

Predicción de índices de contaminación del aire mediante modelos supervisados para la ciudad de Barcelona

**César Fernández Domínguez**

Máster Universitario en Ciencia de Datos (Data Science)

Minería de datos y Machine Learning

**Sergio Trilles Oliver**

**Albert Solé Ribalta**

Junio 2020

  
Esta obra está sujeta a una licencia de Reconocimiento [3.0 España de Creative Commons](http://creativecommons.org/licenses/by/3.0/es/)

**FICHA DEL TRABAJO FINAL**

|  |  |
| --- | --- |
| **Título del trabajo:** | *Predicción de índices de contaminación del aire mediante modelos supervisados para la ciudad de Barcelona* |
| **Nombre del autor:** | *César Fernández Domínguez* |
| **Nombre del consultor/a:** | *Sergio Trilles Oliver* |
| **Nombre del PRA:** | *Albert Solé Ribalta* |
| **Fecha de entrega (mm/aaaa):** | 06/2020 |
| **Titulación:** | *Máster Universitario en Ciencia de Datos* |
| **Área del Trabajo Final:** | *Machine Learning – Modelos predictivos* |
| **Idioma del trabajo:** | *Castellano* |
| **Palabras clave** | *Calidad del aire, predicción, machine learning* |
| **Resumen del Trabajo (máximo 250 palabras):** *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.* | |
| En las últimas décadas han surgido numerosos estudios que relacionan la salud de las personas con la calidad del aire que respiramos. Los cuales han señalado la contaminación atmosférica como causa de mortalidad, principalmente en niños y ancianos con problemas respiratorios. Apoyándose en estos trabajos, o también forzado por los resultados obtenidos en ellos, se ha promovido, por parte de los organismos internacionales y los países en sí, una concienciación para controlar las emisiones de contaminantes al aire que respiramos en nuestro día a día. En España se ha hecho un gran esfuerzo para reducir significativamente estos niveles de contaminación. Sin embargo, la gran concentración de empresas alrededor de las grandes ciudades, y la masificación en torno a ellas, lo cual unido a vehículos contaminantes, ha obligado a que los ayuntamientos de estas ciudades tengan que regular para salvaguardar la calidad del aire de sus conciudadanos.  En este contexto, los organismos tienen que proveerse de herramientas que les permita anticiparse a episodios de alta contaminación y poder actuar en consecuencia.  Este trabajo analiza los datos de contaminación atmosférica recogidos diariamente en determinados puntos de la ciudad de Barcelona, y disponibles abiertamente al público, junto con datos de predicción meteorológica y de densidad de tráfico, para construir un modelo que permita realizar una predicción, suficientemente fiable, de la calidad del aire, con suficiente antelación. | |
| **Abstract (in English, 250 words or less):** | |
| In recent decades, numerous studies have emerged that link people's health to the quality of the air we breathe. They have pointed out air pollution as a cause of mortality, mainly in children and elderly people with respiratory problems. Based on these works, or also forced by the results obtained in them, an awareness has been promoted by international organizations and countries themselves to control emissions of pollutants into the air we breathe in our daily lives. In Spain, a great effort has been made to significantly reduce these pollution levels. However, the great concentration of companies around large cities, and the massification around them, which together with polluting vehicles, has forced the town councils of these cities to regulate in order to safeguard the quality of the air of their fellow citizens.  In this context, organizations must provide themselves with tools that allow them to anticipate episodes of high pollution and be able to act accordingly.  This work analyses the air pollution data collected daily at certain points in the city of Barcelona, and openly available to the public, together with data on weather forecasting and traffic density, to build a model that will make it possible to make a sufficiently reliable prediction of air quality in good time. | |

**Índice**

[1. Introducción 1](#_Toc35790221)

[1.1 Contexto y justificación del Trabajo 1](#_Toc35790222)

[1.2 Objetivos del Trabajo 2](#_Toc35790223)

[1.2.1 Hipótesis (u objetivo principal) 2](#_Toc35790224)

[1.2.2 Objetivos parciales (o preguntas de investigación) 3](#_Toc35790225)

[1.3 Enfoque y método seguido 3](#_Toc35790226)

[1.4 Planificación del Trabajo 4](#_Toc35790227)

[1.5 Breve sumario de productos obtenidos 6](#_Toc35790228)

[1.6 Breve descripción de los otros capítulos de la memoria 6](#_Toc35790229)

[2. Estado del arte 7](#_Toc35790230)

[3. Resto de capítulos 14](#_Toc35790231)

[4. Conclusiones 15](#_Toc35790232)

[5. Glosario 16](#_Toc35790233)

[6. Bibliografía 17](#_Toc35790234)

[7. Anexos 21](#_Toc35790235)

**Lista de figuras**

[Figura 1- Fases de la metodología CRISP-DM (Fuente: [10]) 4](#_Toc35790236)

[Figura 2 - Planificación inicial del proyecto 5](#_Toc35790237)

**Lista de tablas**

[Tabla 1 - Detalle de tareas del plan del proyecto 6](#_Toc35790238)

# 1. Introducción

## 1.1 Contexto y justificación del Trabajo

Hoy en día todos estamos concienciados de que debemos cuidar de nuestro ecosistema. Es ampliamente compartido que si hacemos lo posible por rodearnos de espacios verdes y reducimos las emisiones de contaminantes al aire, ello repercute en nuestra salud. Y, por implicación, en la economía, pues se reducen los problemas de salud generalizados y es necesaria una menor asistencia médica y hospitalaria. Son varios los organismos que actualmente velan por la calidad del aire. Desde organismos supranacionales, como la Organización Mundial de la Salud (World Health Organization, en inglés), el Consejo de Europa, dentro del marco de la Unión Europea, el Gobierno de España, estableciendo leyes de ámbito nacional (Ley 16/2002 [1] y Ley 34/2007 [2]), o los propios ayuntamientos de grandes ciudades.

Según la Organización Mundial de la Salud (OMS), el aire limpio es un requisito básico de la salud y el bienestar humanos. En relación con la contaminación del aire se producen anualmente más de dos millones de muertes mucho antes de su esperanza de vida [3]. La OMS trata anualmente de concienciar a los países para que reduzcan sus niveles de contaminación, para lo cual promueve estudios relacionados con la calidad del aire y define unas guías que tienen por objeto ofrecer una orientación sobre la manera de reducir los efectos de la contaminación del aire en la salud.

En la ciudad de Barcelona, el ayuntamiento ha puesto en marcha distintas iniciativas que buscan reducir la contaminación debida a la circulación de vehículos por el casco urbano. Estas medidas van desde el fomento del uso del transporte público, el uso de aparcamiento disuasorios, la implantación del alquiler de bicicletas, hasta medidas más novedosas como las zonas de bajas emisiones [4], y la restructuración urbana definiendo las llamadas supermanzanas [5].

Las zonas de bajas emisiones (ZBE) de la ciudad de Barcelona tienen por objeto limitar el acceso a vehículos altamente contaminantes al centro de la ciudad. Se establece un perímetro en el cual aquellos vehículos sin distintivo de la DGT no podrán circular, bajo pena de multa. Y se establece limitaciones para el resto.

En cuanto al proyecto de las supermanzanas de Barcelona, este regula el tráfico aprovechando la configuración urbanística de la ciudad, de manera que, sin limitar totalmente el acceso de los vehículos, si se reduce este en gran medida y, además, se amplían las zonas verdes y de esparcimiento de la ciudad.

Los principales contaminantes emitidos directamente por el tráfico son el monóxido de carbono (CO), óxidos de nitrógeno (NOx), compuestos orgánicos volátiles (VOCs), y material particulado (PM). La emisión de estos contaminantes depende en gran medida del tipo de combustible utilizado y de la antigüedad de los vehículos. Actualmente, las factorías de automóviles desarrollan motores más evolucionados que disminuyen las emisiones de partículas contaminantes, además de reducir el consumo. A lo cual se debe añadir la aparición de nuevas tecnologías, tales como los coches híbridos y eléctricos. Sin embargo, aún supone un alto porcentaje los vehículos antiguos que no fueron diseñados para evitar la polución.

Al propio efecto del tráfico rodado, en la calidad del aire, debemos añadir otros factores. Como por ejemplo, la baja tasa de precipitación, lo cual impide el lavado atmosférico e influye en la importancia de procesos de resuspensión por el tráfico rodado del material particulado depositado en los firmes de la carretera [6]; la elevada radiación solar en los meses estivales, que se traduce en un incremento de los niveles de partículas secundarias (como sulfato y nitrato), NO2 y ozono.

Ante este escenario, los ayuntamientos de las grandes ciudades, principalmente, dado que es donde mayor concentración de industria y tráfico suele existir, son responsables de garantizar la calidad del aire de los habitantes de su ciudad. Así, estos, deben de vigilar que no se superan, o incluso se mantienen bajos, los niveles de contaminación según las recomendaciones marcadas por la legislación vigente. Para ello se establecen, por parte de los ayuntamientos, estaciones de medición, distribuidas estratégicamente por la ciudad. Sin embargo, esto únicamente permite obtener una instantánea del estado actual de la calidad del aire en la ciudad. Como mucho, puede servir para realizar estudios sobre la evolución de esta sobre un periodo de medida.

Tal y como se recoge en la Directiva Europea 2008/50/CE [7]: “cuando sea posible, deben aplicarse técnicas de modelización que permitan interpretar los datos puntuales en función de la distribución geográfica de la concentración, lo que podría servir de base para calcular el grado de exposición colectiva de la población residente en la zona.” Por lo tanto, este proyecto tratará de encontrar un modelo predictivo que, a partir de observaciones de los principales contaminantes, recogidos en distintas zonas de la ciudad de Barcelona, y teniendo en cuenta diversos factores que pueden tener relación directa, o indirecta, con la calidad del aire, como por ejemplo, volumen del tráfico rodado o climatología, permita a los organismos públicos adelantarse en la toma de decisiones.

## 1.2 Objetivos del Trabajo

Como ya se ha recogido en el apartado anterior, es responsabilidad de los organismos públicos regular y vigilar para que la calidad del aire que respiramos en las ciudades se mantenga dentro de unos mínimos de contaminación. Estos organismos deberán establecer medidas que mejoren la calidad de vida de sus ciudadanos. Lo cual repercutirá indiscutiblemente en beneficios tanto de salud como económicos para la sociedad en su conjunto.

### 1.2.1 Hipótesis (u objetivo principal)

El objetivo principal que perseguirá este trabajo es el de obtener un modelo, de minería de datos, lo suficientemente fiable, que permita predecir algunos de los factores contaminantes que definen la calidad del aire en la ciudad de Barcelona.

### 1.2.2 Objetivos parciales (o preguntas de investigación)

Para lograr el objetivo principal de este trabajo, el cual se ha establecido en el punto anterior, se marcan una serie de objetivos parciales que se tratará de conseguir. Estos objetivos parciales estarán definidos por:

* Recogida de datos suficientes, en relación con los contaminantes y factores adicionales, que nos permitan realizar una caracterización correcta de la calidad del aire en distintas zonas de la ciudad de Barcelona
* Validación de los datos recopilados para este trabajo. Muy probablemente estos datos tendrán que ser modificados para lograr nuestro modelo, o modelos, de predicción.
* Comprobación de las posibles correlaciones existentes entre los distintos parámetros evaluados para determinar la calidad del aire y, que formarán parte de nuestro modelo. Aquí se intentará comprobar como existe relación entre los niveles alcanzados en distintos contaminantes medidos en las estaciones de medida distribuidos por la ciudad, y otros factores como, por ejemplo: el volumen de tráfico registrado.
* Encontrar una parametrización para nuestro modelo de predicción que obtenga unos resultados que nos permitan dar respuesta al objetivo principal.
* Comparar distintos modelos para encontrar aquel que nos ofrezca mejores resultados.

## 1.3 Enfoque y método seguido

Existen dos estrategias principales para abordar un trabajo de investigación, como el que se plantea en este caso. Una de ellas es la estrategia **cuantitativa o empírico-analítica**, basada en la observación de los hechos empíricos mediante el distanciamiento del observador, la medición matemática y cuantitativa, y la producción de situaciones experimentales controladas sobre las variables que influyen en el fenómeno estudiado [8]. La cual es la estrategia de investigación que ha predominado hasta inicios del siglo XX. A partir de entonces, se plantean nuevas estrategias de investigación orientadas en la comprensión e interpretación émica de los fenómenos humanos y sociales, basándose en como estos se comportan. Estas se han denominado estrategias **cualitativas** (también denominadas comprensivas o constructivistas).

En este trabajo se seguirá un enfoque cuantitativo, pues resulta más apropiado para conseguir nuestro objetivo de obtener un modelo predictivo. Trataremos de obtener datos de los fenómenos que determinan la calidad del aire en las distintas zonas analizadas. A partir de estos datos, los cuales obtendremos de fuentes de datos abiertos al público, se estudiarán, en los siguientes apartados, las relaciones existentes entre estos y se buscará la construcción de un modelo que nos permita dar respuesta al fenómeno analizado: calidad del aire.

Durante el desarrollo este trabajo se ha intentado seguir las recomendaciones, para conseguir la reproducibilidad del conjunto, recopiladas en el artículo publicado en la revista ecosistemas [9]. Todo el código utilizado durante el estudio y los datos asociados, así como los resultados obtenidos, se pueden encontrar en un repositorio de GitHub debidamente dispuestos para su fácil reproducción siempre que sea preciso por cualquiera que quiera comprobar los resultados obtenidos en este proyecto. Para acceder al repositorio con los datos y código utilizados para este trabajo acceder al enlace: <https://github.com/cesisar/TFM_UOC_2020>

## 1.4 Planificación del Trabajo

Para la realización de este trabajo se sigue la metodología de CRISP-DM, la cual se ha convertido de facto en la metodología utilizada en los proyectos de minería de datos [10]. La siguiente figura representa el ciclo de fases que componen esta metodología de diseño.

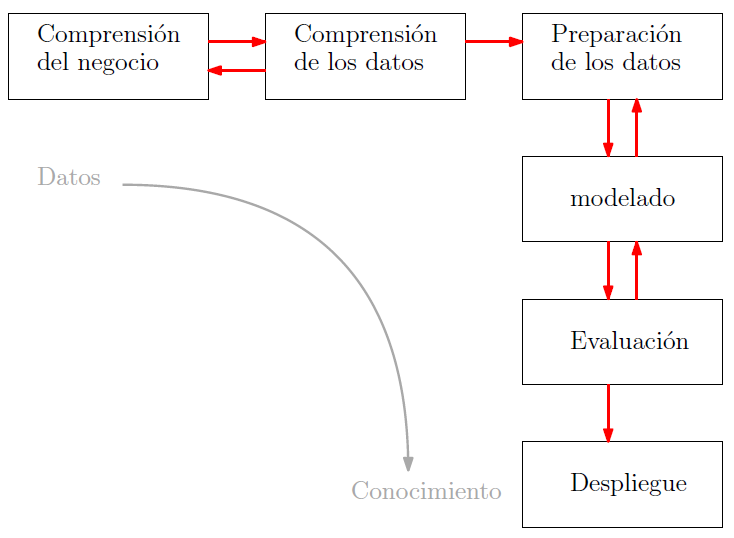


Figura 1- Fases de la metodología CRISP-DM (Fuente: [10])

En esta metodología se definen las siguientes fases, que son las que marcaran la planificación de este Trabajo:

* **Comprensión del negocio**. En esta fase nos enfocaremos en la búsqueda de información, relativa a la calidad del aire, que nos ayude a definir los objetivos propios de nuestro estudio y estableceremos nuestra planificación de trabajo (este punto).
* **Compresión de los datos**. Descargaremos los datos de cada una de las fuentes de datos abiertos disponibles y analizaremos el significado e idoneidad de estos para la realización de nuestro trabajo.
* **Preparación de los datos**. Muy probablemente los datos tendrán que ser modificados para adecuarlos a los modelos que previamente evaluaremos.
* **Modelado**. Esta fase consistirá, en nuestro caso, en la construcción de varios modelos de predicción que incluyan las variables analizadas y que den respuesta al objetivo marcado anteriormente.
* **Evaluación**. Aquí evaluaremos la exactitud de cada uno de los modelos construidos y determinaremos cual será aquel que resulta más conveniente para nuestro propósito.
* **Despliegue**. Esta fase tiene que ver con la puesta en operación de nuestro modelo. Sin embargo, en nuestro caso, al tratarse de un trabajo fin de máster, esta fase terminará con la presentación del trabajo al tribunal examinador.

A continuación, en la siguiente figura, se detalla cuáles son los hitos establecidos para la consecución de este trabajo. También se detalla las fechas de finalización de cada hito, marcadas por las fechas de entrega de cada una de las PECs.

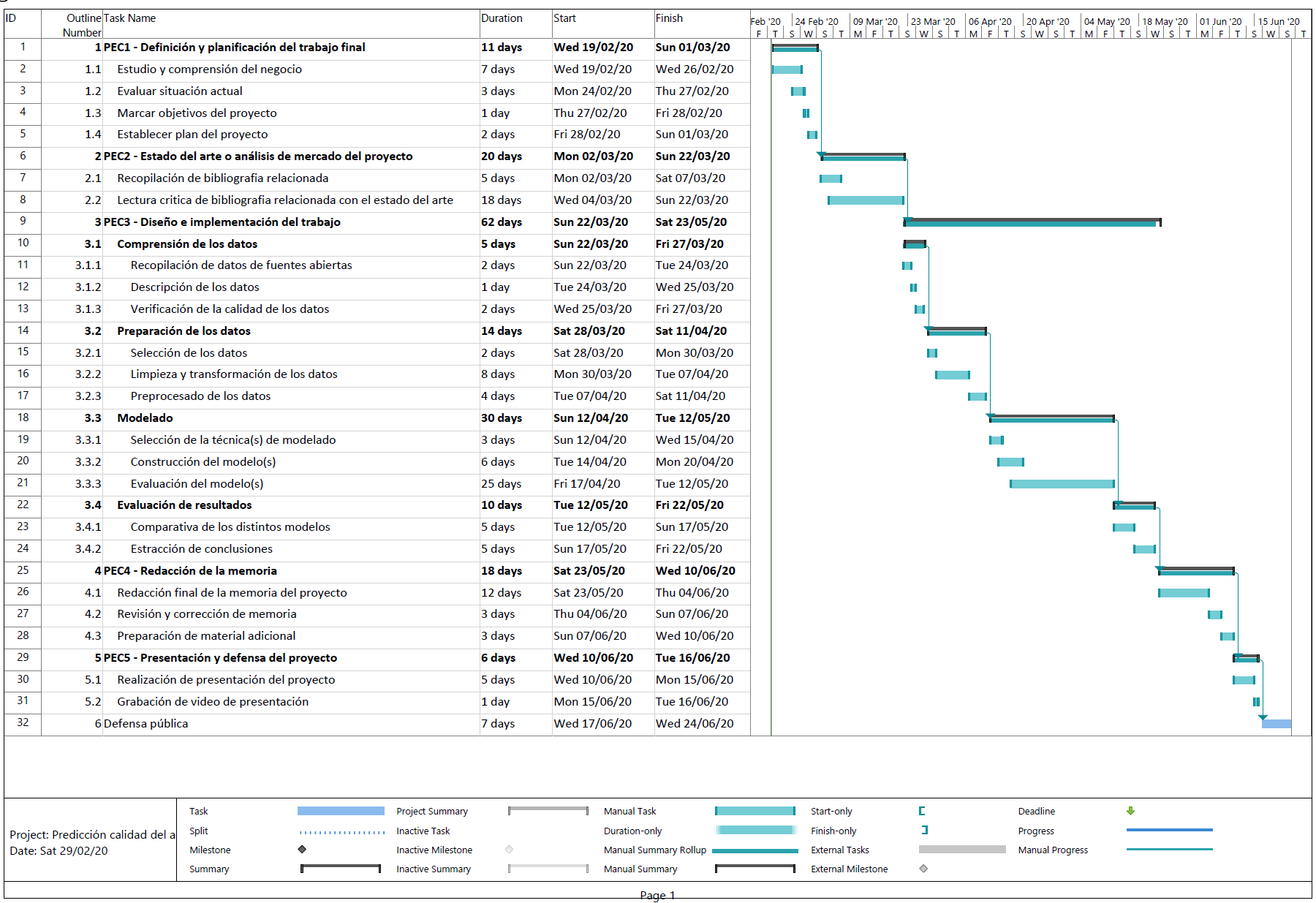


Figura 2 - Planificación inicial del proyecto

Lo cual veremos más claramente en la siguiente tabla:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Outline Number | Task Name | Duration | Start | Finish |
| **1** | **PEC1 - Definición y planificación del trabajo final** | **11 days** | **Wed 19/02/20** | **Sun 01/03/20** |
| 1.1 | Estudio y comprensión del negocio | 7 days | Wed 19/02/20 | Wed 26/02/20 |
| 1.2 | Evaluar situación actual | 3 days | Mon 24/02/20 | Thu 27/02/20 |
| 1.3 | Marcar objetivos del proyecto | 1 day | Thu 27/02/20 | Fri 28/02/20 |
| 1.4 | Establecer plan del proyecto | 2 days | Fri 28/02/20 | Sun 01/03/20 |
| **2** | **PEC2 - Estado del arte o análisis de mercado del proyecto** | **20 days** | **Mon 02/03/20** | **Sun 22/03/20** |
| 2.1 | Recopilación de bibliografía relacionada | 5 days | Mon 02/03/20 | Sat 07/03/20 |
| 2.2 | Lectura crítica de bibliografía relacionada con el estado del arte | 18 days | Wed 04/03/20 | Sun 22/03/20 |
| **3** | **PEC3 - Diseño e implementación del trabajo** | **62 days** | **Sun 22/03/20** | **Sat 23/05/20** |
| **3.1** | **Comprensión de los datos** | **5 days** | **Sun 22/03/20** | **Fri 27/03/20** |
| 3.1.1 | Recopilación de datos de fuentes abiertas | 2 days | Sun 22/03/20 | Tue 24/03/20 |
| 3.1.2 | Descripción de los datos | 1 day | Tue 24/03/20 | Wed 25/03/20 |
| 3.1.3 | Verificación de la calidad de los datos | 2 days | Wed 25/03/20 | Fri 27/03/20 |
| **3.2** | **Preparación de los datos** | **14 days** | **Sat 28/03/20** | **Sat 11/04/20** |
| 3.2.1 | Selección de los datos | 2 days | Sat 28/03/20 | Mon 30/03/20 |
| 3.2.2 | Limpieza y transformación de los datos | 8 days | Mon 30/03/20 | Tue 07/04/20 |
| 3.2.3 | Preprocesado de los datos | 4 days | Tue 07/04/20 | Sat 11/04/20 |
| **3.3** | **Modelado** | **30 days** | **Sun 12/04/20** | **Tue 12/05/20** |
| 3.3.1 | Selección de la técnica(s) de modelado | 3 days | Sun 12/04/20 | Wed 15/04/20 |
| 3.3.2 | Construcción del modelo(s) | 6 days | Tue 14/04/20 | Mon 20/04/20 |
| 3.3.3 | Evaluación del modelo(s) | 25 days | Fri 17/04/20 | Tue 12/05/20 |
| **3.4** | **Evaluación de resultados** | **10 days** | **Tue 12/05/20** | **Fri 22/05/20** |
| 3.4.1 | Comparativa de los distintos modelos | 5 days | Tue 12/05/20 | Sun 17/05/20 |
| 3.4.2 | Extracción de conclusiones | 5 days | Sun 17/05/20 | Fri 22/05/20 |
| **4** | **PEC4 - Redacción de la memoria** | **18 days** | **Sat 23/05/20** | **Wed 10/06/20** |
| 4.1 | Redacción final de la memoria del proyecto | 12 days | Sat 23/05/20 | Thu 04/06/20 |
| 4.2 | Revisión y corrección de memoria | 3 days | Thu 04/06/20 | Sun 07/06/20 |
| 4.3 | Preparación de material adicional | 3 days | Sun 07/06/20 | Wed 10/06/20 |
| **5** | **PEC5 - Presentación y defensa del proyecto** | **6 days** | **Wed 10/06/20** | **Tue 16/06/20** |
| 5.1 | Realización de presentación del proyecto | 5 days | Wed 10/06/20 | Mon 15/06/20 |
| 5.2 | Grabación de video de presentación | 1 day | Mon 15/06/20 | Tue 16/06/20 |
| 6 | Defensa pública | 7 days | Wed 17/06/20 | Wed 24/06/20 |

Tabla 1 - Detalle de tareas del plan del proyecto

## 1.5 Breve sumario de productos obtenidos

No hay que entrar en detalle: la descripción detallada se hará en el resto de los capítulos.

## 1.6 Breve descripción de los otros capítulos de la memoria

Explicación de los contenidos de cada capítulo y su relación con el trabajo en global.

# 2. Estado del arte

El crecimiento de las grandes ciudades y la industrialización en torno a ellas, como ya se ha comentado anteriormente, ha traído consigo un deterioro sustancial de la calidad del aire que respiramos, muy especialmente, en ciudades con mucha densidad de población. Desde hace ya varios años atrás son muchos los gobiernos que han promovido iniciativas para paliar este hecho. Por ejemplo, es el caso del gobierno chino, el cual, dentro de su plan nacional de protección del medio ambiente, para el quinquenio 2006-2010, ha impulsado estudios como el de *Wang et al.* [11] [12] en el cual se modela cual sería el aumento en emisiones de partículas contaminantes, concretamente del dióxido de azufre (SO2), en el caso de que no se realizara ninguna medida correctora y se estuviera en un escenario *Business-As-Usual* [13]. Este estudio analiza cuales son las principales fuentes de emanación de SO2, y diferencia entre transporte terrestre, uso doméstico, industria y plantas de generación de potencia. El objetivo de este estudio es encontrar las condiciones de emisión que permitirían una reducción del 10% en el nivel de SO2 en el aire, respecto al año 2005. Para esto, este estudio, utiliza técnicas de modelado deterministas que tratan de modelar los procesos químicos que intervienen en la generación de los gases contaminantes. En concreto se utiliza el modelo denominado *Community Multi-Scale Air Quality* (CMAQ). Este modelo, junto con otros modelos de predicción meteorológicos, como el MM5 [13] [14] [15], también empleado este artículo para predecir los parámetros climáticos necesarios como input del modelo, es discutido en la colección de dos artículos de *Zhang et al.,* [13] [14] , publicados en la revista *Atmospheric Environment* en el año 2012. En estos dos artículos se analiza el presente y futuro de este tipo de modelos en la predicción en tiempo real de la calidad del aire.

Otros muchos artículos tratan de evaluar la correlación de distintos factores con el aumento o disminución de la concentración de partículas nocivas en el aire, como son: la densidad del tráfico (congestión), meteorología, factores estructurales de las ciudades, así como, incluso, la disposición de árboles a los lados de las calles en las ciudades. Así podemos encontrar estudios como el de *Fiore et al.* [15] en el cual se detallan las conexiones de influencia bidireccional entre el clima y la calidad del aire. En este artículo podemos ver como existe una conexión entre el aumento de radiación solar que se produce en el verano, con un aumento en los niveles de ozono (O3). O como el viento o la lluvia puede influir en la dispersión de partículas contaminantes (*Pollutant Matter* (PM), en inglés). También encontramos artículos como el de *Zhang K. & Batterman S.* [16] en el que se analiza el riesgo para la salud condicionado por la congestión del tráfico en autopistas y en carreteras arteriales de la ciudad. En el primer caso se ve que este riesgo sigue un patrón en forma de “U” según aumenta el volumen de tráfico. Sin embargo, en el caso de carreteras dentro de la ciudad se produce un aumento drástico de estos riesgos para la salud con el aumento de tráfico. Yuan et al. [17] modela la correspondencia entre la distribución de los edificios en las ciudades, y su altura, y los vientos para permitir la dispersión de las partículas contaminantes. Por último, podemos encontrar estudios como el de *Vos et al.* [18], en el que se modeliza la disposición de vegetación en los bordes de las carreteras urbanas, concluyendo que esta, en vez de contribuir a mejorar la calidad del aire, reduce la capacidad de ventilación y, por lo tanto, el efecto de emisión de oxígeno que pudiera producir los árboles, se ve superado por la mayor concentración de contaminantes.

Hoy en día, el problema de la predicción de la calidad del aire puede ser abordada desde tres enfoques muy diferentes: modelos deterministas, métodos puramente estadísticos y *Machine Learning* (ML) ( [19] [20] [21]). En el caso de los modelos deterministas, como puede ser el modelo CMAQ arriba mencionado, tratan de llevar a cabo la predicción del aire mediante la construcción de un modelo de simulación de los procesos de dispersión y transporte de los contaminantes. Estos modelos no suelen necesitar de la existencia de un gran volumen de datos históricos, sin embargo, demandan tener suficiente conocimiento de las fuentes de contaminación y los procesos químicos y físicos que intervienen y, además, suelen ser procesos caros y con gran demanda de recursos para su ejecución. Los modelos estadísticos, en cambio, precisan de datos históricos, aunque parten de la adopción de diferentes suposiciones sobre los datos y un conocimiento previo sobre ellos. Algunos de estos modelos, ampliamente utilizados, son: *Autoregressive Moving Average* (ARMA), *Autoregressive Integrated Moving Average* (ARIMA), *Threshold Autoregressive* (TAR), *Hidden Markov Model* (HMM), o técnicas de interpolación geoestadísticas como: *Ordinary Least Squares* (OLS), *Inverse Distance Weight* (IDW), *Universal Kriging* (UK), *Downscale*r, etc. Por último, recientemente, son cada vez más utilizados enfoques que tienen su base en modelo de ML, o *Deep Learning*. Estos modelos tienen en su favor que no requieren de un conocimiento profundo de los datos, y pueden llegar a obtener resultados con mayor exactitud, ya que se adaptan a la no linealidad de las series temporales de los datos, de forma más eficiente y con un menor coste. Algunas de las técnicas utilizadas aquí son: *Support Vector Machine* (SVM) [22], *Artificial Neural Networks* (ANN), *Fuzzy Logic* (FL), *Random Forest* RF), *Partial Least squares Regression* (PLSR) [23], *Multivariate Polynomial Regression* (MPR) [23], *XGBoost* [24], etc.

La gran potencia de las técnicas de *Machine Learning*, en el campo de la predicción de la calidad del aire, y más concretamente del *Deep Learning*, está corroborada por infinidad de trabajos publicados recientemente. Estos trabajos, en su mayoría, intentan encontrar una técnica que maximice la precisión y exactitud de los resultados obtenidos con respecto a otras técnicas ya probadas anteriormente o, intentan resolver algunos puntos débiles de alguna técnica concreta. A continuación se hace un recorrido por algunos de estos trabajos.

*Giorgio Corani* [25], en su trabajo, utiliza un modelo *Feed-Forward Neural Networks* (FFNN) para predecir lo niveles de Ozono (O3) y PM2.5 en el aire a las 9 am., en la ciudad de Milán, Italia. En este trabajo, se selecciona el modelo FFNN por ser considerado el estado-del-arte en este campo en el momento de su realización. Este modelo es comparado con otros dos modelos que utilizan otras técnicas de aprendizaje automático, *Pruned Neural Networks* (PNN) y *Lazy Learning* (LL), que intentan contrarrestar algunos de los inconvenientes que tiene las FFNNs, como: sobreajuste, dificultad para diseñar la arquitectura de la red neuronal, difícil interpretación del significado de los parámetros o relevancia de estos en el resultado final. El problema del sobreajuste parece ser resuelto con el empleo de ambas técnicas alternativas sin menoscabar en la bondad de los resultados. Sin embargo, la dificultad para realizar el diseño de la arquitectura sigue latente al utilizar la técnica de PNN.

El estudio llevado a cabo por *Voukantsis et al.* [26] compara los índices de contaminación entre dos ciudades costeras del norte y sur de Europa (Helsinki en Finlandia y Tesalónica en Grecia), con muy distintas estructuras geológicas e industriales. Para realizar esta comparación sigue un análisis de componentes principales (PCA) para obtener los principales factores (PC) que explican la variabilidad de los datos de contaminación. En esta parte, el artículo concluye que la principal fuente de contaminación en ambas ciudades es la debida al tráfico urbano (esto explica un 25% de la variabilidad de los datos). El artículo incluye una segunda parte, en la cual se comparan los modelos de aprendizaje automático construidos para predecir la concentración de PM10 y PM2,5,siguiendo un novedoso método, que mezcla análisis lineal y no-lineal, para obtener el mejor modelo en cada caso. A pesar de que los modelos construidos para ambos casos, utilizando este método, parten de diferentes factores, el resultado obtenido es similar.

El artículo de *Singh et al.* [23] compara entre modelos de lineales y no lineales para la predicción de los contaminantes PM, SO2 y NO2, en la ciudad de Lucknow (India), a partir de datos de estaciones de monitorización de la calidad del aire y de datos meteorológicos. Los mejores resultados se obtienen para los modelos no lineales, dada la naturaleza no lineal de cada uno de los parámetros analizados. En concreto, y por orden de peor a mejor, se comparan los modelos, lineal: regresión de mínimos cuadrados parciales y no lineales: regresión polinómica multivariante, perceptrón multicapa, red de función de base radial y red neuronal de regresión generalizada. Las tres últimas son variantes de redes neuronales artificiales (ANN, en sus siglas en inglés).

*Russo et al.* [27] proponen aplicar los recientes avances en análisis de datos estocásticos para reducir el número de variables de entrada para modelos de redes neuronales artificiales para el pronóstico de la calidad del aire, sin alterar el poder predictivo de estos. El hecho de reducir el número de variables de entrada del modelo repercute en la rapidez con la que el modelo se ejecuta. Lo cual permite la implementación de sistemas de alarma más efectivos en el aviso de condiciones de calidad del aire desfavorables.

*Afzali et al.* [28] analizan en su trabajo la relación entre las distintas condiciones meteorológicas y los niveles de contaminación. Para ello realizan una regresión lineal en la cual analizan la influencia de cada variable climática en los niveles de concentración de PM10. Además, este trabajo, analiza dos modelos distintos de ANN: *Feed Forward Network* (FFN) y *Red Elman* para predecir la concentración de PM10 con un día de adelanto utilizando las variables meteorológicas. El estudio concluye que, aunque la relación entre la concentración de PM10 y los parámetros meteorológicos es débil, los modelos de predicción reflejan una mejora al incluir estas variables.

Por su parte, *Feng et al.* [20] nos presentan, en este artículo un modelo híbrido, para predecir la concentración de PM2,5 con dos días de adelanto, que sobre la tecnología de ANN busca mejorar la precisión de estas mediante el modelado de la trayectoria de las masas de aire modificando sus variables de entrada y, por otra parte, aprovecha las propiedades de la transformación de ondículas continuas [29] (*Wavelet Transformation*), que permite expresar una señal continua en el dominio de la frecuencia (igual que la transformada de Fourier) y del tiempo, para obtener representaciones de la variable dependiente más fáciles de predecir. El modelo es probado en tres condiciones: con las variables tal cual, añadiendo un modelo para definir la trayectoria del viento y, añadiendo una descomposición en *wavelet*.

En el trabajo de *Zeng et al.* [29] encontramos una estructura compuesta por cuatro componentes que nos permite predecir la calidad del aire para las próximas 48 horas, en una estación de monitorización a partir de datos de calidad del aire de la propia estación, de estaciones cercanas, datos meteorológicos y predicciones del clima para las próximas tres días. Estos cuatro componentes, del modelo propuesto, son diseñados con un propósito concreto. Así, un primer componente, denominado en el artículo predictor temporal (*Temporal Predictor*, TP), basado en un modelo de regresión lineal, se encarga de modelar los factores locales de la estación para la que se realiza la predicción. El segundo componente se basa en una red neuronal, se denomina predictor espacial (*Spatial Predictor*, SP), y predice la parte correspondiente a las estaciones cercanas. El tercer componente integra las predicciones realizadas por los anteriores predictores. Así recibe el nombre de agregador de predicción (*Prediction Aggegator*, PA). Este componente construye un árbol de regresión a partir de las salidas de los anteriores componentes, TP y SP, y datos meteorológicos de las diferentes estaciones. Por último, se tiene el componente denominado predictor de inflexión (*Inflector Predictor*, IP), el cual tiene como función la detección de cambios bruscos en la variación de la calidad del aire. Este modelo, así formado, es comparado, para probar su eficiencia, con otros modelos de diversa índole: distintas configuraciones del modelo estadístico *Auto-Regression-Moving-Average* (ARMA); distintos modelos de aprendizaje automático, aplicados de forma separada, como: regresión lineal, redes neuronales y árboles de regresión; un clásico modelo de predicción del clima capaz de predecir la calidad del aire y; el propio modelo al que se le quitan alguno de sus componentes.

*Li et al.* [31] proponen también un modelo espacio-temporal capaz de predecir la calidad del aire en múltiples estaciones a la vez. Para ello construye un modelo compuesto por varios *autoencoders*, estructura básica del *Deep Learning*, apilados (*stacked*) de forma tal que se tienen en cuenta, para cada estación, las condiciones de las demás estaciones que componen la red de monitorización. Como en el trabajo anterior, en este caso también se compara el modelo con el resultado de aplicar otros modelos sobre los datos.

*Akbar Rahimi* [32] construye, en este artículo, un modelo de predicción de la concentración óxido de nitrógeno NOx a partir de datos de calidad del aire y meteorológicos medidos en la ciudad de Tabriz, Irán. Para la construcción del modelo se utilizada una red neuronal *feed-forward*, para la cual se realiza una selección efectiva de las variables de entrada y de los parámetros de configuración propios de la red neuronal.

En el artículo realizado por *Athira et al.* [33] , presentado en la Conferencia Internacional sobre Inteligencia Computacional y Ciencia de los Datos (*International Conference on Computational Intelligence and Data Science*, ICCIDS’2018), se analizan tres modelos de predicción de calidad del aire basados en redes neuronales recurrentes (*Recurrent Neural Networks*, RNN). Estos modelos son: las redes neuronales recurrentes, *Long Short-Term Memory* (LSTM) y, *Gated Recurrent Unit* (GRU). Estas últimas son dos tipos de RNNs mejoradas que tratan de mitigar el problema del descenso del gradiente que se produce en el primero. En el trabajo se incluye una comparativa de los tres modelos con distintas configuraciones que afectan a las prestaciones de cada una de las técnicas utilizadas. Los mejores resultados se obtienen con el modelo GRU.

*Akbarzadeh et al.* [22] en su trabajo otra técnica, distinta a las redes neuronales, para la predicción de contaminantes. Se trata de la técnica de *Support Vector Machine* (SVM), la cual es comparada con las técnicas más ampliamente utilizadas en este campo, tales como: ANN y *Adaptive Neuro-Fuzzy Inference System* (ANFIS). Uno de los puntos críticos aquí es la selección de variables de entrada. Para esto se ha optado por emplear un método de selección de variables denominado *Forward Selecction* (FS), el cual se basa en, partiendo de una ordenación de las variables de entradas, o independientes, según su mejor correlación con la variable dependiente, se añaden progresivamente variables una a una al modelo quedándonos el mínimo número de variables de entradas a partir del cual no se produce mejora sustancial en el resultado del modelo (se utiliza el estadístico R-cuadrado para evaluar el mejor modelo). El artículo concluye que los modelos SVM analizados tienen prestaciones similares a los mejores modelos aplicados al campo en cuestión.

Aunque el trabajo de *Feng et al.* [35] se sale un poco de el propósito de este trabajo fin de máster, lo incluyo aquí por lo novedoso del planteamiento. En este caso, este trabajo, plantea la contaminación generada por la combustión de los pastos en determinadas zonas de China. El estudio planteado lo primero que realiza es identificar los pixeles quemados en imágenes satelitales (MODIS) de una determinada zona de China. Finalmente, a partir de estos datos de pixeles quemados, el trabajo presenta un modelo de predicción de contaminación utilizando una red neuronal de tipo *Back-Propagation Neural Networks* (BPNN).

Sun, Xiaotong & Xu, Wei [36] presentan un modelo, que denominan LSTM-DRSL, que se basa en realizar una combinación de varios modelos LSTM. Como entradas utiliza datos de calidad del aire y meteorológicos proporcionados tanto por estaciones de monitorización instaladas en las ciudades como, también, por datos provenientes de satélites de observación (datos de calidad del aire atmosférico y meteorológicos recogidos por el satélite MODIS). Los datos MODIS son disponibles públicamente para cualquiera que desee utilizarlos en sus trabajos. Además de estos datos, este modelo también incluye, lo cual es algo novedoso, datos de emisión de contaminantes proporcionados por industrias chinas. Como es habitual, para validar el modelo propuesto en este trabajo, también se incluye una comparativa con los modelos más utilizados actualmente.

El siguiente trabajo, realizado por *Sayeed et al.* [37], utiliza una red neuronal convolucional (*Convolutional Neural Network*, CNN) de 5 capas para modelar el nivel de Ozono, O3, en varias estaciones de la ciudad de Texas (USA), y así construir un sistema capaz de predecir el nivel de concentración de este gas con 24 horas de anticipación. Durante el estudio se entrena el modelo con datos entre 2014 y 2016 y, se predice el nivel de O3 para todo el año 2017 con el fin de evaluar la respuesta del modelo a los cambios estacionales en cada estación de observación. El modelo es comparado con otros modelos de aprendizaje automático como: perceptrón multicapa, redes neuronales profundas, redes neuronales recurrentes tipo *Gate Recurrent Unit* y modelos de regresión como Lasso y Ridge.

Mediante la unión de varios modelos LSTM para distintos rangos de predicción y la aplicación de la técnica de *Transfer Learning*, que se basa en la reutilización del conocimiento adquirido en conjuntos de datos o modelos para ser transferido a modelos en otro dominio diferente, el trabajo de *Ma et al.* [19] nos presenta un modelo complejo capaz de predecir la concentración de PM2,5 con una antelación de días, semanas e incluso meses. El LSTM bidireccional es una RNN mejorada y es capaz de aprender de las dependencias a largo plazo de los datos de las series temporales de las secuencias tanto hacia adelante como hacia atrás. Además resuelve el problema del descenso del gradiente que suele aparecer en los modelos RNN. El modelo construido obtiene errores de predicción más bajos que los modelos mejores actuales, sobretodo para predicciones a largo plazo.

Liu et al. [38] aplican la novedosa técnica de Sequence-to-Sequence (Seq2Seq), introducida por Google en 2014, para obtener un modelo de predicción de calidad del aire. Liu et al. realizan algunas modificaciones sobre la técnica Seq2Seq para aumentar la rapidez de aprendizaje del modelo y para reducir la acumulación de errores que ocurre en el modelo original. Al comparar el modelo obtenido con otros modelos del estado del arte se ve que, el modelo obtenido mejora los del modelo Seq2Seq sin modificaciones y, comparado con otros modelos, se obtiene mayor exactitud para predicciones a corto plazo, 2 horas de adelanto.

Otros dos trabajos, presentados recientemente, buscan mejorar los resultados de modelos anteriores mediante la descomposición de la señal descrita por los valores de contaminación analizados, de naturaleza fuertemente no lineal, en otras señales de comportamiento lineal, más fáciles de predecir. Así, tenemos el trabajo de *Fang et al.* [39] que, al igual que el trabajo de *Feng et al.* [20], utilizada la transformación de Wavelet para descomponer la serie temporal del parámetro de calidad del aire analizado en otras señales más simples sobre las que se establece un modelo de redes neuronales de *Back-Propagation*. Las salidas de todos estos modelos se suman, aprovechando la propiedad de adición de las wavelets, para obtener la predicción buscada. El otro trabajo, es el de *Jin et al.* [21]. En este caso, se utiliza la técnica de descomposición de señales denominada *Empirical Mode Decomposition* (EMD). La cual tiene algunas similitudes con las *Wavelets*, aunque en este caso una señal se descompone en los denominados componentes *Intrinsic Mode Function* (IMF) con diferentes características de frecuencia. Este trabajo plantea una estructura de redes neuronales mucho más compleja que la anterior, en la cual se distinguen: una fase de descomposición de la serie temporal de entrada en componentes IMF; una segunda parte de clasificación mediante un modelo *Convolutional Neural Network* (CNN), obteniendo tres clústeres; una tercera fase de predicción, a partir de los grupos/clústeres anteriores, mediante sendos modelos *Gate Recurrent Unit* (GRU); y, una última fase de fusión para obtener la predicción final.

Tal y como hemos afirmado arriba, y como hemos visto en los trabajos descritos, la gran mayoría de estos trabajos buscan encontrar el mejor modelo, empleando distintas técnicas de ML y distintas combinaciones de variables independientes, que mejoren en algún aspecto a sus predecesores. Sin embargo, en este abanico de opciones, también encontramos trabajos que buscan resolver problemáticas bien distintas. Por ejemplo, el trabajo de *Fan et al.* [40], presenta una estructura de redes neuronales que es capaz de realizar la predicción de la calidad del aire (en este caso se predice la concentración de partículas PM2.5) en un esquema espacio-temporal, sin importar la existencia de valores perdidos o inexistentes en alguna de las estaciones de medida. Por su parte, *Honarvar & Sami* [41] o *Hsieh et al.* [30], con el objetivo de encontrar las mejores ubicaciones para instalar una estación de medida o, también, con el propósito de poder ahorrarse la necesidad de gastar dinero público en equipos que, a día de hoy, son muy caros, proponen, en su respectivos trabajos, un esquema de BigData que es capaz de predecir la cantidad de partículas PM10, en determinadas zonas, sin necesidad de contar con datos de contaminación previos. En el primer caso, el modelo propuesto utiliza datos de tráfico en las principales arterias de una ciudad (Aarhus, Alemania), identificación de puntos de interés (POI) en la ciudad que puedan relacionarse con la variabilidad de los contaminantes en un determinado tiempo y lugar, y, datos medidos por las estaciones meteorológicas (precipitación, humedad, temperatura, velocidad y dirección del viento, etc.). En el trabajo de *Hsieh et al.* [30], se infieren los valores a partir de los resultados para otras estaciones de medida cercanas.

Aunque puede que se salgan un poco del propósito de este TFM, *Song et al.* [42] y *Rahimi & Sajedi* [43] presentan, en sus respectivos trabajos, dos proyectos muy interesantes de aplicación de técnicas de inteligencia artificial (AI) en el campo de la predicción de la calidad del aire. En el primero de estos artículos, *Song et al.* [42], se realiza un estudio que trata de buscar patrones en las emisiones de contaminantes producidas por el tráfico observado en distintos puntos de la ciudad. Para esto, primero desarrolla un sistema de monitorización de vehículos, el cual, mediante cámaras de video y software de reconocimiento de objetos en imágenes, realiza una clasificación de estos objetos en distintos grupos/clases de vehículos (motocicletas, turismos, camiones grandes y pequeños, etc.). Realiza un conteo de estos vehículos, según cada grupo, y estima la emisión de contaminantes en distintas zonas de la ciudad. De esta forma, el sistema es capaz de evaluar las emisiones producidas por el tráfico en distintos puntos y a diferentes horas del día. La estimación de contaminantes producida en cada momento se realiza por conteo de los distintos tipos de vehículos y la asignación de un coeficiente de emisión a cada grupo. En el otro artículo mencionado, *Rahimi & Sajedi* [43] presentan una herramienta de móvil que mediante una simple foto realizada hacia el cielo es capaz de indicarnos la calidad del aire. En este caso, este estudio simplemente es una prueba de concepto, pues plantea muchos inconvenientes que harían inviable este sistema, como por ejemplo, la existencia de nubes o condiciones climáticas desfavorables.

# 3. Resto de capítulos

En estos capítulos, hay que describir los aspectos más relevantes del diseño y desarrollo del proyecto, así como de los productos obtenidos. **La estructuración de los capítulos puede variar según el tipo de Trabajo**.

En cada apartado es muy importante describir las alternativas posibles, los criterios utilizados para tomar decisiones y la decisión tomada.

En caso de que corresponda, se incluirá un apartado de “Valoración económica del trabajo”. Este apartado indicará los gastos asociados al desarrollo y mantenimiento del trabajo, así como los beneficios económicos obtenidos. Hacer un análisis final sobre la viabilidad del producto.

# 4. Conclusiones

Este capítulo tiene que incluir:

* Una descripción de las conclusiones del trabajo: Qué lecciones se han aprendido del trabajo?
* Una reflexión crítica sobre el logro de los objetivos planteados inicialmente: Hemos logrado todos los objetivos? Si la respuesta es negativa, por qué motivo?
* Un análisis crítico del seguimiento de la planificación y metodología a lo largo del producto: Se ha seguido la planificación? La metodología prevista ha sido la adecuada? Ha habido que introducir cambios para garantizar el éxito del trabajo? Por qué?
* Las líneas de trabajo futuro que no se han podido explorar en este trabajo y han quedado pendientes.

# 5. Glosario

**aire ambiente**

el aire exterior de la troposfera, con exclusión de los lugares de trabajo definidos en la Directiva 89/654/CEE, cuando se apliquen las disposiciones sobre salud y seguridad en el trabajo, a los que el público no tiene acceso habitualmente

**Business-As-Usual**

Se trata de un modelo basado en establecer una línea base sobre la ejecución normal de un fenómeno y determinar las desviaciones sobre este.

**compuestos orgánicos volátiles (COV)**

compuestos orgánicos de fuentes antropogénicas y biogénicas, con excepción del metano, capaces de producir oxidantes fotoquímicos por reacción con los óxidos de nitrógeno bajo el efecto de la luz solar

**contaminante**

toda sustancia presente en el aire ambiente que pueda tener efectos nocivos para la salud humana y el medio ambiente en su conjunto

**MM5**

Del inglés: Fifth-Generation Pen State/NCAR Mesoscale Model, MM5 es un modelo regional a mesoescala para la creación de previsiones meteorológicas y de proyecciones climáticas.

**óxidos de nitrógeno**

suma en partes por mil millones en volumen de monóxido de nitrógeno (óxido nítrico) y dióxido de nitrógeno, expresada en unidades de concentración másica de dióxido de nitrógeno (μg/m3)

**PM10**

partículas que pasan a través del cabezal de tamaño selectivo definido en el método de referencia para el muestreo y la medición de PM10 de la norma EN 12341, para un diámetro aerodinámico de 10 μm con una eficiencia de corte del 50 %

**PM2,5**

partículas que pasan a través del cabezal de tamaño selectivo definido en el método de referencia para el muestreo y la medición de PM2,5 de la norma EN 14907, para un diámetro aerodinámico de 2,5 μm con una eficiencia de corte del 50 %

# 6. Bibliografía

x

|  |  |
| --- | --- |
| 1. | BOE. Ley 16/2002, de 1 de julio, de prevención y control integrados de la contaminación. Gobierno de España; 2002. |
| 2. | BOE. Ley 34/2007, de 15 de noviembre, de calidad del aire y protección de la atmósfera. Gobierno de España; 2007. |
| 3. | WHO. Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de azufre. Actualización mundial 2005 (OMS) OMdlS, editor.; 2006. |
| 4. | Ayuntamiento de Barcelona. ZBE Rondes de Barcelona. [Online].; 2020. Available from: <https://www.zbe.barcelona/>. |
| 5. | Torres ME. Supermanzanas de Barcelona: el exitoso plan anticoches que arrancó con la oposición vecinal. El País. 2019 Oct 9. |
| 6. | Querol X, Viana M, Moreno T, Alastuey A. Bases científico técnicas para un Plan Nacional de Mejora de la Calidad del Aire. Consejo Superior de Investigaciones Científicas (CSIC); 2012. |
| 7. | European Union. Directiva 2008/50/CE relativa a la calidad del aire ambiente y a una atmósfera más limpia en Europa. Parlamento Europeo y Consejo de la Unión Europea; 2008. |
| 8. | Rodríguez Gómez D, Valldeoriola Roquet J. Metodología de la investigación: Universitat Oberta de Catalunya (UOC); 2019. |
| 9. | Rodríguez-Sánchez F, Pérez-Luque AJ, Bartomeus I, Varela S. Ciencia reproducible: qué, por qué, cómo. AEET - Revista Ecosistemas 25(2) [Mayo-Agosto 2016]. 2016;: p. 83-92. |
| 10. | Gironés J, Casas J, Minguillón J, Caihuelas R. Minería de datos - modelos y algoritmos. Primera edición ed.: Editorial UOC; 2017. |
| 11. | Wang L, Jang C, Zhang Y, Wang K, Zhang Q, Streets D, et al. Assessment of air quality benefits from national air pollution control policies in China. Part I: Background, emission scenarios and evaluation of meteorological predictions. Atmospheric Environment. 2010; 44(2010): p. 3442-3448. <https://doi.org/10.1016/j.atmosenv.2010.05.051>. |
| 12. | Wang L, Jang C, Zhang Y, Wang K, Zhang Q, Streets D, et al. Assessment of air quality benefits from national air pollution control policies in China. Part II: Evaluation of air quality predictions and air quality benefits assessment. Atmospheric Environment. 2010; 44(2010): p. 3449-3457. <https://doi.org/10.1016/j.atmosenv.2010.05.058>. |
| 13. | Fei T, Shuang-Qing X. Definition of Business as Usual and Its Impacts on Assessment of Mitigation Efforts. Advances in climate change research. 2012; 3(4): p. 212-219. <10.3724/SP.J.1248.2012.00212>. |
| 14. | Anthes RA, Warner TT. Development of Hydrodynamic Models Suitable for Air Pollution and Other Mesometeorological Studies. Monthly Weather Review. 1978 August; 106(8): p. 1045-78. |
| 15. | Grell GA, Dudhia J, Stauffer DR. A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note. ; 1994. Report No.: NCAR/TN-398þSTR. |
| 16. | Dudhia J. MM5 VERSION 3.7 (THE FINAL VERSION). NCAR/MMM; 2005. |
| 17. | Zhang Y, Bocquet M, Mallet V, Seigneur C, Baklanov A. Real-time air quality forecasting, part I: History, techniques, and current status. Atmospheric Environment. 2012; 60(2012): p. 632-655. <http://dx.doi.org/10.1016/j.atmosenv.2012.06.031>. |
| 18. | Zhang Y, Bocquet M, Mallet V, Seigneur C, Baklanov A. Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects. Atmospheric Environment. 2012; 60(2012): p. 656-676. <https://doi.org/10.1016/j.atmosenv.2012.02.041>. |
| 19. | Fiore AM, Naik V, Leibensperger EM. Air Quality and Climate Connections. Journal of the Air & Waste Management Association. 2015; 65(6): p. 645-685. <http://dx.doi.org/10.1080/10962247.2015.1040526>. |
| 20. | Zhang K, Batterman S. Air pollution and health risks due to vehicle traffic. Science of the Total Environment. 2013; 450-451(2013): p. 307-316. <http://dx.doi.org/10.1016/j.scitotenv.2013.01.074>. |
| 21. | Yuan C, Ng E, Norford LK. Improving air quality in high-density cities by understanding the relationship between air pollutant dispersion and urban morphologies. Building and Environment. 2014; 71(2014): p. 245-258. <http://dx.doi.org/10.1016/j.buildenv.2013.10.008>. |
| 22. | Vos PEJ, Maiheu B, Vankerkom J, Janssen S. Improving local air quality in cities: To tree or not to tree? Environmental Pollution. 2013; 183(2013): p. 113-122. <http://dx.doi.org/10.1016/j.envpol.2012.10.021>. |
| 23. | Ma J, Cheng JCP, Lin C, Tan Y, Zhang J. Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. Atmospheric Environment. 2019; 214(2019). <https://doi.org/10.1016/j.atmosenv.2019.116885>. |
| 24. | Feng X, Li Q, Zhu Y, Hou J, Jin L, Wang J. Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. Atmospheric Environment. 2015; 107(2015): p. 118-128. <http://dx.doi.org/10.1016/j.atmosenv.2015.02.030>. |
| 25. | Jin XB, Yang NX, Wang XY, Bai YT, Su TL, Kong JL. Deep Hybrid Model Based on EMD with Classification by Frequency Characteristics for Long-Term Air Quality Prediction. Mathematics. 2020; 8(214). <https://doi.org/10.3390/math8020214>. |
| 26. | Akbarzadeh A, Vesali Naseh MR, NodeFarahani M. Carbon Monoxide Prediction in the Atmosphere of Tehran Using Developed Support Vector Machine. Pollution. 2019; 6(1): p. 43-57. <https://doi.org/10.22059/poll.2019.279412.618>. |
| 27. | Singh KP, Gupta S, Kumar A, Shukla SP. Linear and nonlinear modeling approaches for urban air quality prediction. Science of the Total Environment. 2012; 426(2012): p. 244-255. <https://doi.org/10.1016/j.scitotenv.2012.03.076>. |
| 28. | Mahanta S, Ramakrishnudu T, Jha RR, Tailor N. Urban Air Quality Prediction Using Regression Analysis. 2019 IEEE Region 10 Conference (TENCON 2019). 2019;: p. 1118-1123. <https://doi.org/10.1109/tencon.2019.8929517>. |
| 29. | Corani G. Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning. Ecological Modelling. 2005 Jan 3; 185(2005): p. 513-529. <https://doi.org/10.1016/j.ecolmodel.2005.01.008>. |
| 30. | Voukantsis D, Karatzas K, Kukkonen J, Räsänen T, Karppinen A, Kolehmainen M. Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2.5 concentrations using artificial neural networks, in Thessaloniki and Helsinki. Science of the Total Environment. 2011; 409(2011): p. 1266-1276. <https://doi.org/10.1016/j.scitotenv.2010.12.039>. |
| 31. | Russo A, Raischel F, Lind PG. Air quality prediction using optimal neural networks with stochastic variables. Atmospheric Environment. 2013; 79(2013): p. 822-830. <http://dx.doi.org/10.1016/j.atmosenv.2013.07.072>. |
| 32. | Afzali A, Rashid M, Sabariah B, Ramli M. PM10 Pollution: Its Prediction and Meteorological Influence in PasirGudang, Johor. Earth and Environmental Science. 2014; 18(2014): p. 012100. <http://dx.doi.org/10.1088/1755-1315/18/1/012100>. |
| 33. | Vázquez Gerardo R. La transformada de ondícula continua. 2016.. |
| 34. | Zheng Y, Yi X, Li M, Li R, Shan Z, Chang E, et al. Forecasting Fine-Grained Air Quality Based on Big Data. KDD '15. 2015; 11-14(2015). <http://dx.doi.org/10.1145/2783258.2788573>. |
| 35. | Li X, Peng L, Hu Y, Shao J, Chi T. Deep learning architecture for air quality predictions. Environ Sci Pollut Res. 2016. <https://doi.org/10.1007/s11356-016-7812-9>. |
| 36. | Rahimi A. Short-term prediction of NO2 and NOx concentrations using multilayer perceptron neural network: a case study of Tabriz, Iran. Ecological Processes. 2017; 6(4). <https://doi.org/10.1186/s13717-016-0069-x>. |
| 37. | Athira V, Geetha P, Vinayakumar R, Soman KP. DeepAirNet: Applying Recurrent Networks for Air Quality Prediction. Procedia Computer Science. 2018; 132(2018): p. 1394-1403. <https://doi.org/10.1016/j.procs.2018.05.068>. |
| 38. | Xu Feng TMF, Cao H, Tian H, Fan Q, Chen X. Neural network predictions of pollutant emissions from open burning of crop residues: Application to air quality forecasts in southern China. Atmospheric Environment. 2019; 204(2019): p. 22-31. <https://doi.org/10.1016/j.atmosenv.2019.02.002>. |
| 39. | Sun X, Xu W. Deep Random Subspace Learning: A Spatial-Temporal Modeling Approach for Air Quality Prediction. Atmosphere 2019. 2019; 10(560). <https://doi.org/doi:10.3390/atmos10090560>. |
| 40. | Sayeed A, Choi Y, Eslami E, Lops Y, Roy A, Jung J. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. Neural Networks. 2019; 121(2020): p. 396-408. <https://doi.org/10.1016/j.neunet.2019.09.033>. |
| 41. | Liu B, Yan S, Li J, Qu G, Li Y, Lang J, et al. A Sequence-to-Sequence Air Quality Predictor Based on the n-Step Recurrent Prediction. IEEE Access. 2019; 7(2019). <https://doi.org/10.1109/ACCESS.2019.2908081>. |
| 42. | Fang Z, Zhang L, Huang Y. A Novel BP Neural Network with Wavelet Transform Inputs for Air Quality Index Prediction. IOP Conference Series: Materials Science and Engineering. 2020; 735(012059). <https://doi.org/10.1088/1757-899X/735/1/012059>. |
| 43. | Fan J, Li Q, Hou J, Feng X, Karimian H, Lin S. A Spatiotemporal Prediction Framework for Air Pollution Based on Deep RNN. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2017; Volume IV-4/W2, 2017. <https://doi.org/10.5194/isprs-annals-IV-4-W2-15-2017>. |
| 44. | Ali Reza Honarvar AS. Towards Sustainable Smart City by Particulate Matter Prediction Using Urban Big Data, Excluding Expensive Air Pollution Infrastructures. Big Data Research. 2018;: p. 1-10. <https://doi.org/10.1016/j.bdr.2018.05.006>. |
| 45. | Hsieh HP, Lin SD, Zheng Y. Inferring Air Quality for Station Location Recommendation Based on Urban Big Data. KDD '15. 2015;(2015): p. 10-13. <http://dx.doi.org/10.1145/2783258.2783344>. |
| 46. | Song J, Zhao C, Lin T, Li X, Prishchepov AV. Spatio-temporal patterns of traffic-related air pollutant emissions in different urban functional zones estimated by real-time video and deep learning technique. Journal of Cleaner Production. 2019; 238(2019). <https://doi.org/10.1016/j.jclepro.2019.117881>. |
| 47. | Rahimi SA, Sajedi H. Monitoring air pollution by deep features and extreme learning machine. Journal of Experimental & Theoretical Artificial. 2019. <https://doi.org/10.1080/0952813X.2019.1572658>. |
| 48. | Xavier Q. La calidad del aire en las ciudades. Un reto mundial fenosa Fg, editor.; 2018. |
| 49. | EEA. Air quality in Europe. European Environment Agency; 2018. |
| 50. | MITECO. Visor a nivel nacional de contaminantes del aire del Ministerio de Transición Ecológica. [Online].; 2020 [cited 2020 02 20. Available from: <https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/calidad-del-aire/visor/>. |
| 51. | Orío Hernández A, Pallarés Querol M. Análisis de la calidad del aire en España. Evolución 2001-2012. Gobierno de España, Ministerio de Agricultura, Alimentación y Medio Ambiente; 2013. |
| 52. | U.S. EPA. A Guide to Air Quality and. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards; 2014. |
| 53. | U.S. EPA. Guidelines for Developing an Air Quality. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards; 2003. |
| 54. | U.S. EPA. Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI). U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards; 2018. |
| 55. | Carslawa DC, Ropkins K. openair d An R package for air quality data analysis. Environmental Modelling & Software. 2011; 27-28(2012): p. 52-61. <https://doi.org/10.1016/j.envsoft.2011.09.008>. |
| 56. | Berrocal VJ, Guan Y, Muyskens A, Wang H, Reich BJ, Mulholland JA, et al. A comparison of statistical and machine learning methods for creating national daily maps of ambient PM2:5 concentration. Atmospheric Environment. 2019;(2019). <https://doi.org/10.1016/j.atmosenv.2019.117130>. |

x

# 7. Anexos

Listado de apartados que son demasiado extensos para incluir dentro de la memoria y tienen un carácter autocontenido (por ejemplo, manuales de usuario, manuales de instalación, etc.)

Dependiente del tipo de trabajo, es posible que no haya que añadir ningún anexo.