 'Fase 1: Identificación y Objetivos' and 'Fase 2: Diseny del proyecto'.

Fase 1: Identificación y Objetivos

- 1.1 - Introducción**: A card with an illustration of a hand holding a tablet, a Wi-Fi symbol, and a cloud. It includes a date of '5 de junio de 2022' and a 'Read more >' link.
- 1.2 - Indicadores de calidad del aire**: A card with an illustration of a cloud, a Wi-Fi symbol, and a cloud with a Wi-Fi symbol. It includes a date of '5 de junio de 2022' and a 'Read more >' link.
- 1.3 - Estaciones**: A card with three photographs of air quality monitoring stations. It includes a date of '5 de junio de 2022' and a 'Read more >' link.

Fase 2: Diseny del proyecto

- 2.1 - Modelado del proyecto de calidad del aire**: A card with an illustration of a hand holding a tablet showing a map. It includes a date of '6 de junio de 2022' and a 'Read more >' link.
- 2.2 - Herramientas y soluciones tecnológicas**: A card with an illustration of a cloud, a Wi-Fi symbol, and a cloud with a Wi-Fi symbol. It includes a date of '6 de junio de 2022' and a 'Read more >' link.
- 2.3 - Necesidades de hardware y software**: A card with an illustration of a circuit board. It includes a date of '6 de junio de 2022' and a 'Read more >' link.
- 2.4 - Flujo de trabajo y Procesamiento de datos**: A card with a diagram showing the data flow from 'Kafka' to 'Ingestion Service', then to 'Hadoop', 'Analytical Data', and finally to 'Hive/Spark/Presto/Notebooks' and 'Dispersal Service'. It includes a date of '6 de junio de 2022' and a 'Read more >' link.



Desarrollo del proyecto

Portal web. Visualizaciones.





Desarrollo del proyecto

LICENCIA

- Creative Commons Atribución/Reconocimiento - No Comercial - Compartir Igual 4.0 Internacional (CC BY-NC-SA)



- Se debe atribuir o reconocer al creador.
- Solo se permiten usos no comerciales de la aplicación.
- Cualquier adaptación del software debe compartirse en los mismos términos.



EVALUACIÓN Y CONCLUSIONES FINALES



Evaluación y conclusiones finales

EVALUACIÓN EJECUCIÓN DEL PROYECTO

- Gran dificultad para recabar los datos, habiendo invertido mucho tiempo en investigar las fuentes disponibles y en las tareas de EDA y ETL
- Se han cubierto todos los objetivos excepto el módulo de predicciones futuras, no terminado por falta de tiempo.
- Ha resultado sorprendente descubrir que no hay correlación significativa entre los indicadores de calidad del aire y la información meteorológica.





Evaluación y conclusiones finales

PROPUESTA DE MEJORAS

- Automatizar la captura e inserción de nuevas mediciones diarias que vayan alimentando y actualizando las fuentes de datos.
- Automatizar los procesos de aplicación de modelos para realizar las nuevas predicciones diarias.
- Automatización de la generación de presentaciones y cuadros de mandos con los datos actualizados.
- Continuar la publicación de los nuevos datos adquiridos de predicción y conjunto de datos en los elementos online dispuestos para el acceso público a los mismos.



Gracias por vuestra atención

Tono, Mikel y Santi

