



Universidad Autónoma de Querétaro

Facultad de Ingeniería

Maestría en Ciencias en
Inteligencia Artificial

Modelación y predicción de excedencias de partículas PM10 utilizando redes recurrentes.

TESIS

Que como parte de los requisitos para obtener el grado de
Maestro en Ciencias en Inteligencia Artificial

Presenta:

Julio Alberto Ramírez Montañez

Dirigido por:

Dr. Marco Antonio Aceves Fernández

SINODALES

Dr. Marco Antonio Aceves Fernández
Presidente

Dr. Jesús Carlos Pedraza Ortega
Secretario

Dr. Efrén Gorrostieta Hurtado
Vocal

Dr. Saúl Tovar Arriaga
Suplente

Dr. Luis Ricardo Corral Velazquez
Suplente

Centro Universitario
Querétaro, QRO
México.
Febrero 2020

Dirección General de Bibliotecas UAQ

Dirección General de Bibliotecas UAQ

This thesis is dedicated to my family

Dirección General de Bibliotecas UAQ

Agradecimientos

Agradezco enormemente a mis profesores que durante todo mi proceso académico me brindaron consejos y motivaciones, exigiéndome a dar lo mejor de mi en cada asignatura, así como en mi propio proyecto de investigación.

Especialmente agradezco a mis padres, quienes siempre me han apoyado y creído en mis capacidades para lograr los objetivos que me he propuesto.

Sin olvidar a mis amigos quienes de igual forma me apoyaron y animaron siempre.

Dirección General de Bibliotecas UAQ

Dirección General de Bibliotecas UAQ

Abstract

This thesis shows the implementation of two deep neural networks (LSTM) one for the modeling of polluting air particles called PM10, while the second network is used to detect the conditions that precede a leave. Since the data used are records of the city of Mexico, the respective Mexican standard was taken as a reference. This is because the monitoring of air quality is a topic of current interest, PM10 particles are air pollutants made up of different components.

The results obtained demonstrate that it is possible to anticipate an excess of PM10 particles 24, 48 or 72 hours before it occurs. Similarly, different algorithms were implemented to validate the results obtained, focusing on the preprocessing of the initial data and the modeling and prediction of the exceedances.

Keywords: Air pollution, LSTM Recurrent Neural Network, Deep Neural Network, PM10.

Dirección General de Bibliotecas UAQ

Resumen

La presente tesis muestra la implementación de dos redes neuronales profundas (LSTM) una para la modelación de las partículas aéreas contaminantes llamadas PM10, mientras que la segunda red es utilizada para detectar las condiciones que anteceden a una excedencia. Dado que los datos utilizados son registros de la ciudad de México, se tomo de referencia la norma mexicana respectiva. Lo anterior debido a que el monitoreo de la calidad del aire es un tema de interés actual, las partículas PM10 son contaminantes aéreos constituidas de diferentes componentes.

Los resultados obtenidos demuestran que es posible anticipar una excedencia de partículas PM10 con 24, 48 o 72 horas previas a que ocurra. De igual manera se implemento distintos algoritmos para validar los resultados obtenidos, enfocándose en el preprocesamiento de los datos iniciales y en la modelación y predicción de las excedencias.

Palabras claves: Contaminación aérea, Red Neuronal Recurrente LSTM, Redes neuronales profundas, PM10.

Dirección General de Bibliotecas UAQ

Índice general

Agradecimientos

Abstract

I

Resumen

III

1. Introducción

1

1.1. Contaminación aérea	1
1.2. Inteligencia artificial	2
1.3. Estado del arte	2
1.4. Descripción del problema	4
1.5. Justificación	5
1.6. Hipótesis	6
1.7. Objetivos	6
1.7.1. Objetivo general	6
1.7.2. Objetivos particulares	6

2. Marco teórico.

7

2.1. Material particulado	7
2.1.1. Partículas PM10	8
2.1.2. Excedencias	8
2.1.3. Daños a la salud	9
2.2. Inteligencia artificial	10
2.2.1. Aprendizaje de maquinas	10
2.2.2. Aprendizaje profundo	11
2.3. Redes neuronales artificiales	11
2.3.1. Redes neuronales recurrentes	13
2.3.2. Red neuronal recurrente LSTM	14
2.3.3. Funciones de activación	16

3. Metodología

19

3.1. Etapa uno de la implementación de la red LSTM	21
3.1.1. Adquisición y preprocesamiento de los datos.	21
3.1.2. Red neuronal de modelación.	24
3.1.3. Determinación de excedencias.	25

3.1.4. Red neuronal de predicción.	26
3.1.5. Evaluación continua.	27
3.2. Etapa dos herramientas de comparación	28
3.2.1. Imputación múltiple por ecuación encadenada	28
3.2.2. Maquinas de soporte vectorial	28
3.2.3. Red Perceptrón Multicapa	30
4. Resultados y Discusión	33
4.1. Resultados	33
5. Conclusiones	39
5.1. Trabajos futuros	39
Referencias	41
Apéndices	44

Dirección General de Bibliotecas UAQ

Índice de figuras

1.1. Contaminación en México (Tomada de Instituto Nacional de Estadística y Geografía).	1
1.2. Ejemplo del comportamiento medio de las concentraciones de partículas PM10 a lo largo del año 2017 en la ciudad de México. (Adaptada de los registros de la base de datos RAMA).	5
2.1. Comparación entre las partículas PM10, PM2.5 y un cabello humano (Adaptada de Concepción 2015).	8
2.2. Ejemplo del origen de las PM (natural y artificial) (Adaptada de Stadlober 2008).	8
2.3. Lugar de daño provocado por las diferentes partículas ambientales (Tomada de Terashima 1997).	9
2.4. Breve línea del tiempo (Adaptada de LeCun 2015).	11
2.5. Comparación entre una neurona artificial y una natural (Tomada de Hopfield 1988).	12
2.6. Red neuronal básica (Adaptada de Hassoun 1995).	13
2.7. Red Neuronal Recurrente (Tomada de Bianchi 2017).	14
2.8. Neurona básica LSTM (Adaptada de Hua 2019).	15
2.9. Función de activación sigmoide (Tomada de Abraham 2005).	17
2.10. Función de activación ReLU (Tomada de Abraham 2005).	17
2.11. Función de activación SELU (Tomada de Ramachandran 2017).	18
3.1. Metodología implementada.	20
3.2. Metodología de comparación de resultados.	20
3.3. Localización de las estaciones de la Red Automática de Monitoreo Atmosférico (RAMA)(Tomada de la base de datos RAMA).	21
3.4. Propuesta de reemplazo de los valores atípicos.	23
3.5. Concatenación de los datos anuales.	24
3.6. Red neuronal LSTM.	25
3.7. Reducción de datos.	26
3.8. Red de predicción.	27
3.9. Red propuesta de forma completa.	27
3.10. Ejemplos de hiperplanos utilizados por una SVM (Tomada de Betancourt 2005).	29
3.11. Red MLP general. Adaptada de Wilamowski 2009.	31
4.1. Modelación de enero del 2017, utilizando la red LSTM propuesta.	33
4.2. Modelación de enero del 2016, utilizando la red LSTM propuesta.	34
4.3. Modelación de enero del 2018, utilizando la red LSTM propuesta.	34

4.4. Valores medios del año 2016.	35
4.5. Valores medios del año 2017.	36
4.6. Valores medios del año 2018.	37
4.7. Resultados de la evaluación continua de la red propuesta junto con sus comparaciones.	38

Dirección General de Bibliotecas UAQ

Índice de tablas

1.1. Tabla comparativas de diferentes publicaciones relacionadas con el estudio de partículas contaminantes en especial PM10.	3
2.1. Clasificación por tamaño del material particulado (Adaptada de Concepción 2015). . .	7
2.2. Clasificación general del material particulado (Adaptada de Davidson 2005).	8
2.3. Tabla de comparación entre la norma mexicana y lo establecido por la OMS (Adaptada del Informe de calidad de aire 2016 CDMX).	9
2.4. Tabla de operación de la compuerta lógica XOR (Tomada de Hopfield 1988).	13
2.5. Comparación entre una red neuronal simple y una red neuronal recurrente (Adaptada de Bianchi 2017).	14
3.1. Ejemplo de una tabla de registros de partículas PM10 del año 2018.	22
4.1. Resultados de modificar el tamaño de muestra con el que se entrena la red LSTM. . .	34

Dirección General de Bibliotecas UAQ

Introducción

1.1. Contaminación aérea

La contaminación aérea o atmosférica es la presencia de materiales o gases, denominados material particulado, los cuales implican algún riesgo o daño para las personas y bienes de cualquier naturaleza. Dicha contaminación es provocada por emisiones de los automóviles, polvo, polen, esporas de moho, entre otros (Davidson, Phalen, y Solomon, 2005).

Desde la segunda mitad del siglo XVIII, con la revolución industrial, el desarrollo de diversas fábricas, el desarrollo del transporte y el uso de combustibles fósiles han provocado el incremento de diversos gases y materiales suspendidos en la atmósfera (Brunekreef y Holgate, 2002).

En la Figura 1.1 se observa un gráfico circular del porcentaje de las diferentes áreas de contaminación en México, siendo la contaminación del aire la más alarmante.

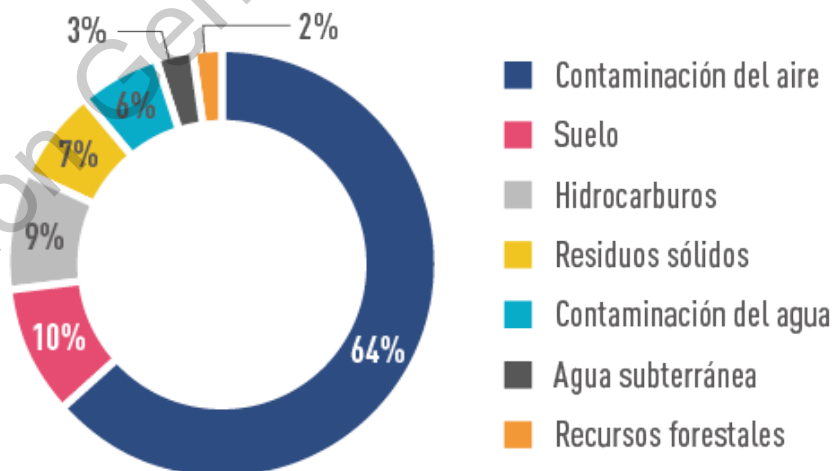


Figura 1.1: Contaminación en México (Tomada de Instituto Nacional de Estadística y Geografía).

1.2. Inteligencia artificial

El termino “inteligencia artificial” fue introducido en 1956 en la conferencia de Dartmouth por John McCarthy. Quien la definió como la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligente; siendo bien recibido en la comunidad investigadora (Buchanan, 2005).

La inteligencia artificial (AI ¹) es la combinación de algoritmos estructurados con el propósito de crear máquinas con capacidades similares a las humanas. Cuyos resultados aun se visualizan lejanos, pero sus avances están presentes en nuestro día a día (Pennachin y Goertzel, 2007).

Stuart Russell y Peter Norvig, clasifican la inteligencia artificial en 4 tipos (Russell y Norvig, 2016):

- **Sistemas que piensan como humanos:** tomando como ejemplo las redes neuronales, estos algoritmos automatizan actividades como lo es el aprendizaje y toma de decisiones.
- **Sistemas que actúan como humanos:** estas son máquinas que realizan tareas de forma similar a como lo haría una persona, un ejemplo son los robots.
- **Sistemas que piensan racionalmente:** englobando a los sistemas expertos, los cuales son capaces de emular el pensamiento lógico racional de los humanos.
- **Sistemas que actúan racionalmente:** en esta clasificación de forma ideal pero aun se esta trabajando en sus inicios; se trata de agentes inteligentes, los cuales imitan de manera racional el comportamiento humano.

Sus orígenes se pueden remontar a los estudios de los antiguos griegos sobre el raciocinio. Pero no fue hasta 1315, cuando Ramon Llull en su libro *Ars magna* tuvo la idea de que el razonamiento podía ser efectuado de manera artificial. En 1936 Alan Turing diseñó formalmente una Máquina universal que demuestra la viabilidad de un dispositivo físico para implementar cualquier cómputo formalmente definido (Buchanan, 2005). Herbet Simon, Allen Newell y J.C. Shaw en 1955, desarrollaron el primer lenguaje de programación orientado para resolver problemas (Buchanan, 2005). A mediados de los 60, surgieron los sistemas expertos, los cuales eran capaces de predecir la probabilidad de una solución dependiendo de un conjunto de condiciones (Buchanan, 2005). Y así sucesivamente surgieron avances hasta la actualidad, existiendo diversos algoritmos que emulan ciertas capacidades de aprender o identificar patrones.

1.3. Estado del arte

Actualmente se ha buscado el desarrollo de modelos que ayuden a entender y predecir el comportamiento del material particulado, sin embargo como punto crítico es determinar las condiciones que causan un nivel superior a lo establecido por normas, denominado excedencia. En la Tabla 1.1 se muestra trabajos donde el estudio del comportamiento del material particulado es el objetivo principal, tanto el daño que provoca en la salud como las condiciones de su comportamiento.

¹Por sus siglas del inglés Artificial Intelligence

Tabla 1.1: Tabla comparativas de diferentes publicaciones relacionadas con el estudio de partículas contaminantes en especial PM10.

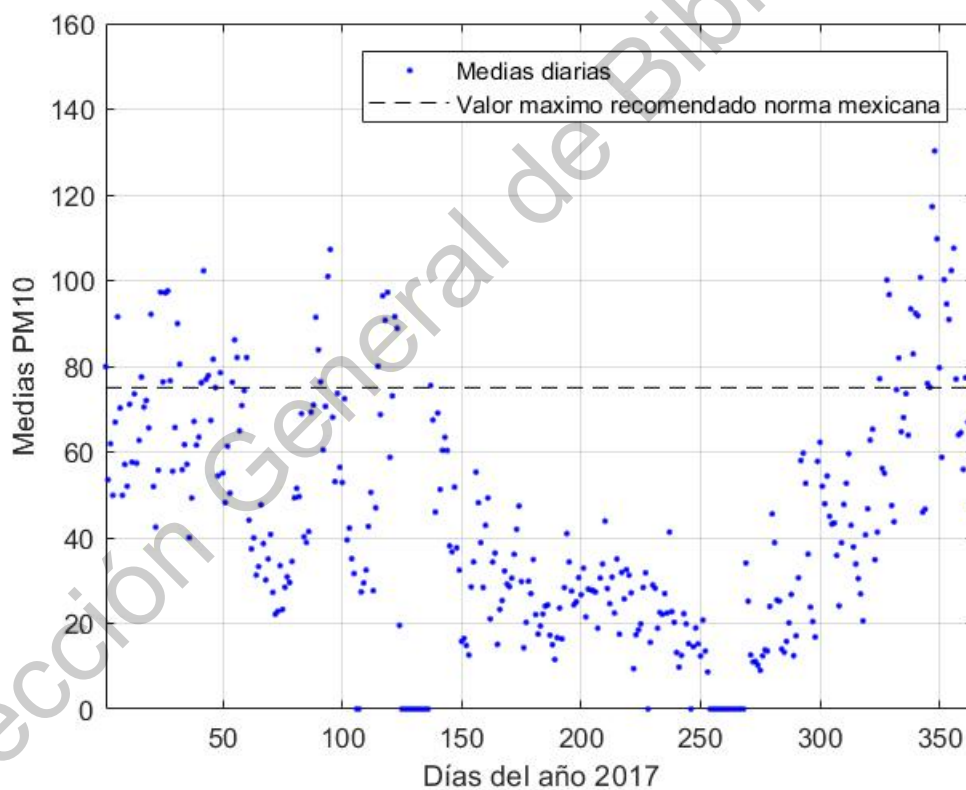
Software	Objetivo	Partícula estudiada	Título	Año	Referencia
R studio	Comparar las concentraciones de PM10 en tres ciudades distintas.	PM10	Variability of PM10 in industrialized-urban areas. New coefficients to establish significant differences between sampling points	2018	(Vicente et al., 2018)
Air-Q2.2.3	Evaluar los impactos en la salud de la exposición a PM10 y SO2	PM10 y SO2	Human health risk assessment due to ambient PM10 and SO2 by an air quality modeling technique	2017	(Khaniabadi et al., 2017)
Phyton	Detectar excedencias de las partículas PM10 detectando las condiciones previas	PM10	PM10 Density Forecast Model Using Long Short Term Memory	2017	(Park et al., 2017)
No mencionado	Generar un modelo matematico robusto el cual puede predecir excedencias de SO2	SO2 y NOx	Three hours ahead prevision of SO2 pollutant concentration using an Elman neural based forecaster	2008	(Brunelli et al., 2008)
No mencionado	En Base a datos previos diseñar un modelo robusto para el analisis de las particulas estudiadas	PM10 y PM2.5	Recursive neural network model for analysis and forecast of PM10 and PM2.5	2008	(Biancofiore et al., 2017)

1.4. Descripción del problema

En las América, 93 000 defunciones anuales en países de ingresos bajo y medios (LMIC) y 44 000 en países de ingresos altos (HI) son atribuidas a la contaminación atmosférica (OPS, 2019). Lo anterior indica, si se sobrepasa el nivel seguro de partículas ambientales, los habitantes de una determinada región corren riesgos de enfermarse gravemente, dichas enfermedades reducirán significativamente la calidad de vida de los individuos afectados y en casos extremos provocar su fallecimiento (Brunekreef y Holgate, 2002).

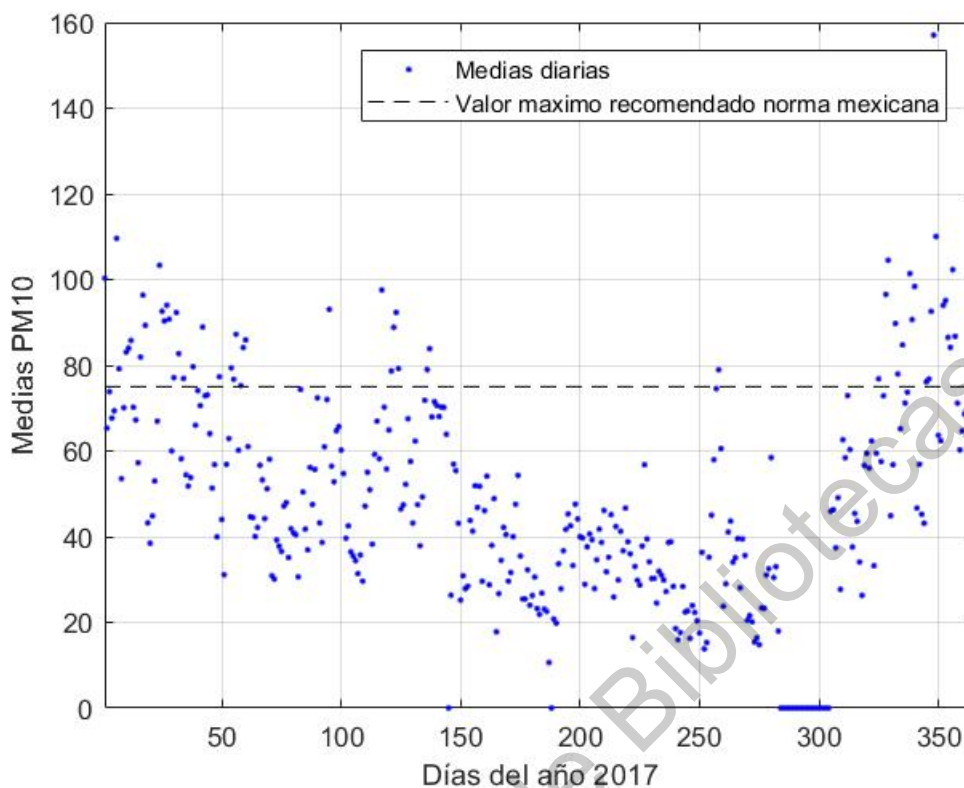
La ciudad de México, México y sus alrededores poseen estaciones de monitoreo; las cuales forman parte de la Red Automática de Monitoreo Atmosférico (RAMA). Dichas estaciones solo son capaces de medir las concentraciones de diversos contaminantes aéreos (CDMX, 2019), pero no son capaces de identificar las condiciones que anteceden a una excedencia.

Como ejemplo en el 2017 dos de las estaciones registraron valores medios diarios que sobrepasan lo establecido por la norma mexicana ² (la cual establece una media diaria de $75\mu\text{g}/\text{m}^3$) como se muestra en la Figura 1.2.



(a) Estación de monitoreo ACO.

²NOM-025-SSA1-2014



(b) Estación de monitoreo TLI.

Figura 1.2: Ejemplo del comportamiento medio de las concentraciones de partículas PM10 a lo largo del año 2017 en la ciudad de México. (Adaptada de los registros de la base de datos RAMA).

1.5. Justificación

La contaminación aérea generada afecta diversos sectores uno de ellos es el referente a la salud. Estudios han demostrado la correspondencia entre la contaminación atmosférica y la mortalidad por enfermedades cardiovasculares y respiratorias (Davidson et al., 2005; Katsouyanni et al., 2001). Datos de la OMS ³ demuestran que en el 2016, el 91 % de la población mundial vivía en lugares donde no se respetaban las direcciones de la OMS sobre la calidad del aire (OMS, 2018).

Actualmente existe una escasez de modelos robustos capaces de detectar las condiciones que anteceden a una situación de excedencia de contaminantes atmosféricos; esto da a entender que para detectar particularmente excedencias de partículas PM10 no existen o simplemente no se han logrado obtener resultados favorables, impidiendo alertar a la población del lugar para que tome sus precauciones necesarias para evitar problemas de salud.

³Organización Mundial de la Salud.

1.6. Hipótesis

Existen condiciones particulares que favorecen o inician una excedencia de partículas PM10, las cuales se podrán catalogar y predecir, permitiendo realizar una alerta temprana. El uso de una topología de red neuronal recurrente permitirá predecir dichas condiciones particulares con una buena tasa de éxito y sera capaz de adaptarse a los constantes cambios ambientales.

1.7. Objetivos

1.7.1. Objetivo general

Como objetivo general se busca identificar con el mayor tiempo de anticipación condiciones de excedencias de PM10 en la ciudad de México, implementando una topología de Red Neuronal Recurrente, la cual permitirá tomar precauciones y medidas para disminuir las partículas en suspensión producidas por el sector industrial.

1.7.2. Objetivos particulares

- Generar un modelo robusto que permita predecir el comportamiento de excedencias de partículas PM10.
- Obtener un porcentaje aceptable de aciertos de excedencias utilizando los dos últimos años (2017 y 2018) de la base de datos RAMA, como medida de comparación.
- Detectar las condiciones presentes en una excedencia de PM10, tales como la humedad, la temperatura, dirección del viento e intensidad del mismo.
- Diseñar un modelo que detecte los inicios de condiciones de excedencias.

Marco teórico.

2.1. Material particulado

El material particulado (PM)⁴ o partículas en suspensión, es un contaminante aéreo el cual se conforma de una gran variedad de diminutos cuerpos sólidos o de gotas de líquidos, cuyos orígenes varían entre uno natural y uno artificial (Davidson et al., 2005). Lo anterior indica que sus características fisicoquímicas son muy diferentes entre sí provocando distintos problemas de salud.

En general, la composición del material particulado varía entre nitratos, sulfatos, metales, compuestos orgánicos, entre otros. Debido a su tamaño, estas partículas pueden estar suspendidas en el aire por amplios periodos de tiempo dispersándose a una cierta distancia de su origen; los patrones de viento y las condiciones atmosféricas pueden hacer que la concentración de partículas en el aire ambiental varíe día a día e incluso hora con hora (Stadlober, Hörmann, y Pfeiler, 2008). Algunas de estas partículas son tan diminutas que fácilmente son inhaladas logrando entrar al torrente sanguíneo, imitando el comportamiento del oxígeno, pero a diferencia de este generando daños al organismo (Concepción Jiménez, 2015; Stadlober et al., 2008).

En la Tabla 4.1 se muestra una de las clasificaciones del material particulado, dicha clasificación se basa en el tamaño aerodinámico de dichas partículas dividiéndolas en tres grupos, el nombre de estos grupos se forman con el acrónimo “PM”, seguido por el número que indica el tamaño máximo de las partículas en dicha clasificación (Concepción Jiménez, 2015). Mientras en la Tabla 2.2 se muestra otra forma de clasificación de forma más genérica (Davidson et al., 2005). Estas clasificaciones son utilizadas con propósitos de medición de contaminación ambiental y por los efectos de salud provocados (Davidson et al., 2005; Concepción Jiménez, 2015; Stadlober et al., 2008).

Tabla 2.1: Clasificación por tamaño del material particulado (Adaptada de Concepción 2015).

Clasificación	Tamaño
PM10	$\leq 10\mu m$
PM2.5	$\leq 2,5\mu m$
PM0.1	$\leq 0,1\mu m$

⁴Por sus siglas en inglés *Particulate Matter*

Tabla 2.2: Clasificación general del material particulado (Adaptada de Davidson 2005).

Designación	Tamaño
Material particulado ultrafino	PM < 0,1
Material particulado fino	PM 0.1–2.5
Material particulado grueso	PM 2.5–10

En la Figura 2.1 se observa mejor el tamaño de las partículas, siendo comparadas con el grosor de un cabello humano ⁵ (Concepción Jiménez, 2015). En la imagen falta las partículas ultrafinas, esto se debe al tamaño tan reducido que estas tienen.

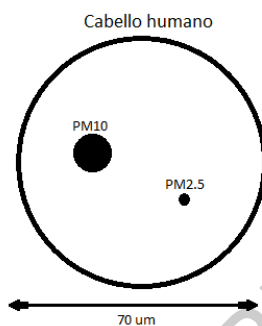


Figura 2.1: Comparación entre las partículas PM10, PM2.5 y un cabello humano (Adaptada de Concepción 2015).

2.1.1. Partículas PM10

Las partículas PM10 forman del caso de estudio del presente trabajo, las cuales están conformadas principalmente por el polen de diversas flores, rocas trituradas, tierra, contaminantes generados por el uso de combustibles fósiles, entre otros (Stadlober et al., 2008). En la Figura 2.2 se muestra algunos ejemplos de dichas partículas.

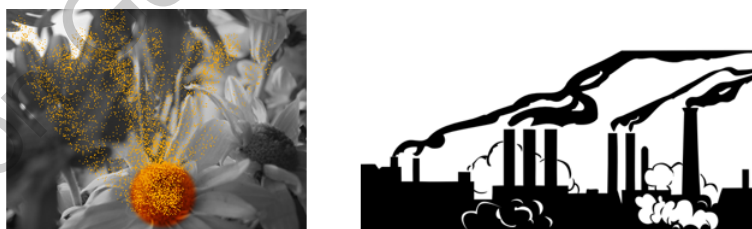


Figura 2.2: Ejemplo del origen de las PM (natural y artificial) (Adaptada de Stadlober 2008).

2.1.2. Excedencias

Una excedencia se denomina cuando los valores de una variable sobrepasa el valor máximo permitido. En este caso la variable se refiere a la concentración media diaria de partículas PM10 y

⁵El cabello humano tiene un grosor promedio de 70 μm

el límite permitido lo define la norma mexicana NOM-025-SSA1-2014, la cual establece una media diaria de $75 \mu\text{g}/\text{m}^3$ y una media anual de $40 \mu\text{g}/\text{m}^3$. Como la norma establece un valor medio, es necesario tomar los registros ambientales cada cierto tiempo e ir calculando la media en base a los registros. Este no es un tema de interés local, a nivel mundial la Organización Mundial de la Salud (OMS) igualmente indica los valores máximos que considera seguro para no afectar la salud pública. En la tabla 2.3 se observa la diferencia entre la norma mexicana y lo dicho por la OMS.

Tabla 2.3: Tabla de comparación entre la norma mexicana y lo establecido por la OMS (Adaptada del Informe de calidad de aire 2016 CDMX).

Limite	NOM		OMS	
	Media diaria (24h)	Media anual	Media diaria (24h)	Media anual
	$75\mu\text{g}/\text{m}^3$	$40\mu\text{g}/\text{m}^3$	$50\mu\text{g}/\text{m}^3$	$20\mu\text{g}/\text{m}^3$

2.1.3. Daños a la salud

La exposición prolongada o de forma repetitiva al material particulado puede provocar efectos nocivos en el sistema respiratorio y cardiovascular (Brunekreef y Holgate, 2002; Terashima, Wiggs, English, Hogg, y van EEDEN, 1997). Debido a que la mayoría del material particulado se precipita en la tierra, provocando una capa de polvo en la superficie, este puede afectar seriamente a la salud tanto de organismos terrestres como a organismos acuáticos (Davidson et al., 2005).

Las personas con alguna enfermedad previa, niños, ancianos y familias de bajos ingresos son más vulnerables a padecer por los efectos negativos de la sobre exposición (Terashima et al., 1997). Con la clasificación del material particulado por su tamaño se estudio el daño provocado a la salud de forma independiente. Aquellas partículas mayores a $10 \mu\text{m}$ son retenidas por la mucosa nasal mientras que las PM10 por su tamaño pueden penetrar hasta los bronquios mientras que las partículas pueden incluso llegar hasta los alveolos (la parte más profunda del aparato respiratorio) (Davidson et al., 2005; Terashima et al., 1997). En la Figura 2.3 se observa los diferentes lugares donde las partículas generan su principal daño en el sistema respiratorio.

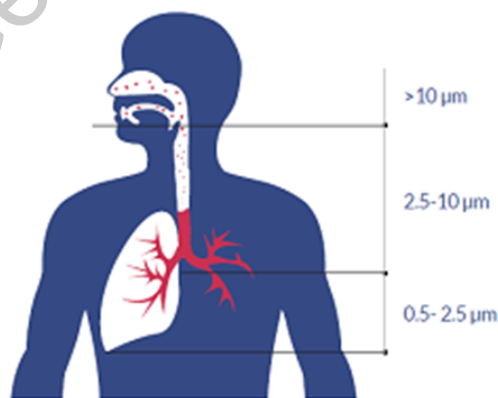


Figura 2.3: Lugar de daño provocado por las diferentes partículas ambientales (Tomada de Terashima 1997).

2.2. Inteligencia artificial

Los principales campos de aplicación de la inteligencia artificial según el reporte TechRadar de Forest (2018) (Gualtieri, 2017) son los siguientes:

- **1. Generación de lenguaje natural.** En esta aplicación las computadoras convierten datos en texto, permitiendo comunicar ideas. Actualmente, se utilizan en el servicio al cliente para generar informes de mercado por empresas como Narrative Science, Yseop, Lucidworks, Digital Reasoning y SAS.
- **2. Reconocimiento de voz.** El principal ejemplo de esto es “Siri” el sistema que puede entender lo que se le dice. No es el único sistema de este tipo, existiendo diversas compañías que ofrecen este servicio como lo son NICE, Nuance Communications, OpenText y Verint Systems.
- **3. Biométricas.** Con esta tecnología se puede identificar, medir y analizar el comportamiento humano y los aspectos físicos de la estructura y forma del cuerpo, permitiendo interacciones naturales entre los seres humanos y la máquinas.
- **4. Automatización de procesos robóticos.** Desde la implementación de robots en la industria el uso de la inteligencia artificial ha disminuido los riegos y las actividades repetitivas de los operadores humanos. Advanced Systems Concepts, Automation Anywhere, Blue Prism son algunos ejemplos de empresas de automatización de procesos.
- **5. Procesamiento de lenguaje natural (NLP⁶) y análisis de texto.** El análisis de texto y el NLP se utiliza en sistemas de seguridad y detección de fraudes. De igual manera se utiliza para extraer datos no estructurados de forma automática.
- **6. Defensa cibernética.** La defensa cibernética es un mecanismo de defensa de redes informáticas que se centra en prevenir, detectar y proporcionar respuestas oportunas ante ataques o amenazas hacia la infraestructura e información. Las redes neuronales recurrentes son usadas ampliamente mediante un entrenamiento supervisado, las cuales han revelado actividades sospechosas en los usuarios y detectan hasta el 85 % de todos los ciberataques.

2.2.1. Aprendizaje de maquinas

El aprendizaje de maquinas ⁷ o aprendizaje automático es el subcampo de las ciencias computacionales y una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras “aprendan” (Dietterich, 2002). En este caso el concepto de aprender se refiere al echo de que una computadora genere cierta experiencia en la realización de una tarea específica mediante una supervización humana (Michie, Spiegelhalter, Taylor, et al., 1994).

A grandes rasgos el uso básico del aprendizaje de maquinas es el uso de algoritmos para el manejo de datos, aprender de ellos y luego ser capaces de hacer una predicción o sugerencia de algo (Dietterich, 2002). Desde el inicio de la inteligencia artificial, los algoritmos han evolucionado con el objetivo de analizar y obtener mejores resultados: árboles de decisiones, redes Bayesianas, entre otros (Michie et al., 1994; Dietterich, 2002; LeCun, Bengio, y Hinton, 2015). En la Figura 2.4 se

⁶Por sus siglas del inglés Natural Language Processing

⁷Del inglés *Machine Learning*

muestra un breve desarrollo de la inteligencia artificial en los últimos años, denotando el origen del aprendizaje de maquinas y el actual aprendizaje profundo.

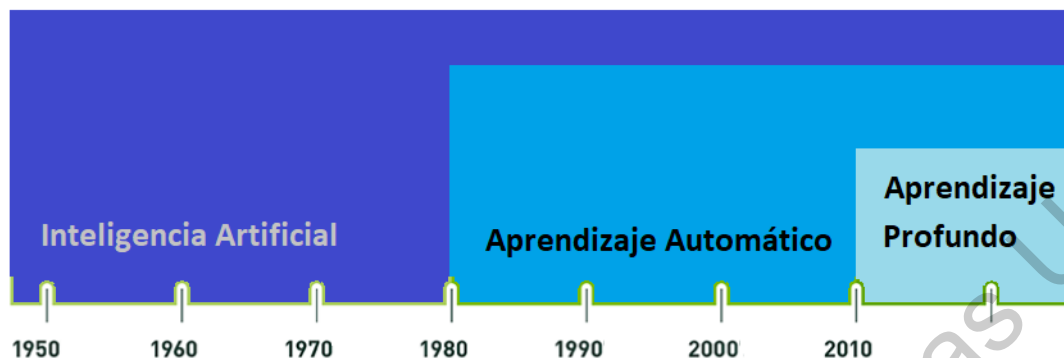


Figura 2.4: Breve línea del tiempo (Adaptada de LeCun 2015).

2.2.2. Aprendizaje profundo

El aprendizaje profundo ⁸ es una subsección del aprendizaje automático, siendo relativamente nuevo. Permite que modelos computacionales compuestos de múltiples capas de procesamiento aprendan representaciones de datos con múltiples niveles de abstracción (LeCun et al., 2015). Mejorando exponencialmente el estado del arte en reconocimiento de voz, detección de objetos, modelación de patrones no lineales, entre otros (Schmidhuber, 2015; LeCun et al., 2015).

El aprendizaje profundo recibe su nombre ya que utiliza redes profundas. El número de capas dentro de una red neuronal define la profundidad de la red. Dichas capas tienen diferentes funciones y escalas que cambian la entrada a medida que pasa a través de ellas, eventualmente la red produce una predicción (Schmidhuber, 2015).

La principal diferencia entre el aprendizaje profundo y el aprendizaje automático se basa en que el aprendizaje automático requiere cierta supervisión externa al algoritmo implementado, mientras el aprendizaje profundo tiene etapas de retropropagación con el cual puede aprender por si mismo resolver el algoritmo implementado, de la misma manera la cantidad de datos utilizados en el aprendizaje profundo es mayor a los usados en el aprendizaje automático. Por ende se a usado principalmente en el área de “grandes datos” ⁹; grandes empresas como el caso de Google usa el aprendizaje profundo para el reconocimiento de voz y de imágenes (Ahmad, Farman, y Jan, 2019).

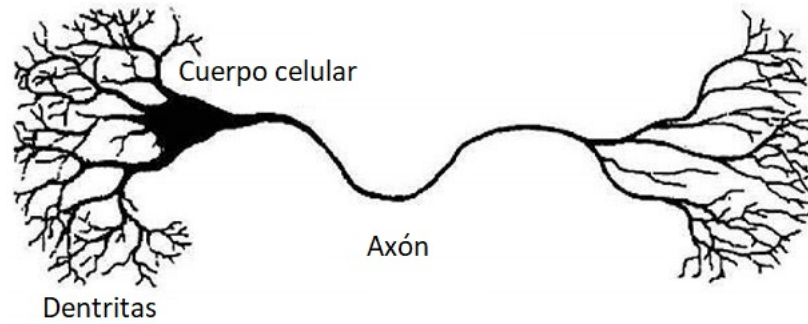
2.3. Redes neuronales artificiales

Las redes neuronales artificiales son un modelo computacional inspirado en el cerebro humano, más específicamente en una red de neuronas (Hopfield, 1988). En la Figura 2.5 se puede observar la comparación entre las partes de una neurona natural y una representación esquemática de un modelo de una neurona artificial. De esta manera se observa que ambas neuronas comparten ciertas características, tales como las dendritas, el cuerpo celular y el axón; los cuales en la neurona artificial

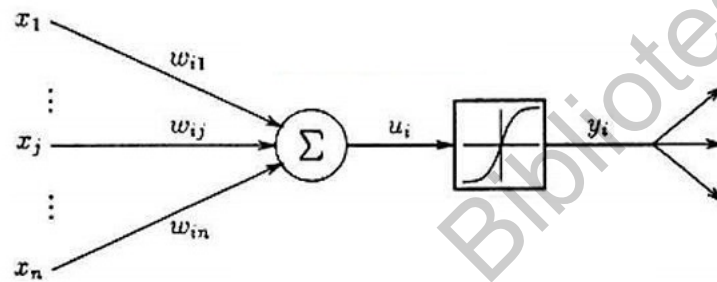
⁸Del inglés *Deep Learning*

⁹En inglés Big Data

se representan por las entradas x_n y los pesos w_{in} , la sección de sumatoria Σ y la función de activación respectivamente.



(a) Neurona natural



(b) Neurona artificial

Figura 2.5: Comparación entre una neurona artificial y una natural (Tomada de Hopfield 1988).

En esta área, Warren McCulloch y Walter Pitts (1943) fueron pioneros en el desarrollo de un modelo informático de una neurona artificial, elemento básico de una red neuronal, llamado umbral lógico (Hassoun et al., 1995). En 1958, Frank Rosenblatt definió el perceptrón (un tipo de neurona), como un algoritmo de reconocimiento de patrones basado en una red de aprendizaje de dos capas (Rojas, 2013). Estas neuronas al ser conectadas con otras neuronas se denominan red neuronal. Término que solo fue reportado hasta 1965 por Ivakhnenko y Lapa (Schmidhuber, 2015) La información de entrada atraviesa la red, sometándose a diversas operaciones, produciendo unos valores de salida (Abraham, 2005).

Cada neurona está conectada con otras a través de unos enlaces. En los enlaces el valor de salida de la neurona anterior es multiplicado por un valor de peso. Estos pesos en dichos enlaces pueden incrementar o disminuir el estado de activación de las neuronas adyacentes (Abraham, 2005; Hassoun et al., 1995). Dichas redes se conforman por una capa de entrada, n capas ocultas y una capa de salida. En la Figura 2.6 se muestra un ejemplo de una red neuronal con tres capas, con la finalidad de visualizar la capa de entrada, una capa oculta y una capa de salida.

Un avance clave fue el algoritmo de propagación hacia atrás ¹⁰, el cual resuelve eficazmente el problema de la programación de XOR ¹¹, este proceso utiliza la diferencia entre el resultado producido y el resultado deseado para cambiar los pesos de las conexiones entre las neuronas

¹⁰Del inglés *backpropagation*

¹¹Compuerta lógica usada para evaluar el correcto funcionamiento de una red neuronal

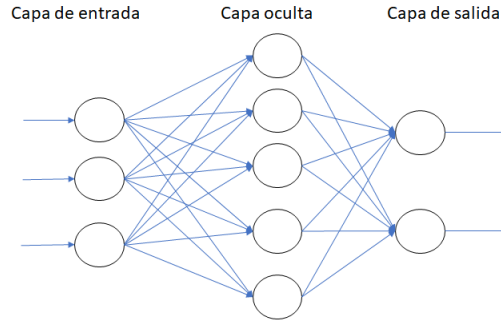


Figura 2.6: Red neuronal básica (Adaptada de Hassoun 1995).

artificiales (Rojas, 2013). La compuerta XOR es una compuerta lógica que requiere dos entradas distintas para que su salida sea verdadera. En la Tabla 2.4 se muestra su comportamiento, teniendo como entrada las variables “X” y “Y”.

Tabla 2.4: Tabla de operación de la compuerta lógica XOR (Tomada de Hopfield 1988).

X	Y	Salida
0	0	0
0	1	1
1	0	1
1	1	0

Actualmente existen diversas estructuras o topologías de redes neuronales, en la rama inteligencia artificial se les considera resultado de la interpretación del funcionamiento del cerebro humano, pero no está claro que topología refleja realmente el funcionamiento cerebral (Abraham, 2005; Hopfield, 1988).

Algoritmo backpropagation

En español propagación hacia atrás de errores o retropropagación, es un método de cálculo del gradiente utilizando en el entrenamiento de redes neuronales artificiales. De forma simple, al alimentar a una red con un patrón establecido, este se propagará desde la capa de entrada a la capa de salida. En la capa de salida se compara la señal obtenida con la salida deseada y se calcula una señal de error al cual retrocede hasta la capa de entrada. Las capas intermedias solo reciben una fracción de la señal del error, basándose en la contribución relativa que haya aportado cada neurona a la salida original (Abraham, 2005; Rojas, 2013). La eficacia de este algoritmo se basa en que cada vez que se realice la retropropagación el error irá disminuyendo.

2.3.1. Redes neuronales recurrentes

Las redes neuronales recurrentes (RNN ¹²) son una herramienta del aprendizaje de máquinas, las cuales calculan recursivamente nuevos estados aplicando funciones de transferencia a estados anteriores y a sus entradas. Las funciones de transferencia generalmente están compuestas por una

¹²Del inglés Recurrent Neural Network

transformación afín seguida de una función no lineal, determinada por la naturaleza del problema en cuestión (Bianchi, Maiorino, Kampffmeyer, Rizzi, y Jenssen, 2017).

En 2007, Maass demostró que las RNNs poseen la llamada propiedad de aproximación universal, la cual establece la capacidad de aproximación a sistemas dinámicos no lineales arbitrarios con cierta precisión arbitraria, mediante la realización de mapeos complejos desde secuencias de entrada hasta secuencias de salida (Maass, Joshi, y Sontag, 2007).

En la Tabla 2.5 se muestra algunas características propias de una RNN comparada con una red neuronal simple (Bianchi et al., 2017). La RNN funciona como una red con múltiples copias de si misma, cada una con un mensaje a su sucesor, ampliamente utilizadas en listas o en datos temporales. En la Figura 2.7 se observa claramente una red recurrente desplegada de forma simple.

Tabla 2.5: Comparación entre una red neuronal simple y una red neuronal recurrente (Adaptada de Bianchi 2017).

Red Neuronal Simple	Red Neuronal Recurrente
Los vectores de entrada producen vectores de salida	Tratan datos secuenciales de forma eficiente
No mezclan información entre ejecuciones	Utilizan las salidas anteriores como entrada
Tratan una secuencia de datos de una sola vez	Tratan secuencias de datos, elemento a elemento

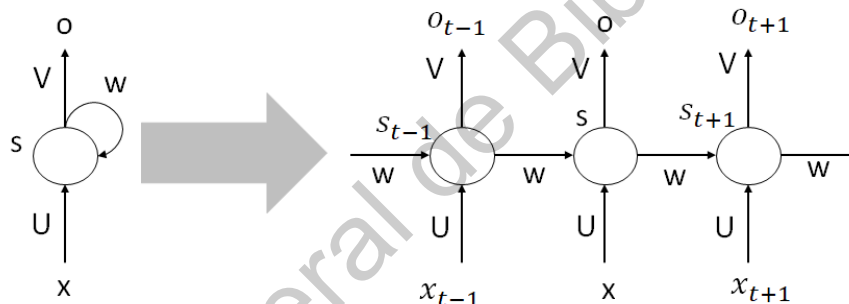


Figura 2.7: Red Neuronal Recurrente (Tomada de Bianchi 2017).

Las RNN no cuentan con una estructura de capas definida, lo cual permite conexiones arbitrarias entre neuronas, permitiendo crear cierta temporalidad, generando que la red tenga memoria. Existen diversos tipos de redes recurrentes dependiendo del número de capas y la forma de realizar la retropropagación (Guyon, 1991).

2.3.2. Red neuronal recurrente LSTM

La red neuronal recurrente LSTM¹³ o red neuronal recurrente de largo y corto plazo de memoria, propuesta por Hochreiter y Schmidhunder en 1997 (Lipton, Berkowitz, y Elkan, 2015), forman parte del aprendizaje profundo. Esta red recurrente es la utilizada principalmente en el presente trabajo.

A diferencia de una red neuronal recurrente simple que forma una memoria a largo plazo en forma de pesos entre las neuronas, los cuales se van modificando durante el entrenamiento de la red y una memoria a corto plazo definida en las funciones de activación entre la comunicación de los nodos de las neuronas. El modelo LSTM introduce un bloque de memoria interno, compuesto por

¹³Por sus siglas en inglés Long Short-Term Memory

bloques simples conectados de forma específica (ver Figura 2.8), cada uno de los cuales se describen a continuación (Hua et al., 2019):

- **Nodo de entrada:** representado por “a1” es un nodo cuya activación suele ser una sigmoide, que pondera los valores de entrada.
- **Compuerta de entrada:** “a2” es una característica propia de las LSTM. Toma los puntos actuales junto con los datos de un paso anterior. Gestionando un indicador lógico si es cero corta el flujo de datos entre nodos y si es uno los datos pasan a través de él. Actuando como un bloque de memoria.
- **Estado interno:** es la parte más importante, referenciada con una “S”, este nodo tiene una conexión recurrente, esta conexión suele denominarse carrusel de errores constantes, evitando que el error aumente.
- **Compuerta de olvido:** “a4” introducido por Ger en el 2000, proporciona un método por el cual la red puede ajustarse al contenido del estado interno. Usado especialmente para series de tiempo.
- **Compuerta de salida:** h_t es el valor producido por el valor del estado interno multiplicado por el valor de la compuerta de salida “a3”, para asegurar esta salida la función de activación del primer estado es una función tanh.

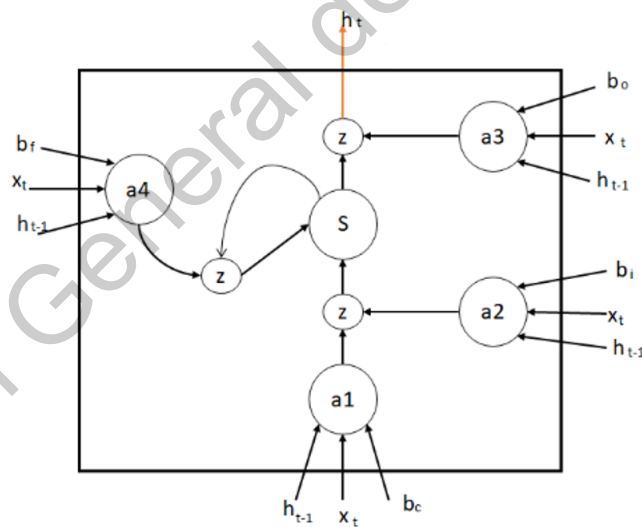


Figura 2.8: Neurona básica LSTM (Adaptada de Hua 2019).

En conclusión la neurona dispone de un optimizador de las ponderaciones basadas en el error de salida de la red resultante, que controla cada compuerta. Las ecuaciones de la 2.1 a la 2.5 describen el comportamiento de la estructura interna (Hua et al., 2019).

$$i_t = \sigma(W_{xi} * x_t + W_{hi} * h_{t-1} + W_{ci} * C_{t-1} + b_i) \quad (2.1)$$

$$f_t = \sigma(W_{xf} * x_t + W_{hf} * h_{t-1} + W_{cf} * C_{t-1} + b_f) \quad (2.2)$$

$$c_t = i_t * \tanh(W_{xc} * x_t + W_{hc} * h_{t-1} + b_c) + f_t * c_{t-1} \quad (2.3)$$

$$o_t = \sigma(W_{x0} * x_t + W_{h0} * h_{t-1} + W_{c0} * c_t + b_0) \quad (2.4)$$

$$h_t = o_t * \tanh(c_t) \quad (2.5)$$

Donde:

- x_t : es el vector de entrada de la neurona.
- f_t : es el vector de activación de la compuerta de olvido.
- i_t : es el vector de activación de la compuerta de entrada/actualización.
- o_t es el vector de estado oculto, mejor conocido como vector de salida de la unidad LSTM.
- W : son los pesos o pesos de matrices implementados.

2.3.3. Funciones de activación

Tanto en las redes neuronales biológicas como en las redes neuronales artificiales, una neurona no solo transmite la entrada que recibe. Existe un paso intermedio, denominado función de activación (Guyon, 1991; Rojas, 2013; Abraham, 2005; Ramachandran, Zoph, y Le, 2017). Las funciones de activación en las redes neuronales artificiales son funciones no lineales. Si no existiera una función de activación, todas las capas se pueden contraer en una sola capa (Rojas, 2013). Las redes neuronales son capaces de trabajar con espacios muestrales no lineales, solo por las funciones de activación (Guyon, 1991). Existen diversas funciones de activación, en el presente trabajo se implementaron una sigmoide, una ReLU ¹⁴ y una SELU ¹⁵.

Históricamente, la función sigmoide es la función de activación mas antigua y popular. Se define con la ecuación 2.6.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.6)$$

Donde: e denota la constante exponencial ¹⁶. Una neurona que use la sigmoide como función de activación se le llama *neurona sigmoide* (Guyon, 1991; Abraham, 2005). En la Figura 2.9 se observa de forma gráfica dicha función.

$\sigma(z)$ actúa como una función compactadora, comprimiendo la salida a un rango de 0 a 1. En el centro, donde $z = 0$, $\sigma(0) = 1/(1 + e^0) = 0.5$. Para valores negativos grandes de z , el termino e^{-z} en el denominador crece exponencialmente, y $\sigma(z)$ se aproxima a 0. Al contrario, valores positivos grandes de z reducen e^{-z} hacia 0, y $\sigma(z)$ se aproxima a 1 (Abraham, 2005).

¹⁴Por sus siglas en inglés Rectified Linear Unit

¹⁵Por sus siglas en inglés Scaled Exponential Linear Unit

¹⁶Aproximadamente 2.71828

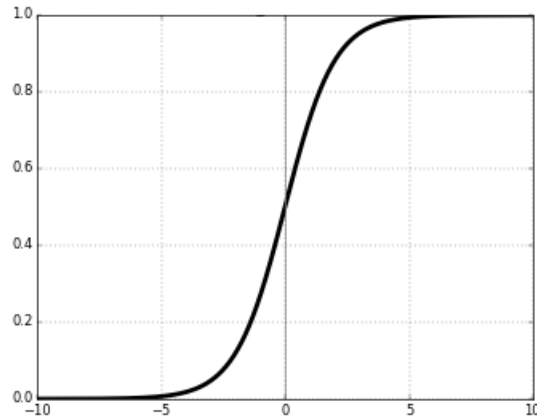


Figura 2.9: Función de activación sigmoide (Tomada de Abraham 2005).

Posteriormente se propuso la función ReLU (en español Unidad Lineal Rectificada), la cual reduce el problema del descenso del gradiente. El cual es un problema provocado por lo difícil de entrenar una red neuronal de múltiples capas utilizando la función sigmoide. La función ReLU se define como la ecuación 2.7. En la Figura 2.10 se muestra su representación gráfica.

$$R(z) = \max(0, z) \quad (2.7)$$

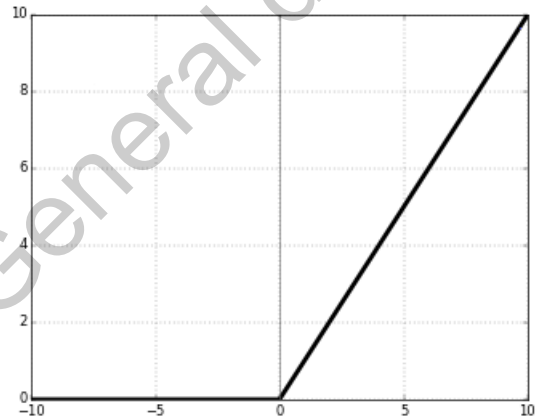


Figura 2.10: Función de activación ReLU (Tomada de Abraham 2005).

La superioridad de ReLU se basa en que su rango de capacidad de respuesta es más útil. Mientras que la capacidad de respuesta de una función sigmoide decae relativamente rápido en ambos extremos, la función ReLU permite el paso de todos los valores positivos sin cambiarlos, asignando todos los negativos a 0 (Rojas, 2013).

Recientemente Djork-Arne Clevert, Thomas Unterthiner y Sepp Hochreiter en 2015, propusieron la ELU ¹⁷, en español Unidad Lineal Exponencial. Al igual que la ReLU mejora el problema del

¹⁷Por sus siglas del inglés Exponential Linear Unit

gradiente descendente a través de la identificación de los valores positivos. En contraste con la ReLU tiene valores negativos que le permite empujar las activaciones de las unidades medias más cerca de cero, pero con una menor complejidad computacional (Ramachandran et al., 2017).

La SELU ¹⁸, en español Unidad Lineal Exponencial Escalada es una variante de la ELU. Propuesta en 2017, por G. Klambauer, T. Unterthiner, A. Mayr y S. Hochreiter. Es como una red neuronal autonormalizada con una varianza media y unitaria cercana a cero, convergiendo hacia la media y varianza unitaria cuando se propaga a través de múltiples capas durante el entrenamiento, ver Figura 2.11. Lo que la hace adecuada para el aprendizaje profundo (Ramachandran et al., 2017). La función se describe a continuación:

$$f(x) = \lambda \begin{cases} x, & x > 0 \\ \alpha e^x - \alpha, & x \leq 0 \end{cases} \quad (2.8)$$

Donde, λ es el factor de escala y $\alpha \approx 1,6733$

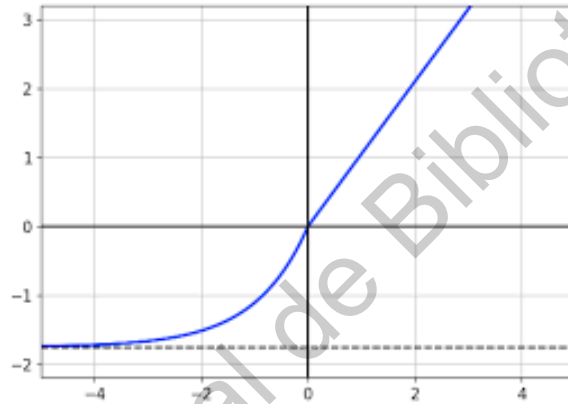


Figura 2.11: Función de activación SELU (Tomada de Ramachandran 2017).

¹⁸Scaled Exponential Linear Unit

Metodología

En el desarrollo de este trabajo se propone el uso de dos redes neuronales recurrentes LSTM para la modelación y predicción de excedencias de partículas PM10 de la Ciudad de México. La primera red se implementara para la modelación del comportamiento de las partículas PM10, con la finalidad de no afectar la predicción de excedencias cuando exista la falta de datos reales, provocados por cualquier eventualidad, de este modo la segunda red mantendrá un registro de datos completos para la predicción. Al dividir las etapas de modelación y predicción se busca que la red sea robusta permitiendo mantener una repetitividad en sus resultados.

La metodología implementada se dividió en dos etapas, la primera etapa describe el proceso de la implementación de la red LSTM propuesta, ver Figura 3.1. En la cual se puede distinguir cinco secciones importantes, la adquisición y preprocesamiento de los datos, la primera red o red de modelación, la caracterización de excedencias, la segunda red o red de predicción y la evaluación continua.

Mientras la segunda etapa describe el proceso de validación de los resultados obtenidos con la propuesta de la red LSTM (ver Figura 3.2), utilizando métodos de imputación como lo es el algoritmo MICE y SVM para proponer datos faltantes; y en el caso de la predicción se comparo con una red MLP.

Los métodos de imputación estadísticamente hablando, se utilizan para completar datos faltantes de registros de forma que estos datos propuestos mantengan similitud del comportamiento con los datos reales. En otras palabras los datos propuestos mantienen la varianza de los datos reales permitiendo una continuidad homogénea (Zhang, 2016).

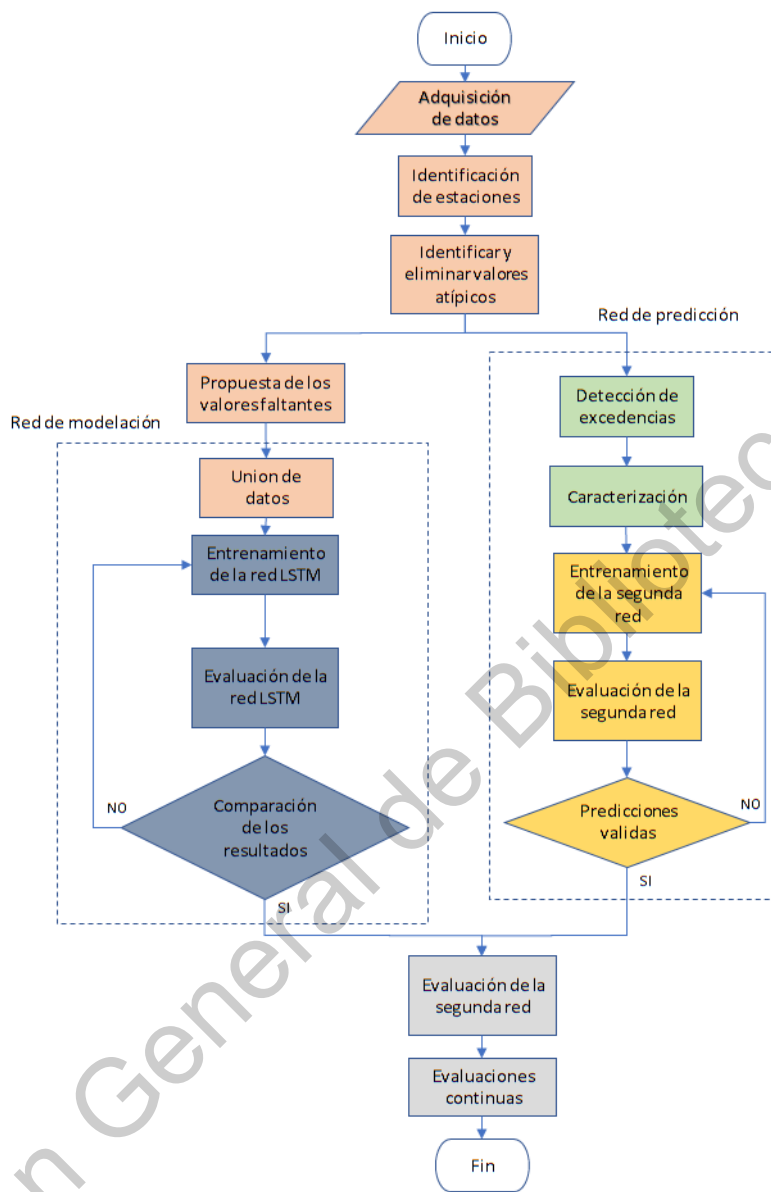


Figura 3.1: Metodología implementada.

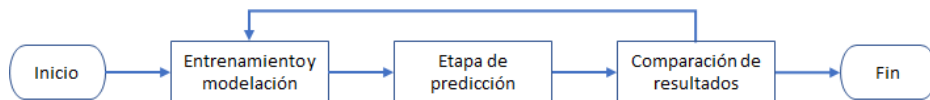


Figura 3.2: Metodología de comparación de resultados.

3.1. Etapa uno de la implementación de la red LSTM

3.1.1. Adquisición y preprocesamiento de los datos.

Los datos fueron obtenidos de la base de datos RAMA (Red Automática de Monitoreo Atmosférico) de la Ciudad de México, la cual cuenta con registros de diversos contaminantes aéreos como lo son el dióxido de nitrógeno, óxido de nitrógeno, monóxido de nitrógeno, dióxido de azufre, monóxido de carbono, partículas PM10, partículas PM2.5, entre otras (CDMX, 2019). A su vez cuenta con diversas estaciones de monitoreo distribuidas en distintos sectores, como se muestra en la Figura 3.3. Los nombres de dichas estaciones son siglas del lugar donde se encuentran localizadas, como ejemplo la estación con la clave MER, llamada Merced en la delegación de Venustiano Carranza.

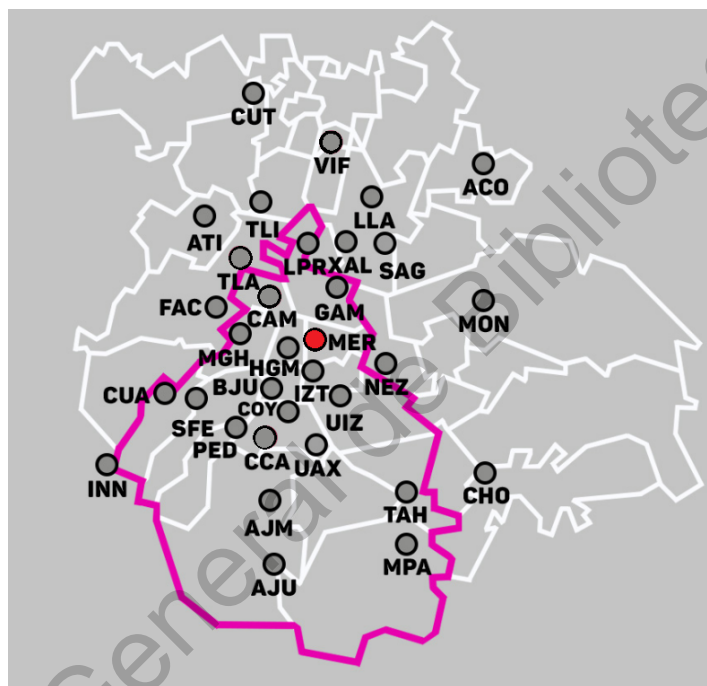


Figura 3.3: Localización de las estaciones de la Red Automática de Monitoreo Atmosférico (RAMA) (Tomada de la base de datos RAMA).

De forma individual cada estación registra las concentraciones de contaminantes aéreos con una frecuencia de muestreo de una hora. Los registros utilizados en el presente trabajo son del año 2002 al 2018. En la Tabla 3.1 se muestra una sección de los registros del año 2018, a modo de ejemplo. Debido al crecimiento de la ciudad no siempre sean contados con las 24 estaciones actuales, en el 2002 solo estaban en operación 15 estaciones.

En la tabla se observan dos tipos de registros, los números positivos representan la concentración de partículas PM10 en $\mu g/m^3$ en ese momento determinado, mientras que los valores negativos representan valores atípicos, los cuales indican errores de medición producidos por un mal funcionamiento en el sensor, desconexión por mantenimiento del mismo, entre otros factores (CDMX, 2019).

Tabla 3.1: Ejemplo de una tabla de registros de partículas PM10 del año 2018.

FECHA	HORA	ACO	AJM	ATI	BJU	CAM	CHO	CUA	CUT	FAC	HGM	INN	IZT	MER	MGH	MPA
01/01/2018	1	125	92	-99	78	-99	-99	137	100	52	96	48	-99	114	91	-99
01/01/2018	2	98	95	-99	73	-99	-99	111	187	37	85	47	-99	108	67	-99
01/01/2018	3	88	81	-99	82	-99	-99	85	100	64	92	36	-99	110	74	-99
01/01/2018	4	82	59	55	87	-99	-99	58	75	48	102	31	-99	113	80	-99
01/01/2018	5	96	48	57	69	-99	-99	52	134	42	95	23	-99	107	64	-99
01/01/2018	6	88	46	51	62	-99	-99	84	144	39	55	25	-99	78	62	-99
01/01/2018	7	86	71	26	55	-99	-99	62	132	31	71	21	-99	77	66	-99
01/01/2018	8	71	100	32	52	-99	-99	36	152	-99	65	19	-99	70	64	-99
01/01/2018	9	55	102	72	54	-99	-99	58	217	-99	62	17	-99	-99	59	-99
01/01/2018	10	47	96	76	67	-99	-99	65	32	54	80	27	-99	69	32	-99
01/01/2018	11	37	102	33	70	-99	-99	37	-99	78	75	54	-99	34	79	-99
01/01/2018	12	45	76	46	87	-99	-99	86	-99	56	60	57	-99	49	62	-99
01/01/2018	13	77	85	63	70	-99	-99	36	-99	30	78	109	-99	82	69	-99
01/01/2018	14	54	70	56	87	-99	-99	78	-99	98	73	48	-99	69	68	-99
01/01/2018	15	57	85	75	102	-99	-99	96	-99	78	-99	50	-99	102	89	-99
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
31/12/2018	14	72	50	40	40	84	64	74	60	88	45	-99	86	-99	35	-99
31/12/2018	15	70	55	55	75	102	62	45	85	53	76	-99	21	-99	30	-99
31/12/2018	16	64	80	84	91	45	97	90	96	75	35	-99	42	-99	68	-99
31/12/2018	17	54	42	63	36	81	82	96	55	53	68	-99	72	-99	98	-99
31/12/2018	18	61	49	68	48	43	20	60	45	80	45	28	56	-99	81	-99
31/12/2018	19	77	51	62	46	48	52	36	83	102	58	20	76	54	39	-99
31/12/2018	20	81	52	88	35	52	42	40	113	70	42	18	64	61	41	-99
31/12/2018	21	51	58	86	39	63	-99	32	93	58	48	22	70	86	67	-99
31/12/2018	22	45	49	76	56	74	81	79	74	73	46	32	66	69	64	-99
31/12/2018	23	-99	46	85	67	91	76	140	52	57	51	31	74	63	78	-99
31/12/2018	24	-99	42	92	64	122	81	-99	103	64	60	34	-99	75	91	-99

Al trabajar en un intervalo de tiempo 16 años, se tuvo que identificar aquellas estaciones que se han mantenido a lo largo de los años. Posteriormente se identifico la cantidad de valores atípicos presentes en dichas estaciones con la finalidad de trabajar con la mayor cantidad de datos reales. En este punto los datos se dirigen a dos etapas diferentes, a la detección de excedencias (la cual sera explicada posteriormente) y a la sección de propuesta de valores faltantes.

La primera red LSTM ocupa una continuidad en los datos de entrada; por ende, los valores atípicos fueron reemplazados por valores medios entre los registros anteriores y posteriores a dicho evento. Este método de generación de datos faltantes resulta ser simple, de bajo coste computacional y mantiene el comportamiento estudiado de las partículas (Pratama, Permanasari, Ardiyanto, y Indrayani, 2016).

En la Figura 3.4 se puede visualizar de forma gráfica, los recuadros de color naranja selecciona los valores atípicos mientras que los recuadros azules seleccionan valores anteriores y posteriores a dichos valores.

FECHA	HORA	ACO	AJM	ATI	BJU	CAM	CHO	CUA	CUT	FAC	HGM	INN	IZT	MER
01/01/2018	1	125	92	-99	78	-99	-99	137	100	52	96	48	-99	114
01/01/2018	2	98	95	-99	73	-99	-99	111	187	37	85	47	-99	108
01/01/2018	3	88	81	-99	82	-99	-99	85	100	64	92	36	-99	110
01/01/2018	4	82	59	55	87	-99	-99	58	75	48	102	31	-99	113
01/01/2018	5	96	48	57	69	-99	-99	52	134	42	95	23	-99	107
01/01/2018	6	88	46	51	62	-99	-99	84	144	39	55	25	-99	78
01/01/2018	7	86	71	26	55	-99	-99	62	132	31	71	21	-99	77
01/01/2018	8	71	100	32	52	-99	-99	36	152	-99	65	19	-99	70
01/01/2018	9	55	102	72	54	-99	-99	58	217	-99	62	17	-99	-99
01/01/2018	10	47	96	76	67	-99	-99	65	32	54	80	27	-99	69
01/01/2018	11	37	102	33	77	-99	-99	36	-99	30	75	54	-99	102
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
31/12/2018	19	77	51	62	46	48	52	36	83	53	45	20	76	54
31/12/2018	20	81	52	88	35	52	42	40	113	70	42	18	64	61
31/12/2018	21	51	58	86	39	63	-99	32	93	58	48	22	70	86
31/12/2018	22	45	49	76	56	74	81	79	74	73	46	32	66	69
31/12/2018	23	-99	46	85	67	91	76	140	52	57	51	31	74	63
31/12/2018	24	-99	42	92	64	122	81	-99	103	64	60	34	-99	75

Figura 3.4: Propuesta de reemplazo de los valores atípicos.

El algoritmo implementado para esto se explica a continuación:

Inicio: Determinación de la longitud del vector que representa los valores atípicos, dicha longitud determinara los valores anteriores y posteriores a utilizar.

```

for n ≤ Length do
  Afterdata = Yeardata - (Yeardata - (Length - n))
  Beforedata = Yeardata - (Yeardata - (Length + n))
  Newdata[n] = (Afterdata + Beforedata) / 2
end for

```

Donde:

Length = longitud del vector de valores atípicos
 Afterdata = Selección del valor anterior
 Beforedata = Selección del valor posterior
 Yeardata = Longitud del vector de registros del año
 Newdata = Nuevos valores

El próximo paso es pasar de 16 vectores a uno solo, agrupando de forma ordenada los registros con la finalidad de obtener un vector el cual suministrara los datos a la primera red LSTM. En la Figura 3.5 se observa el proceso de concatenación de forma gráfica.

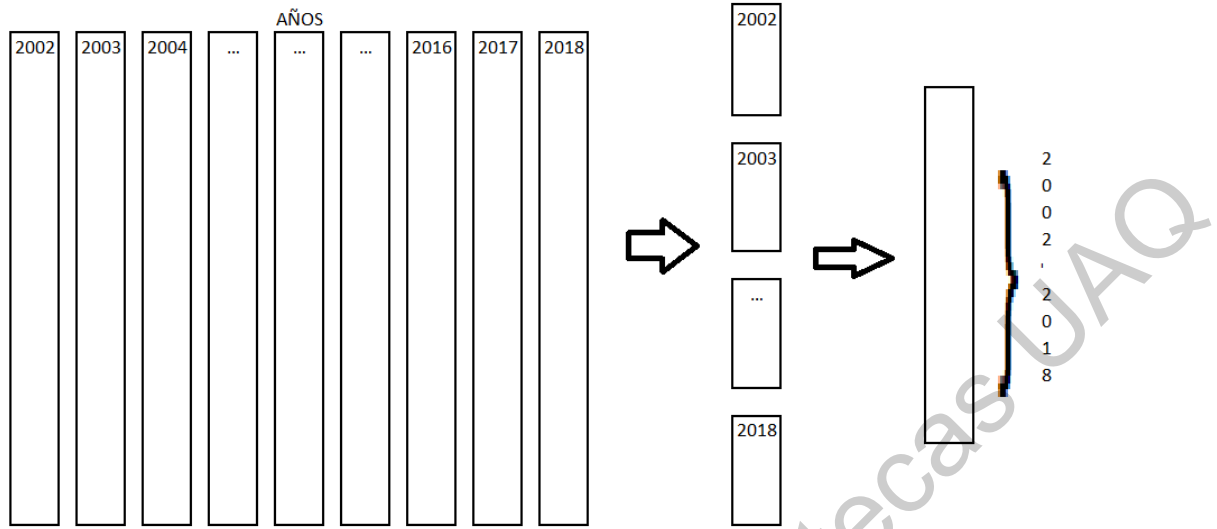


Figura 3.5: Concatenación de los datos anuales.

3.1.2. Red neuronal de modelación.

El vector de datos completos previamente formado; en la sección anterior, se escalo a rangos de -1 a 1, con la ecuación 3.1. Con la finalidad de reducir la amplitud entre los valores máximos y mínimos registrados.

$$\frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (3.1)$$

De la longitud del vector se seleccionara el 80 % para el entrenamiento y el 20 % restante sera para la validación de la red. El próximo paso es la construcción de la red neuronal; la cual contara con tres capas, la capa de entrada que estará formada por 50 neuronas LSTM, una capa oculta de 256 neuronas LSTM y la capa de salida sera una neurona dense, interconectadas todas con todas(full connection)(Hua et al., 2019).

En la Figura 3.6 se observa la estructura general de la red neuronal LSTM. Igualmente se definen las funciones de activación de las neuronas la cual será una función lineal exponencial escalada (por sus siglas en ingles SELU “Scaled Exponential Linear Unit”) y su evaluación de reducción de error sera con un error cuadrático medio. Se implemento la SELU debido a su versatilidad de manejar tanto valores positivos como negativos.

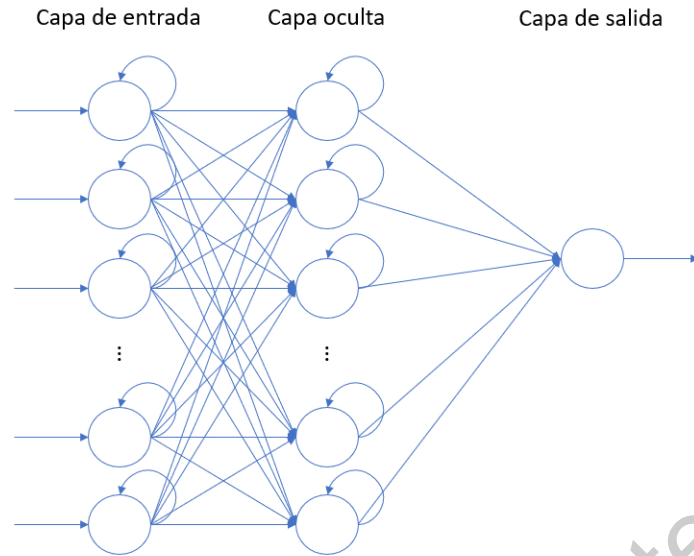


Figura 3.6: Red neuronal LSTM.

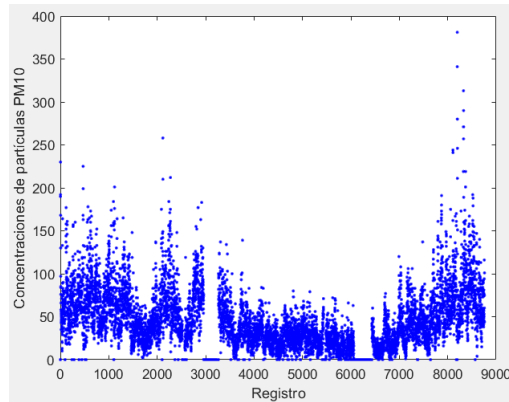
De forma general las neuronas evalúan conjuntos de 50 datos del vector inicial, un número definido de épocas, en este caso de 11.

A la red le serán suministrados conjunto de 50 datos con la finalidad de predecir el dato 51, dicho dato sera comparado con el real. El error permitirá ajustar los pesos de las conexiones de las neuronas, de forma interna las neuronas evaluarán los datos, con la finalidad de determinar aquellos que le permitan de mejor forma de predecir el siguiente dato.

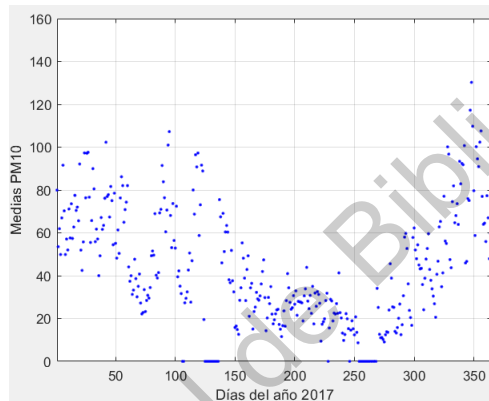
El error también es una condición de termino de entrenamiento de la red neuronal, dicho error es evaluado al finalizar una época; es decir, al finalizar de evaluar todos los conjuntos de datos comparando los resultados obtenidos con los reales.

3.1.3. Determinación de excedencias.

En esta sección se utilizan los datos originales sin la sustitución de los valores atípicos. Estos datos fueron agrupados en conjuntos de 24 datos, con la finalidad de conocer la media registrada y así disminuir la cantidad de datos de 8760 por año a 365. Dado la existencia de valores atípicos, el calculo de la media fue realizado con los datos validos de cada grupo. En la Figura 3.7 se observa la reducción de los datos.



(a) Registros.



(b) Medias.

Figura 3.7: Reducción de datos.

Utilizando de referencia la norma mexicana NOM-025-SSA1-2014, la cual determina el medio diario de una excedencia en $75 \mu\text{g}/\text{m}^3$. El registro de todos los días se compararon con el valor de la norma con la finalidad de identificar los días que presentan excedencias. El próximo paso fue guardar las condiciones registradas previamente a la excedencia; la longitud de dichos vectores sera de 24, 48 o 72 registros previos, es decir, las horas previas con las que se desea trabajar la predicción de las excedencias.

3.1.4. Red neuronal de predicción.

Esta red neuronal solo cuenta con dos capas, la capa de entrada y la capa de salida, como se muestra la Figura 3.8. La cantidad de neuronas en la capa de entrada dependerá de las longitudes de los vectores a analizar (24, 48 ó 72). Debido a que solo se busca identificar que condiciones anteceden a una excedencia se implemento una clasificación Booleana; es decir, 1 si el día presenta una excedencia y 0 si no. Las funciones de activación para esta red fueron: para la capa de entrada se implemento una activación ReLU (por la linealidad de esta permitiendo la distinción de excedencias más fácilmente), mientras que para la capa de salida se implemento una función sigmoide (Dado que es una función simétrica).

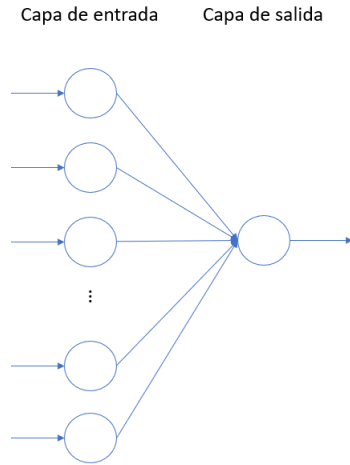


Figura 3.8: Red de predicción.

Antes de la evaluación continua ambas redes se implementaron juntas (ver Figura 3.9), debido a que la primera red tiene el objetivo de modelar el comportamiento de las partículas PM10 sin valores atípicos y la segunda red utilizará estos nuevos datos para predecir que días se presentará una excedencia.

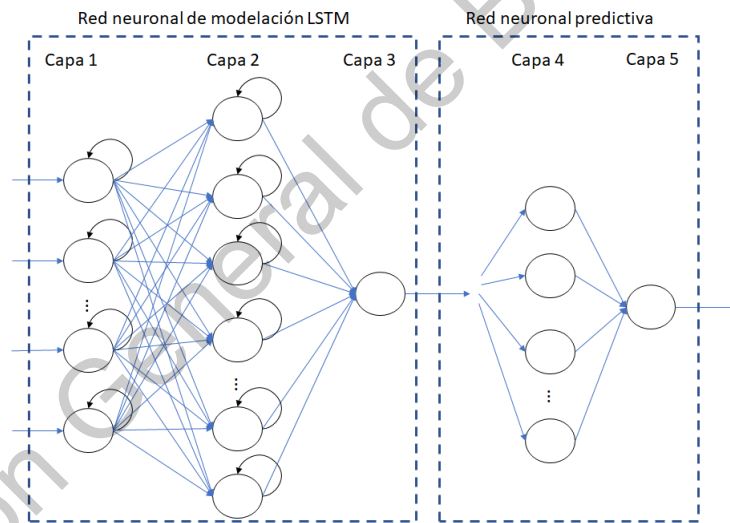


Figura 3.9: Red propuesta de forma completa.

3.1.5. Evaluación continua.

Para verificar la repetitividad del porcentaje de predicción de la red propuesta completa se evaluó de forma continua 50 veces, guardando la precisión con la cual es capaz de predecir los valores de excedencia. En otras palabras la primera red LSTM modela el comportamiento de las partículas para que los registros correspondientes a los años 2016, 2017 y 2018, se calculó la raíz cuadrada del error cuadrático medio, solo para verificar el modelado. Para evaluar la repetitividad del algoritmo, la red se evaluó en 50 ocasiones con datos tomados de forma aleatoria lo cual indicará

la robustes del algoritmo completo.

3.2. Etapa dos herramientas de comparación

Con la finalidad de evaluar la red LSTM descrita anteriormente se utilizaron tres herramientas; dos como preprocesamiento de datos y la tercera como comparación de funcionalidad. Para el preprocesamiento de los datos iniciales se implementaran un algoritmo MICE y una SVM; mientras una red MLP será usada para comparar la funcionalidad la red LSTM propuesta.

3.2.1. Imputación múltiple por ecuación encadenada

La imputación múltiple por ecuaciones encadenadas (MICE ¹⁹) también llamado imputación múltiple de regresión secuencial, surgió para tratar de resolver el problema de falta de datos en la investigación. La creación de imputaciones múltiples, a diferencia de las imputaciones individuales, explica la incertidumbre estadística de las imputaciones (Azur, Stuart, Frangakis, y Leaf, 2011). Estadísticamente hablando, una imputación es la sustitución de valores no informados en una observación por otros (Galván, 2007).

MICE crea conjuntos de datos (denominados “m”) imputando los valores faltantes; es decir, un valor que falta en el conjunto de datos originales se reemplaza por “m” valores imputados plausibles (Zhang, 2016). En otras palabras, la estadística de interés se estima a partir de cada conjunto de datos y luego se combinan en uno final. En comparación con la imputación única, este procedimiento tiene en cuenta la incertidumbre de la estimación del valor faltante. Este proceso se lleva a cabo creando primero un modelo de predicción para la variable objetivo de todas las demás variables. Por ende, se utiliza la media predictiva para las variables continuas y la regresión logística para la variable dicotómica (Azur et al., 2011; Zhang, 2016).

3.2.2. Maquinas de soporte vectorial

La maquina de soporte de vectores (SVM ²⁰) construye un conjunto de hiperplanos en un espacio de dimensionalidad muy alta que puede ser utilizado en problemas de clasificación o regresión. Cuyos datos de entrada son un vector de n-dimensionalidad (Raicharoen, Lursinsap, y Sanguanbhokai, 2003). Dado un conjunto de datos de entrenamiento se pueden etiquetar las clases y entrenar la SVM para construir un modelo que prediga la clase de la nueva muestra.

Este tipo de algoritmo busca el hiperplano que tenga la máxima distancia con los puntos que estén más cerca de él mismo. De esta forma, los puntos del vector que son etiquetados con una categoría estarán a un lado del hiperplano y los casos que se encuentren en la otra categoría estarán al otro lado. En la Figura 3.10 se observa tres distintos hiperplanos para el conjunto de datos mostrado. A los vectores formados por los puntos más cercanos al hiperplano se le llama vector de soporte (Betancourt, 2005).

¹⁹Por sus siglas del inglés Multiple Imputation by Chained Equations

²⁰Por sus siglas en inglés Support Vector Machine

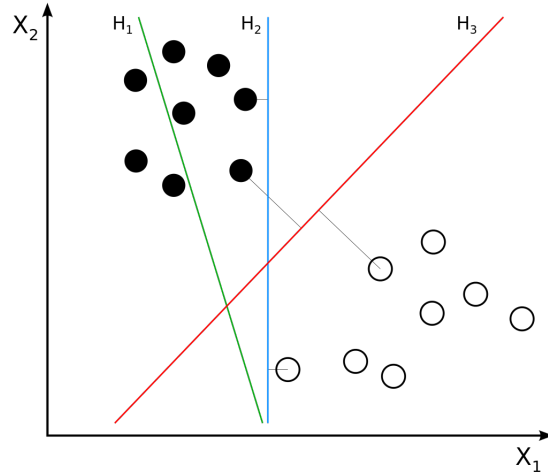


Figura 3.10: Ejemplos de hiperplanos utilizados por una SVM (Tomada de Betancourt 2005).

Soporte vectorial de regresión

Con el objetivo de realizar regresiones las SVM se modificaron poco, generando el soporte vectorial de regresión (SVR ²¹). Dado un conjunto de datos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) (X_i \in X \subseteq \mathbb{R}^n, y_i \in Y \subseteq \mathbb{R}, n$ es el número total de muestras de entrenamiento) genera de forma aleatoria e independiente a partir de una función desconocida, la SVR aproxima la función utilizando la forma:

$$f(x) = w \cdot \phi(x) + b \quad (3.2)$$

Donde $\phi(x)$ representa los espacios de características de alta dimensión que se asigna de forma no lineal iniciando desde el espacio de entrada x . El coeficiente w y b son estimados por minimización de la función de riesgo 3.3.

$$\text{minimize } \frac{1}{2} \|w\|^2 + C \frac{1}{n} \sum_{i=1}^n N_\varepsilon(y_i, f(x_i)) \quad (3.3)$$

$$L_\varepsilon(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (3.4)$$

El primer término $\|w\|^2$ es llamado término de regularización. Minimizando $\|w\|^2$ hará la función lo más plana posible. El segundo término $\frac{1}{n} \sum_{i=1}^n N_\varepsilon(y_i, f(x_i))$ es el error empírico medio por la ε -ésima función de pérdida. C se conoce como la constante de regularización.

Para obtener los estimadores w y b , 3.3 se transforma en la función objetivo primordial 3.6 al introducir la holgura positiva $\xi_i^{(*)}$ ($(*)$ denota las variables con y sin).

$$\text{minimize } \frac{1}{2} \|w\|^2 + C \frac{1}{n} \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3.5)$$

²¹Por sus siglas en inglés Support Vector Regression

sujeto por

$$\begin{aligned} y_i - w \cdot \phi(x_i) - b &\leq \varepsilon + \xi_i \\ w \cdot \phi(x_i) + b - y_i &\leq \varepsilon + \xi_i^* \quad i = 1, \dots, n \\ \xi_i^* &\geq 0 \end{aligned} \quad (3.6)$$

Finalmente son introducidos multiplicadores de Lagrange, la función de decisión 3.2 tiene la siguiente forma explícita (Cao y Tay, 2003).

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) K(x_j, x) + b \quad (3.7)$$

Donde: a_i^* son los multiplicadores de Lagrange y $K(x_i, x)$ define la función kernel. El valor del kernel es igual al producto interno entre dos vectores x_i y x_j en el espacio característico $\phi(x_i)$ y $\phi(x_j)$.

Kernels

El kernel se puede traducir como núcleo, el cual por definición la función kernel es $K : X \times X \rightarrow R$ que asigna a cada par de elementos del espacio de entrada, X , un valor real correspondiente al producto escalar de las imágenes de dichos elementos en un nuevo espacio F .

$$K(x, x') = \langle \Phi(x), \Phi(x') \rangle = (\phi_1(x)\phi_1(x') + \dots + \phi_m(x)\phi_m(x')) \quad (3.8)$$

donde $\Phi : X \rightarrow F$. Por tanto, una función kernel puede sustituir convenientemente al producto escalar:

$$D(x) = \sum_{i=1}^n \alpha_i^* y_i K(x, x_i) \quad (3.9)$$

Por tanto, dado el conjunto de funciones base, $\Phi(x) = \phi_1(x), \dots, \phi_m(x)$, el problema a resolver es encontrar los parámetros $\alpha_i^*, i = 1, \dots, n$ que optimiza el problema expresado por:

$$\begin{aligned} \max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.a.} \sum_{i=1}^n \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C, i = 1, \dots, n \end{aligned} \quad (3.10)$$

3.2.3. Red Perceptrón Multicapa

El perceptrón multicapa (MLP ²²) es capaz de actuar como un aproximador universal de funciones: al ser una red backpropagation contiene al menos una capa oculta con suficientes unidades no lineales puede aproximar cualquier tipo de función. En 1986, Rumelhart propuso un método para que una red del tipo perceptrón multicapa aprendiera la asociación que existe entre un conjunto de patrones de entrada y sus salidas correspondientes, denominado backpropagation error (propagación del error hacia atrás) (Wilamowski, 2009).

Una MLP está compuesta por una capa de entrada, una capa oculta (en la mayoría de las ocasiones) y una capa de salida (ver Figura 3.11). Las conexiones entre neuronas son siempre hacia delante; es decir las conexiones van desde las neuronas de una determinada capa hacia la siguiente,

²²Por sus siglas en inglés Multi Layer Perceptron

sin existir conexiones laterales o hacia atrás (Wilamowski, 2009; Salas, 2004). De esta forma la información fluye desde la capa de entrada a la capa de salida.

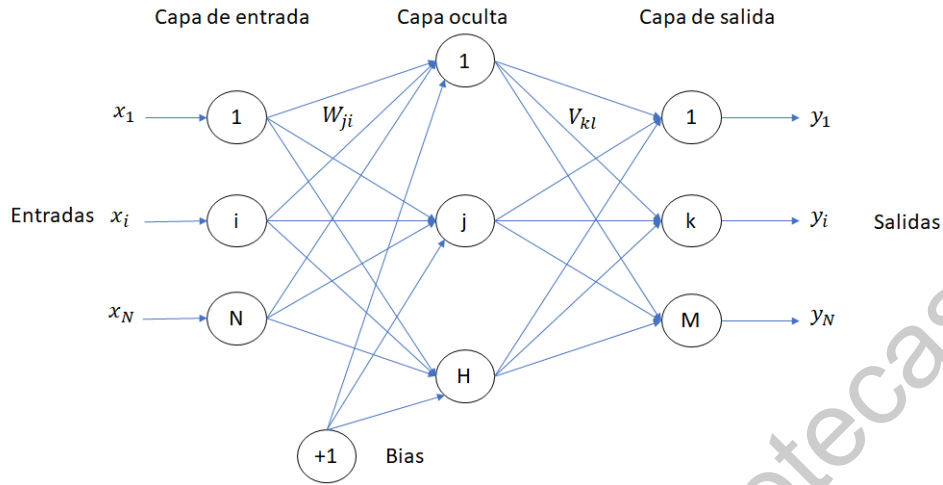


Figura 3.11: Red MLP general. Adaptada de Wilamowski 2009.

En su etapa de funcionamiento, cuando se presenta un patrón “p” de entrada $X^p : x_1^p, \dots, x_i^p, \dots, x_N^p$, éste se transmite a través de los pesos W_{ji} desde la capa de entrada hacia la capa de oculta. Las neuronas de esta capa intermedia transforman las señales recibidas mediante la aplicación de una función de activación proporcionando un valor de salida. Este se transmite a través de los pesos V_{kl} hacia la capa de salida, repitiendo la operación anterior (Wilamowski, 2009).

Finalmente la etapa de aprendizaje, el objetivo es minimizar el error entre la salida obtenida por la red y la salida deseada por el usuario, definida por un conjunto de patrones denominado grupo de entrenamiento (Wilamowski, 2009). Principalmente la MLP intenta resolver dos tipos de problemas: problemas de predicción y problemas de clasificación (Salas, 2004).

Dirección General de Bibliotecas UAQ

Resultados y Discusión

4.1. Resultados

Los resultados se dividen en tres secciones, la primera muestra los resultados obtenidos en la modelación de las partículas PM10 implementando únicamente la red neuronal LSTM propuesta, lo cual se puede ver de la Figura 4.1 a la Figura 4.3.

Las figuras muestran los datos obtenidos por la modelación de la red en color naranja y los datos reales en color azul, las tres corresponde aun tamaño de lote de entrenamiento de 96 (el cual indica que se usaron los datos de 4 días por sección).

En las gráficas se observa como los valores obtenidos por la red siguen el comportamiento de los datos originales, y en aquellos lugares donde los datos reales son inexistentes los datos de la red se acoplan sin problemas en el inicio y fin de los datos faltantes.

Las líneas azules rectas que no siguen el comportamiento de las partículas están representadas de esta forma por el programa python solo para tener cierta continuidad de los datos.

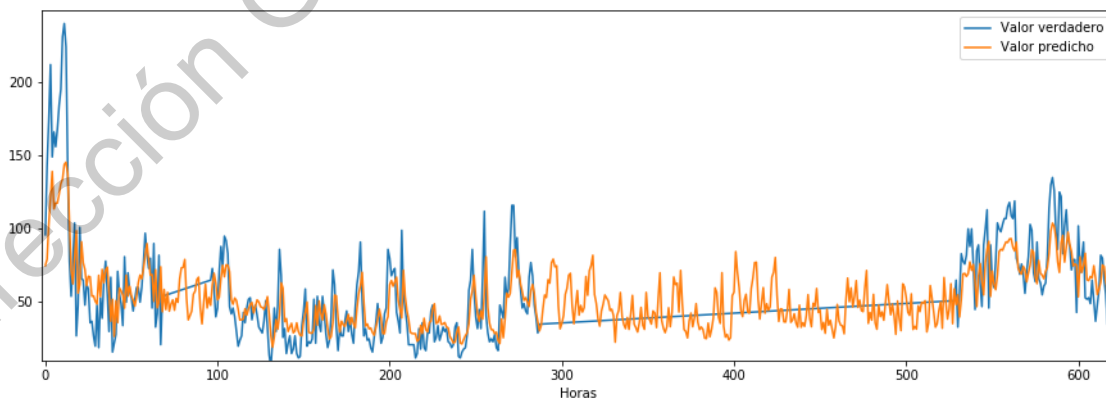


Figura 4.1: Modelación de enero del 2017, utilizando la red LSTM propuesta.

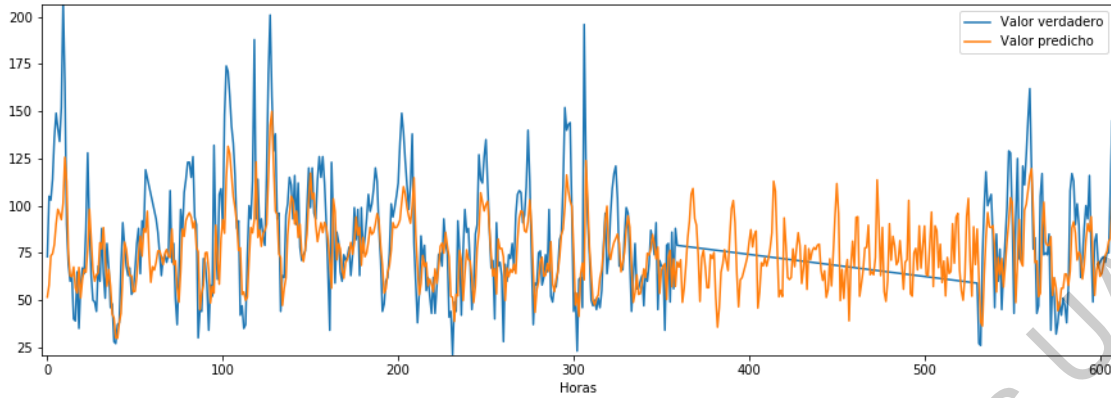


Figura 4.2: Modelación de enero del 2016, utilizando la red LSTM propuesta.

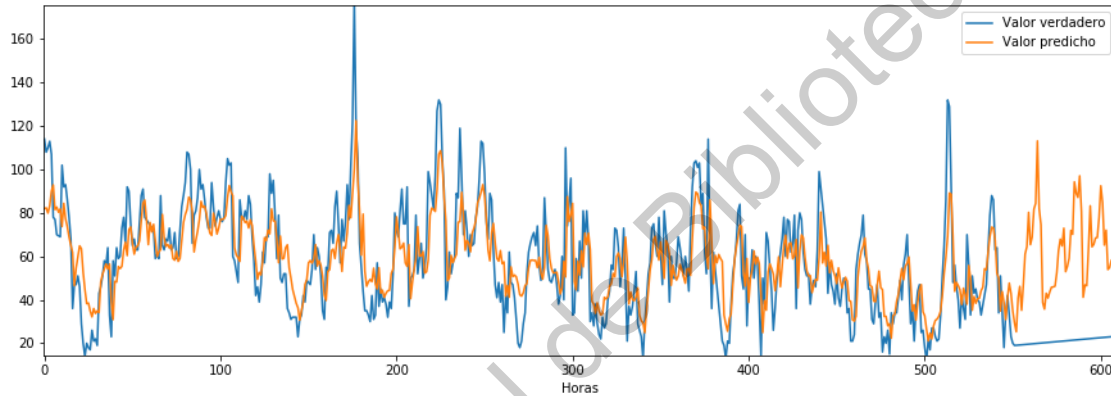


Figura 4.3: Modelación de enero del 2018, utilizando la red LSTM propuesta.

Al modificar el tamaño de lote por los habituales usados en una red LSTM los cuales son 250 y 500 se obtuvieron los siguientes resultados medidos mediante el MSE y el RMSE.

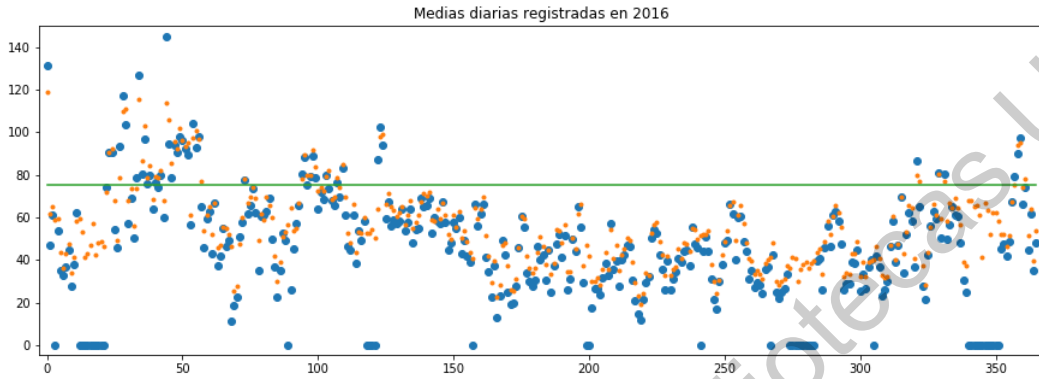
Tabla 4.1: Resultados de modificar el tamaño de muestra con el que se entrena la red LSTM.

Tamaño de lote	MSE	RMSE
96	316.2749	17.7841
250	324.6615	18.0183
500	334.1415	18.2795

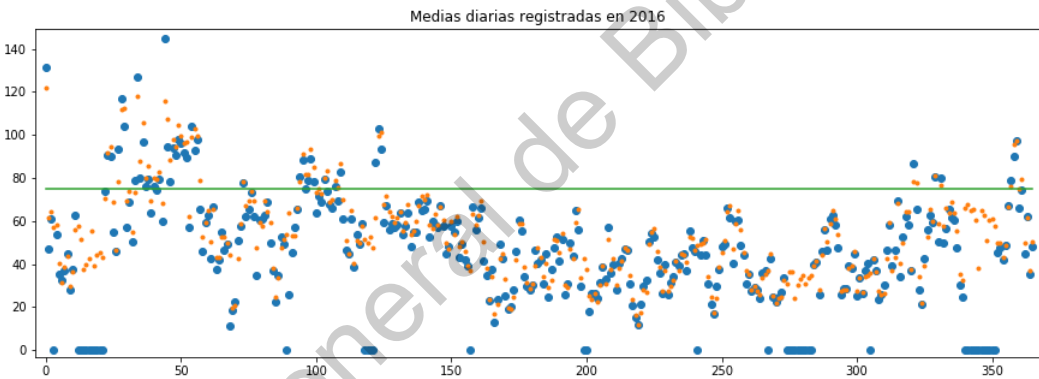
En la Figura 4.4, Figura 4.5 y en la Figura 4.6 se muestran el comportamiento de los niveles medios de las partículas PM10 del año 2016 al 2018 respectivamente; en color azul se muestran los valores reales, de color naranja los valores estimados por los diferentes algoritmos implementados y la línea verde representa el valor de excedencia establecido por la norma mexicana; es decir, marca el valor límite antes de considerarse una excedencia, si el valor no sobrepasa dicho límite no es considerado una excedencia.

Entre las gráficas que componen cada imagen se pueden percibir sutiles pero importantes diferencias, la primera gráfica (a) representa los resultados de solo usar la red LSTM, la segunda

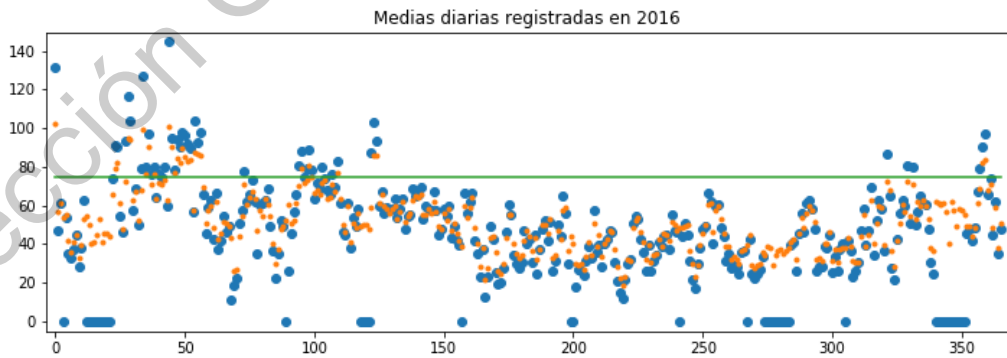
(b)) corresponde a los resultados de implementar el algoritmo MICE en el preprocesamiento de los datos antes de implementar la red LSTM y la ultima (c)) representa el uso de una SVM en el preprocesamiento de los datos. Dicha diferencia radica en la similitud de los puntos reales con los obtenidos por los distintos algoritmos, siendo mejor los obtenidos por el algoritmo MICE y la red LSTM, seguidos por la red LSTM de forma individual y al final la SVM en conjunto con la red LSTM.



(a) Implementando la red LSTM por si sola.

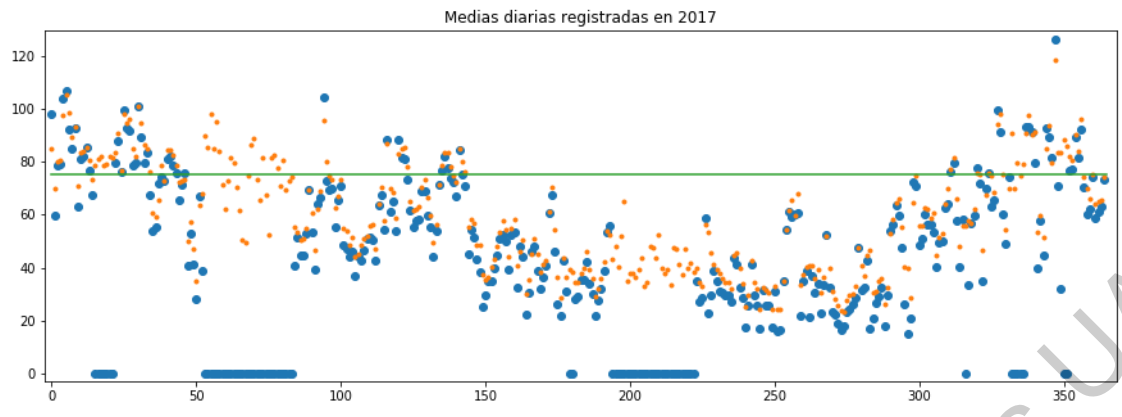


(b) Implementando una LSTM y MICE.

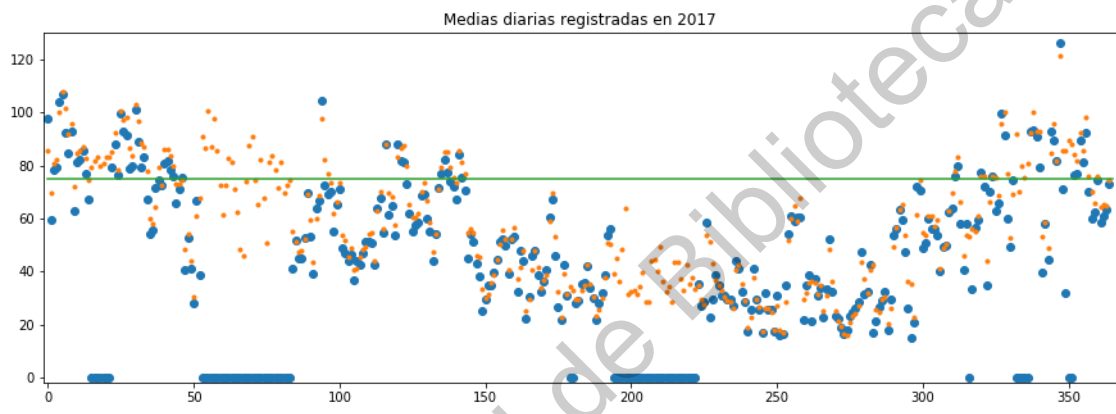


(c) Implementando una red MLP.

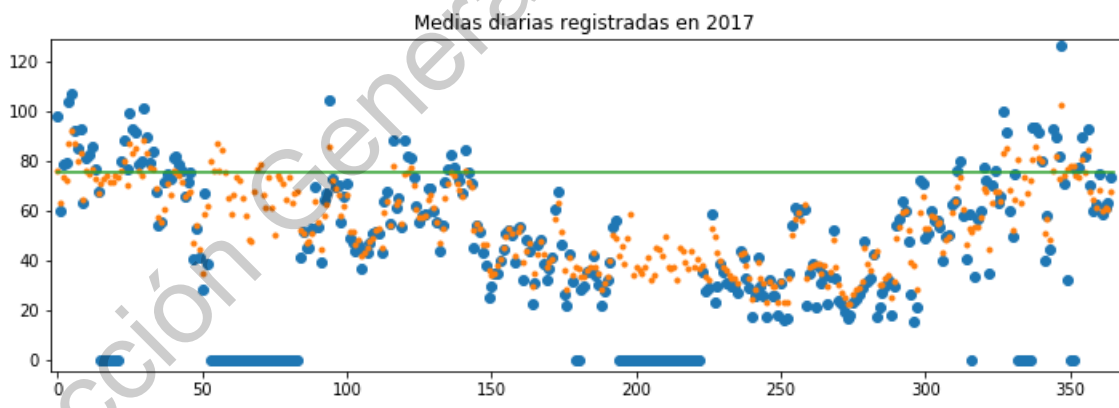
Figura 4.4: Valores medios del año 2016.



(a) Implementando la red LSTM por si sola.

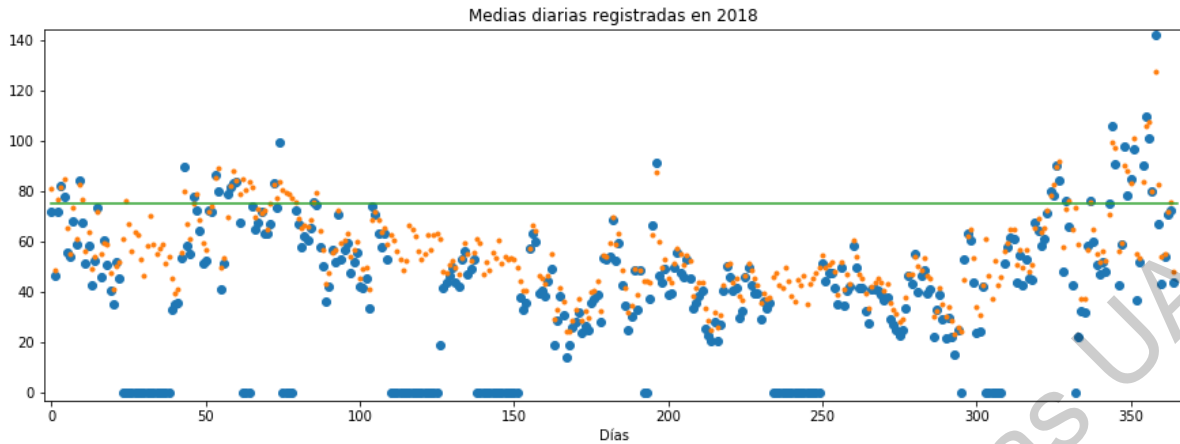


(b) Implementando una LSTM y MICE.

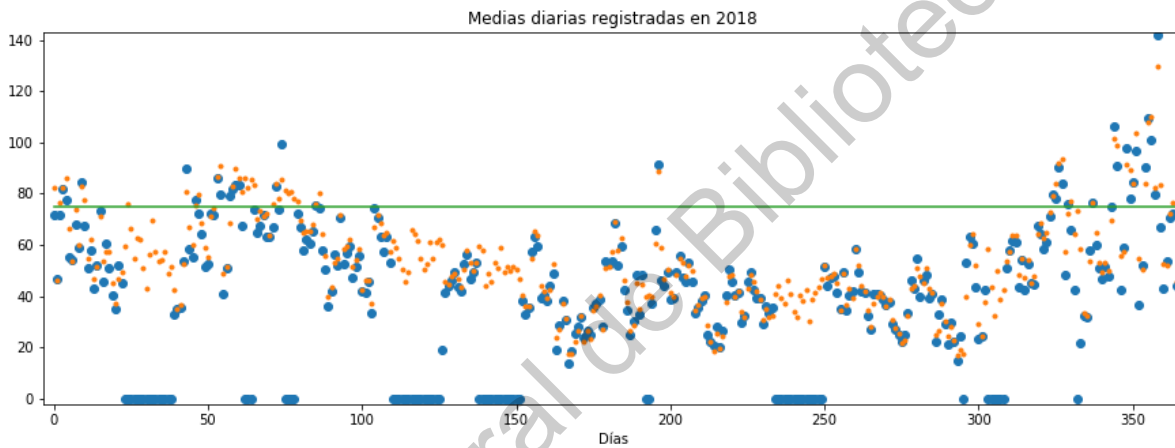


(c) Implementando una red MLP.

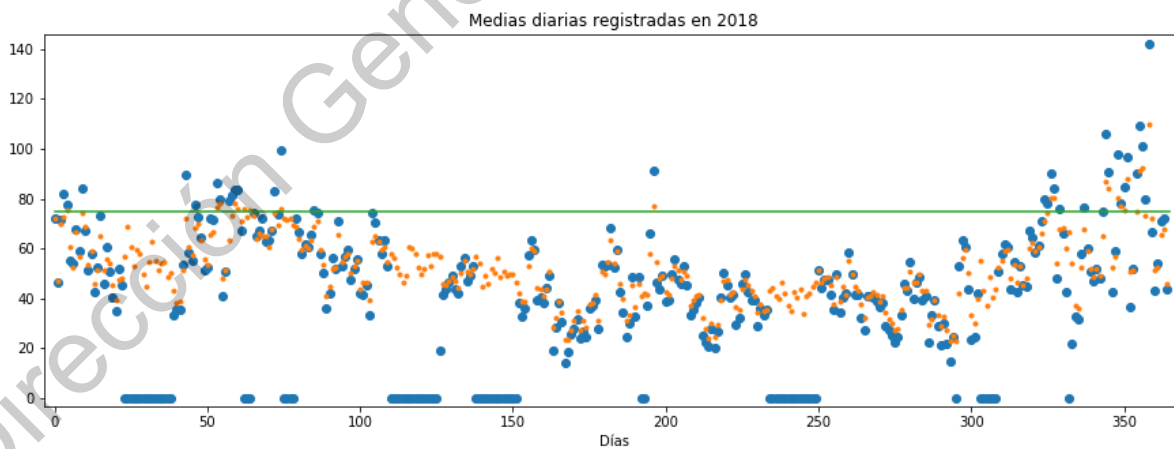
Figura 4.5: Valores medios del año 2017.



(a) Implementando la red LSTM por si sola.



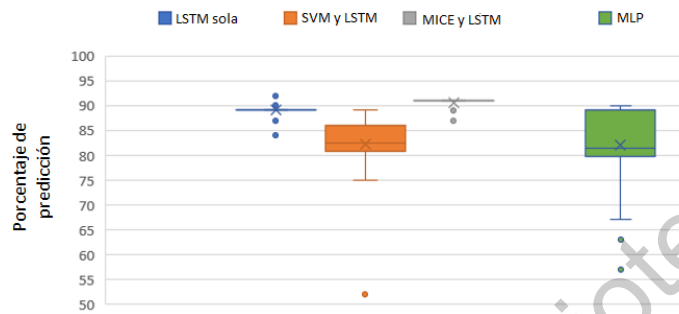
(b) Implementando una LSTM y MICE.



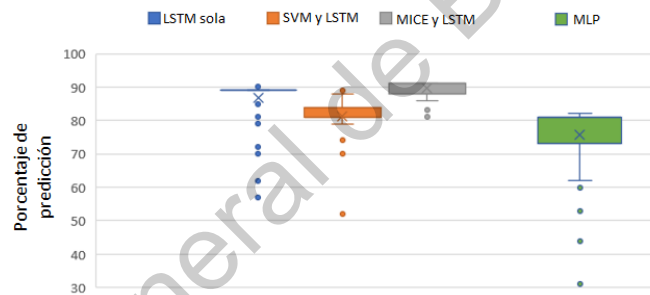
(c) Implementando una red MLP.

Figura 4.6: Valores medios del año 2018.

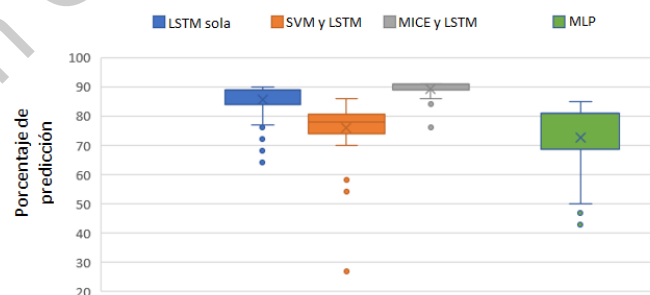
En la figura 4.7 muestra la repetitividad de los resultados con diferentes tiempos de anticipación. En las tres secciones se observa que la red LSTM cuenta con los mejores resultados pero a su vez cuenta con un rango de variabilidad, para la reducción de esta variabilidad el método de imputación MICE resulto ser muy bueno, a su vez se percibe una mejoría en el porcentaje de predicción con 48 y 72 horas de anticipación. Mientras que los porcentajes bajos fue una característica de la implementación de una SVM con la red LSTM, sin embargo cuenta con una compacticidad mayor que la red MLP, la cual resulta ser al algoritmo con un mayor rango de variabilidad.



(a) Comparación con 24 horas de anticipación entre la implementación de la red LSTM y tres diferentes formas de preprocesar los datos iniciales con una red MLP.



(b) Comparación con 48 horas de anticipación entre la implementación de la red LSTM y tres diferentes formas de preprocesar los datos iniciales con una red MLP.



(c) Comparación con 72 horas de anticipación entre la implementación de la red LSTM y tres diferentes formas de preprocesar los datos iniciales con una red MLP.

Figura 4.7: Resultados de la evaluación continua de la red propuesta junto con sus comparaciones.

Conclusiones

La red propuesta logro los resultados esperados al logran predecir una excedencia con 24, 48 y 72 horas de anticipación a que dicha excedencia ocurra, el algoritmo de imputación MICE ayudo a mejorar la repetitividad de la red disminuyendo la variación de los porcentajes de predicción obtenidos en los tres intervalos de tiempo. Esto es debido a que el algoritmo MICE propone de una mejor forma los valores faltantes en los registros en comparación con una media simple de los datos anteriores y posteriores. Al tener la mayor cantidad de datos reales, los propuestos influyen en menor forma permitiendo a la red una mejor modelación del comportamiento y esto hara que la segunda red pueda distinguir las diferencias sutiles que anteceden a una excedencia o a un día sin excedencia.

Al observar la alta repetitividad y compacticidad de los porcentajes de predicción permite afirmar la robustes del algoritmo al menos para este caso de estudio, tratandose de las condiciones particulares presentes en la ciudad de México, México. Con esto se logra cumplir el primer objetivo particular del proyecto.

Al ser comparados los resultados con los obtenidos mediante una red MLP se observa la mejora significativa de la compacticidad de los porcentajes obtenidos, la red MLP es muy usada para comparar nuevas topologias de red, al ser esta una topologia clásica que ha mostrado su versatilidad en diferentes áreas.

De los algoritmos implementados el que arrojó los peores resultados con la red LSTM fue el uso de SVM para preprocesar los datos, esto puede ser debido a una modelación deficiente en comparación con los otros métodos utilizados, esto demuestra que la correcta modelación del comportamiento de las partículas PM10 asegura la mejor predicción de excedencias.

Finalmente uno de los objetivos era determinar las condiciones que anteceden a una excedencia, como la humedad, la temperatura, etc, pero como las partículas PM10 son muy diversas los elementos que las conforman no son igualmente afectados por dichas condiciones.

5.1. Trabajos futuros

- Uno de los posibles trabajos futuros seria la implementación de la red en un sistema embebido para poder monitorear y generar alertas tempranas para tomar precauciones.
- Comparar los resultados obtenidos con otros métodos utilizados en esta área, debido a que una red MLP a pesar de su amplio uso, ya no es posible comparar sus resultados con topologías

de redes más robustas.

- Probar con más registros ambientales ya sean de otro estado de la república o de algún otra zona de mundo.
- Comprobar la red neuronal propuesta con diversos contaminantes ambientales, con la finalidad de comprobar la robustes del algoritmo.

Dirección General de Bibliotecas UAQ

Referencias

- Abraham, A. (2005). Artificial neural networks. *Handbook of measuring system design*.
- Ahmad, J., Farman, H., y Jan, Z. (2019). Deep learning methods and applications. En *Deep learning: Convergence to big data analytics* (pp. 31–42). Springer.
- Azur, M. J., Stuart, E. A., Frangakis, C., y Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research*, 20(1), 40–49.
- Betancourt, G. A. (2005). Las máquinas de soporte vectorial (svms). *Scientia et technica*, 1(27).
- Bianchi, F.-M., Maiorino, E., Kampffmeyer, M., Rizzi, A., y Jenssen, R. (2017). Recurrent neural network architectures. En *Recurrent neural networks for short-term load forecasting* (p. 23–29). Springer. doi: 10.1007/978-3-319-70338-1.3
- Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., . . . Di Carlo, P. (2017). Recursive neural network model for analysis and forecast of PM10 and PM2.5. *Atmospheric Pollution Research*, 8(4), 652–659. doi: 10.1016/j.apr.2016.12.014
- Brunekreef, B., y Holgate, S. T. (2002). Air pollution and health. *The lancet*, 360(9341), 1233–1242.
- Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., y Vitabile, S. (2008). Three hours ahead prevision of SO2 pollutant concentration using an Elman neural based forecaster. *Building and Environment*, 43(3), 304–314. doi: 10.1016/j.buildenv.2006.05.011
- Buchanan, B. G. (2005). A (very) brief history of artificial intelligence. *Ai Magazine*, 26(4), 53–53.
- Cao, L.-J., y Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on neural networks*, 14(6), 1506–1518.
- CDMX. (2019). *Bases de datos - red automática de monitoreo atmosférico (rama)*. <http://www.aire.cdmx.gob.mx/default.php?opc=%27aKBh%27>. (Accedido 08-09-2019)
- Concepción Jiménez, M. d. l. (2015). Definición y validación de una metodología para correlacionar concentración de contaminantes atmosféricos e ingresos hospitalarios.
- Davidson, C. I., Phalen, R. F., y Solomon, P. A. (2005). Airborne particulate matter and human health: a review. *Aerosol Science and Technology*, 39(8), 737–749.
- Dietterich, T. G. (2002). Machine learning for sequential data: A review. En *Joint iapr international workshops on statistical techniques in pattern recognition (spr) and structural and syntactic pattern recognition (sspr)* (pp. 15–30).
- Galván, M. (2007). *Imputación de datos: teoría y práctica* (Vol. 54). United Nations Publications.
- Gualtieri, M. (2017). The forrester waveTM: Predictive analytics and machine learning solutions, q1 2017. *Forrester Research*.
- Guyon, I. (1991). Neural networks and applications tutorial. *Physics Reports*, 207(3-5), 215–259.
- Hassoun, M. H., y cols. (1995). *Fundamentals of artificial neural networks*. MIT press.
- Hopfield, J. J. (1988). Artificial neural networks. *IEEE Circuits and Devices Magazine*, 4(5), 3–10.
- Hua, Y., Zhao, Z., Li, R., Chen, X., Liu, Z., y Zhang, H. (2019). Deep learning with long short-term memory for time series prediction. *IEEE Communications Magazine*.
- Katsouyanni, K., Touloumi, G., Samoli, E., Gryparis, A., Le Tertre, A., Monopoli, Y., . . . others (2001). Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 european cities within the aphea2 project. *Epidemiology*, 12(5), 521–531.

- Khaniabadi, Y. O., Polosa, R., Chaturkova, R. Z., Daryanoosh, M., Goudarzi, G., Borgini, A., ... Naserian, P. (2017). Human health risk assessment due to ambient PM10 and SO₂ by an air quality modeling technique. *Process Safety and Environmental Protection*, 111, 346–354. Descargado de <https://doi.org/10.1016/j.psep.2017.07.018> doi: 10.1016/j.psep.2017.07.018
- LeCun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436–444.
- Lipton, Z. C., Berkowitz, J., y Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*.
- Maass, W., Joshi, P., y Sontag, E. (2007). Computational aspects of feedback in neural circuits. *PLoS Computational Biology*, 3(1). doi: 10.1371/journal.pcbi.0020165.eor
- Michie, D., Spiegelhalter, D. J., Taylor, C., y cols. (1994). Machine learning. *Neural and Statistical Classification*, 13.
- OMS. (2018). *Departamento de salud pública, medio ambiente y determinantes sociales de la salud*. <http://www.who.int/phe/health-topics/outdoorair/databases/health-impacts/es/index2.html>. (Accedido 20-11-2018)
- OPS. (2019). *Contaminación del aire ambiental*. http://https://www.paho.org/hq/index.php?option=com_content&view=article&id=12918:ambient-air-pollution&Itemid=72243&lang=es. (Accedido 11-10-2019)
- Park, J.-H., Yoo, S.-J., Kim, K.-J., Gu, Y.-H., Lee, K.-H., y Son, U.-H. (2017). Pm10 density forecast model using long short term memory. En *2017 ninth international conference on ubiquitous and future networks (icufn)* (pp. 576–581).
- Pennachin, C., y Goertzel, B. (2007). Artificial general intelligence. En (Vol. 2, pp. 1–30). Springer.
- Pratama, I., Permanasari, A. E., Ardiyanto, I., y Indrayani, R. (2016). A review of missing values handling methods on time-series data. En *2016 international conference on information technology systems and innovation (icitsi)* (pp. 1–6).
- Raicharoen, T., Lursinsap, C., y Sanguanbhokai, P. (2003). Application of critical support vector machine to time series prediction. En *Proceedings of the 2003 international symposium on circuits and systems, 2003. iscas'03*. (Vol. 5, pp. V–V).
- Ramachandran, P., Zoph, B., y Le, Q. V. (2017). Searching for activation functions. *arXiv preprint arXiv:1710.05941*.
- Rojas, R. (2013). *Neural networks: a systematic introduction*. Springer Science & Business Media.
- Russell, S. J., y Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,.
- Salas, R. (2004). Redes neuronales artificiales. *Universidad de Valparaíso. Departamento de Computación*, 1.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85–117.
- Stadlober, E., Hörmann, S., y Pfeiler, B. (2008). Quality and performance of a pm10 daily forecasting model. *Atmospheric Environment*, 42(6), 1098–1109.
- Terashima, T., Wiggs, B., English, D., Hogg, J. C., y van EEDEN, S. F. (1997). Phagocytosis of small carbon particles (pm10) by alveolar macrophages stimulates the release of polymorphonuclear leukocytes from bone marrow. *American journal of respiratory and critical care medicine*, 155(4), 1441–1447.
- Vicente, A. B., Juan, P., Meseguer, S., Díaz-Avalos, C., y Serra, L. (2018). Variability of PM10 in industrialized-urban areas. New coefficients to establish significant differences between

- sampling points. *Environmental Pollution*, 234, 969–978. doi: 10.1016/j.envpol.2017.12.026
- Wilamowski, B. M. (2009). Neural network architectures and learning algorithms. *IEEE Industrial Electronics Magazine*, 3(4), 56–63.
- Zhang, Z. (2016). Multiple imputation with multivariate imputation by chained equation (mice) package. *Annals of translational medicine*, 4(2).

Dirección General de Bibliotecas UAQ

Apéndices

Artículo publicado

2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)
Mexico City, Mexico. September 11-13, 2019

Evaluation of a Recurrent Neural Network LSTM for the Detection of Exceedances of Particles PM10.

Julio Alberto Ramírez Montañez.
Facultad de ingeniería.
Universidad Autónoma de Querétaro.
Querétaro, México.
julio.ramirez27@gmail.com

Marco Antonio Aceves Fernandez.
Facultad de ingeniería.
Universidad Autónoma de Querétaro.
Querétaro, México
marco.aceves@gmail.com

Saúl Tovar Arriaga
Facultad de ingeniería.
Universidad Autónoma de Querétaro.
Querétaro, México.
saul.tovar@uaq.mx

Juan Manuel Ramos Arreguin
Facultad de ingeniería.
Universidad Autónoma de Querétaro
Querétaro, México.
jsistdig@yahoo.com.mx

Giovanni Angelo Salini Calderon
Facultad de Ingeniería.
Universidad Católica de la Santísima Concepción.
Chile
giovanni.angelo@gmail.com

Abstract—Monitoring air quality is a topic of current interest, since poor quality has a negative impact on health. Air quality is affected by different pollutants, such as particulate matter and gases, produced by the growing industrial development. As a preventive measure, Mexico established different standards in order to control airborne pollution. In this paper, we propose a methodology based upon a recurrent long-term/short-term memory network for the prediction of exceedances of PM10 (particles of less or equal diameter than 10 micrometers) with time intervals of 72, 48 and 24 hours in advance. Obtaining a satisfactory percentage of prediction as a whole a minimum variability in repetitive experimental runs.

Keywords— Air pollution, neural network (NN), Deep Network.

I. INTRODUCTION

The industrial development achieved today has not only brought great benefits and facilities for daily life, but has also caused environmental problems. Air pollution is one of these problems, which seriously affects health, generating various ills [1]. In response, environmental monitoring facilities have been developed, which record the various concentrations of pollutants in a defined time interval. Like several variables that affect these concentrations such as: temperature, relative humidity, wind direction, and the geographical area where the measurement is recorded [2]. The interaction of these variables makes the behavior of the particulate matter highly random, increasing the complexity of modeling their behavior. The particulate material being an air-pollutant analyzed in this article. As they are values recorded at defined time intervals, recurrent neural networks have demonstrated their efficiency in modeling the behavior of data recorded with the same capture frequency [3].

First of all, the Particulate material includes all suspended particles, whether its origin is natural or artificial. Therefore, the mechanisms to which these particles are dispersed through the atmosphere are different in comparison to most of the harmful gasses in the atmosphere. Some of these particles are of such a

small size that they are able to enter the bloodstream by breathing them, causing serious health problems [1,2].

Some of these particles include: PM10, PM2.5 and ultra-fine. Where PM means "Particulate Matter" and the number then determines the maximum size in microns (μm) as mentioned by Concepción-Jiménez [1]. From the above mentioned classification PM10 particles were selected, as the case study, which are mainly composed of substrates, nitrates, mineral powders, and pollen [4]. In the case of the predictive analysis of the neural network to be used, the focus was on modelling the behaviour of PM10 particles as well as on the characteristics that precede their exceedance.

They are called exceedance when the PM10 particle value exceeds the values defined by some standard. For example, the Mexican standard NOM-025-SSA1-2014 establishes a daily average of $75 \mu\text{g}/\text{m}^3$ and $40 \mu\text{g}/\text{m}^3$ annually. By definition of the standard, it is necessary to know the registered values of pollutants during a day to determine if the average is below or above the value of the norm, in the same way to know the annual average value.

The main reason for the study of the exceedances is that it has been shown that prolonged exposure to high concentrations of particles in the environment (PM10) is related to increased respiratory and cardiovascular problems. Similarly, the World Health Organization (WHO) has proposed standards, even though it is estimated that 91% of the world population lives in places where the established guidelines are not followed with respect to the established levels [5].

The neural network artificial are inspired by the very neurons that make up a brain, they were developed to simulate the ability to learn, using simple elements. These elements can be classified into input, processing and output elements; all these can be interconnected or not [3,6]. In Figure 1, a schematic of a basic neural network is shown.

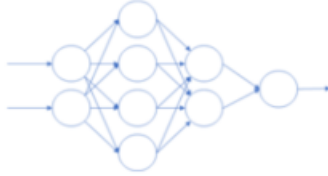


Fig. 1 Example of a four-layer neural network (figure adapted from [3]).

Warren McCulloch and Walter Pitts pioneers in the development of a computer model for an artificial neuron, the basic element of a neuron network, called threshold logic [7]. Over the years, different network topologies have been proposed, each with advantages to solve different problems, whether they are classification, detection, characterization, etc.

In 1997, Hochreiter and Schmidhuder proposed the LSTM model (Long Short-Term Memory), unlike a simple recurrent neural network that has a long-term memory in the form of weights, which are modified during the training of the network and short-term memory defined in activation functions between the communication of the nodes of neurons [8]. The LSTM model introduces a block of internal memory, composed of simple blocks connected in a specific way (Figure 2), each of them are described as follows [9]:

- Input node: represented by "a1" is a node whose activation is usually a sigmoid, which weights the input values.
- Input gate: "a2" is a characteristic of the LSTM. Take the current points in conjunction with the data from a previous step. Managing a logical flag if it is zero cuts the data flow between the nodes and if one is the data pass through it. Acting as a block memory control.
- Internal state: it is the most important part, referenced with an "S", this node has a recurring connection, this connection is usually called a carousel of constant errors. This block prevents the error from increasing.
- Oblivion gate: "a4" introduced by Ger in 2000, provides a method by which the network can adjust to the content of the internal state. Especially for the use of time series.
- Output gate: "output" is the value produced by the value of the internal state multiplied by the value of the output gate "a3", to ensure this output the activation function of the first state is a function tanh.

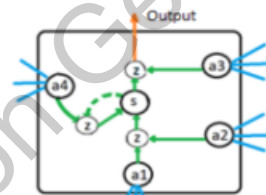


Fig. 2 Basic block of an LSTM (figure adapted form [9]).

Currently, there are few works on the prediction of exceedances of PM10 particles, from the last developed works we found the one made by Park [3], which consisted of the

prediction of exceedances based on the records of the previous days, comparing a linear regression, with a LSTM network, using data recorded in Seoul, Korea, between January 2015 and March 2016.

The LSTM network was evaluated with different hidden neuron values in conjunction with different lot size values. The best result obtained was 3.54 of MSE and 1.86 of RMSE, in the modeling of the behavior of PM10 particles, with these results the behavior predicted by the network allowed the prediction of exceedances in a previously defined day [3].

Another work of 2002, the author Brunelli generated a mathematical model for the prediction of exceedances of SO₂, being able to predict them three hours in advance [6]. The main disadvantage of the previous work is that the SO₂ particles are easily modelled by means of the behavior of the temperature in the environment.

II. MATERIAL AND METHODS.

A. Materials.

Mexico has a constant monitoring network in Mexico City (CDMX) called the Atmospheric Monitoring Network (RAMA). This network has 24 stations distributed in different municipalities. Each station registers in each hour the levels of different atmospheric pollutants such as the level of ozone, nitrogen dioxide, nitrogen oxides, nitrogen monoxides, sulphur dioxide, carbon monoxides, PM10, PM2.5, among others [10].

This database is available on RAMA [10]. In this work used the data from the Merced monitoring station (MER), from the year 2000 to 2018 for training purposes. In Figure 3, it is observed its location being a central point.

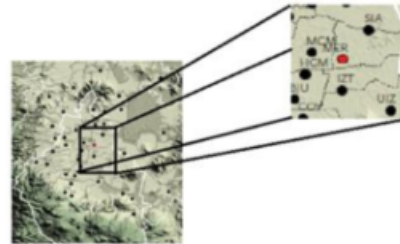


Fig. 3 Location of monitoring stations and MER station (red color) (figure adapted from 11).

B. Model construction.

Before defining the structure of the network, it is necessary to scale the training data in a range of -1 to 1, with the purpose of decreasing the non-linearity of the initial data, facilitating the training of the LSTM neural network.

- Modeling neural network.

This network has 50 neurons in the input layer followed by a hidden layer of 256 neurons and ends with a complete connection layer to which it produces the output vector [12].

- ❖ Layer 1 consists of 50 LSTM neurons, which will take the first 50 data and expand them to feed the second layer, generating a record of them.
- ❖ Layer 2 consists of 256 LSTM neurons.

- ❖ Layer 3 is a simple neuron, which based on the previous recorded data will generate a new value, successively.

- Predictive neural network.

For this the number of incoming neurons is determined by the length of the input vector determined by the number of values prior to evaluating, since 24, 48 and 72 hours, each one represents the previous hours registered to an exceedance and days without exceedances. In its second layer, it is an output layer of a neuron.

- ❖ Layer 1 each neuron receives a value from the input data vector, which generates an output response.
- ❖ Layer 2 receives the results of the first network and generates a classification, given that initially the days with and without exceedances are known, the network uses a continuous regression to adjust its weights and obtain the expected result.
- Complete model.

Figure 4 shows the connection between the two neural networks; the first three layers are from the modeling network of the behavior of the PM10 and the next two layers are from the classification network. Once the two networks previously included are present, the first network supplies a previously established length vector (between 72, 48 and 24) and the second network provides a flag if the day after the length of the vector will be an exceedance or not.

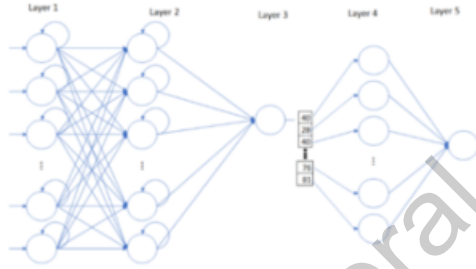


Fig. 4 Complete model.

C. Evaluation of the model.

To evaluate the correct modeling generated by the LSTM network, the Mean Square Error (MSE) and the Root Mean Square Error (RMSE) were implemented, the MSE measures the average of the squared errors, in other words, it measures the difference of the real data and the estimates, its equation is represented below:

$$MSE = \frac{1}{n} \sum_{i=1}^n (YM_i - YR_i)^2 \quad (1)$$

Where: YM is the vector of values modeled by the LSTM network, YR represents the vector of the actual comparison values, and n represents the length of the analyzed vector, both the YM and YR vector must be of the same length

Meanwhile, the RMSE represents the standard deviation between actual and predicted values, and it is simply to apply a square root to the MSE, generating the following equation:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (YM_i - YR_i)^2} \quad (2)$$

D. Methodology.

The following diagram (Figure 5) describes the methodology used during the development of this work.

The study data were acquired from the RAMA database of Mexico City, from January 2000 to December 2018. Due to the industrial and residential development of the city, several monitoring stations were incompatible or removed from the network during the course of the years, the first step is to detect the constant stations from 2000 to date, thus reducing the possible stations for the study. Subsequently, these records were evaluated, in order to detect that station with the highest number of valid records with which to train the neural network.

In the sections where the data is not valid, it is proposed to replace it with values generated by an average between the data before and after said section, in order to continue in a certain way with the behavior of PM10 particles. Before supplying the values to the network, it was necessary to generate a single data vector, in order to have a continuous vector from 2000 to 2017.

The LSTM network uses an 80-20 ratio for training and validation, 80 represents the percentage of data used for training while the remaining 20 is used for the validation of the model, the results will be evaluated through the MSE and the RMSE.

In a parallel cycle the original data will be grouped into sections of 24 data, this will allow to know the average recorded individually for the days from 2000 to 2015. Identifying the days with exceedances and also keeping a vector of the conditions prior to the registration of said leave. These records will be used to train the second network.

Having the modeling of PM10 particles provided by the first network and the second being able to anticipate the day of leave, the years 2017 and 2018 were used to evaluate both networks and observe their ability to predict together.



Fig. 5 Methodology used.

III. RESULTS AND DISCUSSIONS.

The recurrent network LSTM shows a mean root mean square error (RMSE) of 18.0342, both are relatively good results, in Figure 6 we can appreciate the behavior of the data recorded by the station, blue and orange color shows the behavior predicted by the network, it may not reach the highest peaks but you can perceive a similarity in the behavior of both graphs.

By increasing the scale of the graph, it is possible to observe the similarities in the behavior of both registers, in Figure 7 the behavior can be observed in the first 500 hours of the year.

In the areas where the sensor had measurement problems, registering outliers, the predicted values are able to complete a behavior similar to the one that has no valid data.

In Figure 8 the exceedances registered by the monitoring station are observed in blue, while in orange the ones registered by the data supplied by the LSTM network and evaluated by the classification network are shown. The green line shows the maximum value allowed by the Mexican norm.



Fig. 6 Comparison between true value and predict value.

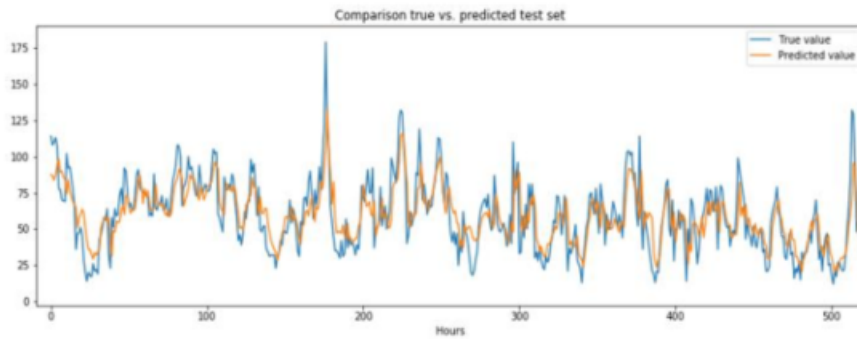


Fig. 7 Increased scale.

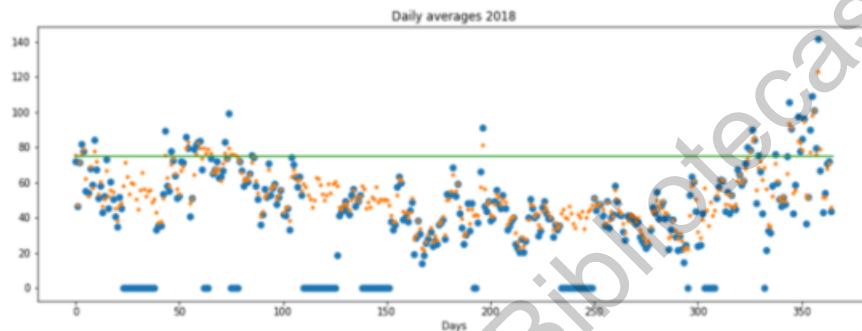


Fig. 8 Comparison of actual and predicted surpluses by the neural network (self-image).

The results of the prediction vary depending on the anticipation time with which work is carried out, the shorter the time the more stable the predictions are when evaluating repetitively. However, 72 hours is perceived in occasions considerably lower, obtaining 9% prediction for a prediction. The reason for this, needs further investigation. In the three time intervals have a steady result of around 89% prediction, which shows consistency for most test runs. In Figure 9, it is shown the variations obtained by evaluating the network 50 times.

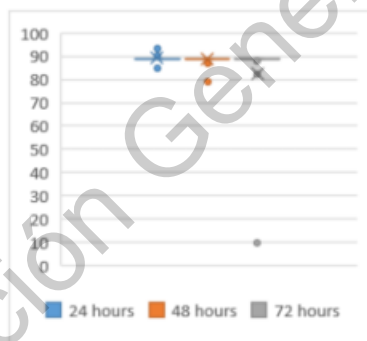


Fig. 9 Variations in the prediction percentage.

IV. CONCLUSIONS.

With the present work, the great versatility of an LSTM network and a simple neural network for the prediction of exceedances in an interval of 48 and 24 hours previous was demonstrated. In the case of 72 hours in advance, it will be crucial to investigate the reasons for some predictions lower than expected and improve upon the modeling of the behavior of the PM10 particles in order to reduce or eliminate the low percentages of prediction obtained regularly.

Alternatives of solution can be several, initially it is possible to revise the proposal of the missing data not registered by the stations, in this work an average was made between the behavior before and after the registration of outliers. Another solution could be to generate an average by comparing the behavior at the precise moment in which the sensor recorded the -99 with the value recorded previous or later years.

The next step would be to review the training of the LSTM network, with numerous possible combinations in the parameters that characterize the network architecture.

REFERENCES

- [1] Concepción Jiménez, M. (2015). Definition and validation of a methodology to correlate concentration of atmospheric pollutants and hospital admissions. (Master's Thesis). University of Seville, Seville.

- [2] Stadlober, E., Hörmann, S., & Pfeiler, B. (2008). Quality and performance of a PM10 daily forecasting model. *Atmospheric Environment*, 42 (6), 1098-1109.
- [3] Park, J.-h., Lee, K.-h., Kim, K.-j., y Son, U.-h. (2017). PM10 Density Forecast Model Using Long Short Term Memory, 576-581.
- [4] Air quality and health. Obtained from the World Health Organization. Retrieved from [www.who.int/en/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](http://www.who.int/en/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health).
- [5] OMS. (2018a). Calidad del aire y salud. [urlhttp://www.who.int/es/news-room/factsheets/detail/ambient-\(outdoor\)-air-quality-and-health](http://www.who.int/es/news-room/factsheets/detail/ambient-(outdoor)-air-quality-and-health). (Accedido 20-11-2018).
- [6] Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., y Vitabile, S. (2008). Three hours ahead prevision of SO2 pollutant concentration using an Elman neural based forecaster. *Building and Environment*, 43 (3), 304-314. doi:10.1016/j.buildenv.2006.05.011.
- [7] Hassoun, M. H. (1995). *Fundamentals of artificial neural networks*. MIT press.
- [8] Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*.
- [9] Hua, Y., Zhao, Z., Li, R., Chen, X., Liu, Z., & Zhang, H. (2019). Deep Learning with Long Short-Term Memory for Time Series Prediction. *IEEE Communications Magazine*.
- [10] Databases - Automatic Air Monitoring Network (RAMA). Obtained from the World Health Organization. Retrieved from www.aire.cdmx.gob.mx/default.php?opc=%27aKBh%27
- [11] Monitoring stations. Obtained from www.aire.cdmx.gob.mx/default.php?opc=%27ZaBhnmI-&dc=%27ZA=
- [12] Chen, K., Zhou, Y., & Dai, F. (2015, October). A LSTM-based method for stock returns prediction: A case study of China stock market. In *2015 IEEE International Conference on Big Data (Big Data)* (pp. 2823-2824). IEEE.
- [13] Lenschow, P., Abraham, H. J., Kutzner, K., Lutz, M., Preuß, J. D., & Reichenbacher, W. (2001). Some ideas about the sources of PM10. *Atmospheric Environment*, 23-33
- [14] Brunekreef, B., & Holgate, S. T. (2002). Air pollution and health. *The lancet*, 360(9341), 1233-1242.

Dirección General de Bibliotecas UAQ



**Centro de Investigación y de Estudios Avanzados
del Instituto Politécnico Nacional**



The Organizing Committee of the 2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE) awards the present

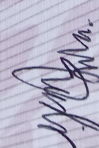
RECOGNITION TO:

Julio Alberto Ramírez-Montañez, Marco Aceves-Fernandez, Saul Toyar-Arriaga, Juan-Manuel Ramos-Arreguin and Giovanni Angelo Salini Calderon

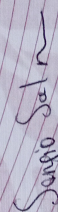
For the presentation of the paper entitled:

Evaluation of a Recurrent Neural Network LSTM for the Detection of Exceedances of Particles PM10.

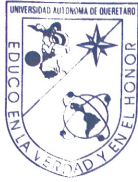
Mexico City, Mexico
September 11-13, 2019


Dr. Gerardo Silva Navarro
Co-Chair of CCE


Dr. Wen Yu Liu
Co-Chair of CCE


Dr. Sergio Salazar Cruz
Co-Chair of CCE

Requisito de manejo de la lengua inglesa



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO FACULTAD DE LENGUAS Y LETRAS



A QUIEN CORRESPONDA:

La que suscribe, Directora de la Facultad de Lenguas y Letras, hace **C O N S T A R** que

RAMIREZ MONTAÑEZ JULIO ALBERTO

Presentó el **Examen de Manejo de la Lengua** efectuado el día veintidós de octubre de dos mil diecinueve, en el cual obtuvo la siguiente calificación:

8-

Se extiende la presente a petición de la parte interesada, para los fines escolares y legales que le convengan, en el Campus Aeropuerto de la Universidad Autónoma de Querétaro, el día veintiocho de octubre de dos mil diecinueve.

Atentamente,
"Enlazar Culturas por la Palabra"



LIC. LAURA PÉREZ TELLEZ

LPT/evm*CL*FLL-C.-705



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

Campus Aeropuerto, Anillo Vial Fray Junípero Serra S/N, Querétaro, Qro.
C.P. 76140, Tel. 01 (442) 192 12 00 Dirección Ext. 61010,
Secretaría Administrativa 61300, Posgrado 61140, Licenciatura 61070,
Centro de Lenguas 61050, Secretaría Académica 61100 y Planeación 61110

SOMOS UAQ
EDUCAR CRECER CONSOLIDAR

Dirección General de Bibliotecas UAQ