
This is the **published version** of the master thesis:

Serna Pérez, Verena; Ramos Terrades, Oriol , tut. Utilización del modelo autorregresivo de series temporales ARIMA para la predicción de visitas en el servicio de urgencias del Hospital Clínico Universitario de Valladolid. 2024. 10 pag. (Màster en Intel·ligència Artificial i Big Data en Salut)

This version is available at <https://ddd.uab.cat/record/304358>

under the terms of the  license

Utilización del modelo autorregresivo de series temporales ARIMA para la predicción de visitas en el servicio de urgencias del Hospital Clínico Universitario de Valladolid

Alumna: Verena Serna Pérez

Tutor: Oriol Ramos Terrades

Resumen

El aumento de la demanda de servicios sanitarios es un reto importante para los servicios de urgencias de un hospital. Esta demanda genera presión sobre los recursos y la calidad de atención y hace que la gestión del flujo de pacientes y la asignación de recursos sea elementos clave para el buen funcionamiento del servicio. Los servicios de urgencias son el primer punto de contacto para los usuarios que buscan atención médica urgente. La predicción precisa de la afluencia de pacientes puede resultar muy útil para optimizar el uso de los recursos y mejorar la atención al usuario. En este proyecto se propone la utilización de inteligencia artificial (IA) para desarrollar un modelo de aprendizaje automático basado en la regresión de series temporales que permita predecir el número de visitas que recibirá el servicio de urgencias del Hospital Clínico Universitario de Valladolid en 2023, con los datos de visitas realizadas en 2021 y 2022.

Palabras clave: Urgencias, Big Data, Aprendizaje Automático, Regresión de Series Temporales, Modelización Predictiva, Gestión Sanitaria, ARIMA.

Abstract

The increasing demand for healthcare services poses a significant challenge to hospital emergency departments (EDs). This surge in demand strains resources, compromises quality of care, and highlights the importance of patient flow management and resource allocation for effective ED operation. EDs serve as the first point of contact for patients seeking urgent medical attention. Accurate patient influx prediction can prove invaluable in optimizing resource utilization and enhancing patient care. This project proposes the employment of artificial intelligence (AI) to develop a time series regression-based machine learning model for predicting the number of patients visits to the ED of Hospital Clínico Universitario de Valladolid in 2023, having the data from 2021 and 2022.

Keywords: Emergency Department, Big Data, Machine Learning, Time Series Regression, Predictive Modeling, Healthcare Management, ARIMA.

1- Introducción

La gestión eficiente de los servicios de urgencias hospitalarios es crucial para garantizar una atención de calidad y poder dar respuesta a las necesidades de los pacientes que acuden a ellos [1]. Sin embargo, los servicios de urgencias enfrentan desafíos significativos debido al aumento de la demanda [2,3].

La justificación detrás de este proyecto surge de la necesidad de aportar una posible herramienta que pueda ayudar en la gestión de este servicio y en la toma de decisiones. Se propone una herramienta basada en Inteligencia Artificial (IA), concretamente en la utilización del modelo autorregresivo de series temporales ARIMA.

Dado el crecimiento exponencial que ha experimentado el campo del aprendizaje automático y su acogida en el ámbito de las ciencias de la salud, se ha considerado explotar la utilidad de la IA para crear esta herramienta. Con el aumento en la disponibilidad de datos y registros médicos electrónicos como la historia clínica de los pacientes, imágenes médicas, secuencias genómicas y mucha más información accesible gracias a los dispositivos inteligentes y el Internet de las Cosas (IoT), la IA puede extraer patrones, identificar correlaciones y predecir resultados de salud con gran precisión y velocidad. Esto puede traducirse en diagnósticos más precisos, pronósticos más fiables y tratamientos personalizados que se adaptan a las necesidades individuales de los pacientes [4]. Así, la inteligencia artificial se ha posicionado como una herramienta especialmente útil en el ámbito de la salud, transformando la forma en que se diagnostican, tratan y previenen enfermedades [5].

El interés y uso de la IA en la investigación médica se refleja en la imagen 1, que muestra la evolución del número de artículos de investigación médica que incluyen la palabra “inteligencia artificial” en el repositorio de PubMed. Se observa un notable incremento en la cantidad de artículos desde 2017 hasta

la actualidad, pasando de 8,240 artículos a 38,260 artículos en 2023 [Imagen 2].

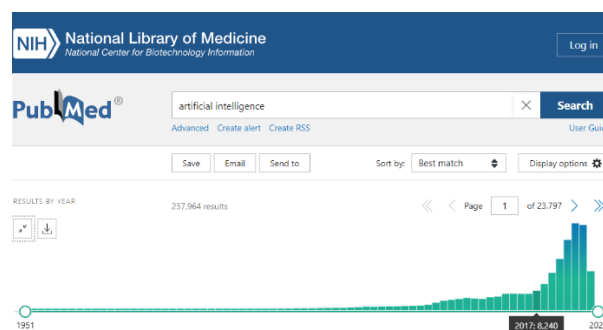


Imagen 1. Gráfico extradido de la web PubMed que muestra el resultado de la búsqueda de ‘artificial intelligence’ desde su primera mención en un artículo, hasta el año 2024, que está en curso. Se muestran los registros de 2017 resaltados.

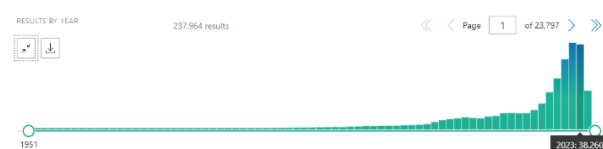


Imagen 2. Gráfico extradido de la web PubMed que muestra el resultado de la búsqueda de ‘artificial intelligence’ desde su primera mención en un artículo, hasta el año 2024, que está en curso. Se muestran los registros de 2013 resaltados.

2- Hipótesis y objetivos

Utilizar las variables obtenidas de los registros oficiales en un modelo de aprendizaje automático autorregresivo por series temporales predice el volumen de pacientes que recibirá el servicio de urgencias.

Existen variables que se pueden relacionar con las visitas por urgencias de los usuarios.

3- Estado del arte

Diversos estudios en la literatura han explorado el uso de inteligencia artificial para predecir el volumen de pacientes en los servicios de urgencias, los flujos de pacientes y las admisiones utilizando datos de triaje, entre otros. Estos estudios han demostrado que el uso de técnicas de aprendizaje automático mejora

significativamente la precisión de la predicción de la demanda en los servicios de urgencias hospitalarias. Por ejemplo, el estudio realizado por Furian mostró que los modelos de aprendizaje automático (en inglés “*Machine Learning*”) pueden predecir de manera precisa la admisión hospitalaria utilizando información del triaje y el historial del paciente, mejorando así el rendimiento del servicio de urgencias [6]. Este estudio respalda la intención de utilizar datos históricos para mejorar la precisión de nuestra predicción.

Otro estudio relevante es la investigación de Yue Hu, quien indicó que los predictores en tiempo real, como los recuentos recientes de llegadas de pacientes, el clima, las tendencias de Google y la información de comorbilidades, tienen un poder predictivo significativo y mejoran la precisión de los modelos de predicción de la demanda del servicio de urgencias [7].

La mejora de los flujos de pacientes y la capacidad de predecir el volumen de visitas al servicio de urgencias es una temática recurrente en la literatura. Por ejemplo, el estudio de Sharafat y Bayati comparte la creación de “*PatientFlowNet*”, un modelo de red neuronal convolucional. Este modelo demostró la capacidad de los modelos de aprendizaje profundo para predecir con precisión las tasas de flujo de pacientes utilizando múltiples variables y dependencias temporales. Este enfoque resalta la importancia de considerar una variedad de factores y técnicas avanzadas para mejorar las predicciones [8].

Los servicios de urgencias hospitalarias son los más frecuentados por la población y una tercera parte de los casos se consideran de utilización inadecuada [9].

El incremento del número de visitas que recibe año tras año, tensiona cada vez más la capacidad de respuesta y la calidad de la atención en el servicio de urgencias, especialmente aumenta la presión asistencial que tiene personal de enfermería, enfermeras/os y TCAIs (técnico/a en cuidados auxiliares de enfermería) y supone

un desafío en la gestión del personal necesario para cubrir el servicio. Predecir el número de visitas que recibirá el servicio podría ser clave para poder gestionar adecuadamente el equipo profesional, además de una herramienta para la gestión del material fungible y el “*stock*” de farmacia en la unidad.

Es por ello que se ha decidido investigar la posibilidad de crear un algoritmo que permita predecir el número de visitas que recibirá el servicio de urgencias del Hospital Clínico Universitario de Valladolid. El Hospital tiene un área de población asignada de aproximadamente 235.000 habitantes [10].

La opinión del personal que trabaja en los servicios de urgencias es que hay ciertas variables que parecen influir en la afluencia de pacientes, y esta opinión se forja según la experiencia de los profesionales en el servicio. Se podría considerar que en los servicios de urgencias de diferentes hospitales elucubran sobre las mismas líneas. Prueba de ello podría ser el equipo del hospital de Bilbao, que realizó un estudio donde querían comprobar el efecto del fútbol en el número de visitas a su servicio. Este reveló diferencias cuando había un partido en juego, se reducían las visitas al servicio de urgencias cuando había un partido de alta relevancia [11].

La temperatura y el clima también son variables que han sido estudiadas como posible factor determinante para afectar al número de visitas a servicios de atención sanitaria. Se realizó un estudio sobre el impacto del clima extremo en el número de ingresos en Catalunya dando como resultado diferencias palpables según la temperatura ambiental [12].

Otra teoría generalizada es que el día de la semana es determinante. Por ello, en el hospital Hartford, hospital norteamericano terciario con más de 75.000 pacientes recibidos en urgencias al año, se realizó un estudio para comprobar cómo afecta el día de la semana a las visitas recibidas. El estudio reflejó un aumento de visitas

significativo durante los lunes, y un aumento de revisitas durante el fin de semana [13].

Los días con un nivel de contaminación ambiental más elevado también repercuten sobre las visitas de urgencias, siendo los principales contaminantes atmosféricos según el departamento de Medi Ambient de la Generalitat de Catalunya y la OMS, son el dióxido de nitrógeno (NO₂), el dióxido de azufre (SO₂) el Ozono (O₃) y las partículas en suspensión (PM₁₀) [14,15].

Dada la proliferación de estudios basados en las creencias del propio personal sobre las causas de afluencia a urgencias, no es descabellado pensar en esos motivos como variables a tener en cuenta. Así pues, este proyecto contará con variables que pueden influir en el hecho que una persona decida acudir al servicio de urgencias.

Se ha decidido la implementación de un modelo predictivo basado en el modelo ARIMA (*“Autoregressive Integrated Moving Average”*). Es una clase de modelo estadístico utilizado para la predicción de series temporales debido a su capacidad para modelarlas y capturar de manera precisa la estructura temporal de los datos. Combina la auto-regresión (AR), la adaptación a tendencias (I) y el aprendizaje de errores (MA) para pronosticar series temporales de manera efectiva, son capaces de capturar patrones de dependencia temporal en los datos [16,17]. Eso los convierte en una buena herramienta para pronosticar valores futuros en series temporales como las visitas a urgencias. Esta predicción se podría ver beneficiada por la aportación al modelo de datos demográficos, climatológicos, de contaminación ambiental y eventos. Los estudios revisados mencionados, proporcionan una base sólida para nuestra metodología y respaldan la relevancia de utilizar la inteligencia artificial junto una amplia gama de datos/variables para la realización de predicciones.

4- Material y métodos

Los datos de los registros de las visitas a urgencias, se han obtenido de la página web del ministerio de transformación digital. Se trata de datos de carácter público que publica la Junta de Castilla y León y recoge las urgencias hospitalarias atendidas durante el último mes, actualizándose el día 8 de cada mes. En concreto se han obtenido los registros de las visitas recibidas en el servicio de urgencias durante los años 2021, 2022 y 2023.

También se han obtenido los registros de mortalidad y natalidad mensual de hombres y mujeres, los datos de habitantes totales, los registros meteorológicos de Valladolid también de los años 2021 y 2022, así como el registro de contaminantes ambientales de ambos años y por último el registro de los partidos jugados por el equipo local de fútbol.

4.1- Datos

El primer paso realizado con los datos obtenidos de las visitas es analizar su contenido y proceder a su limpieza. Se han buscado valores faltantes (*“missings”*) y outliers, y han sido eliminados o modificados teniendo en cuenta el tipo de dato que es. En el caso de datos faltantes, en ambos registros se han localizado diversos missings en cinco columnas diferentes. De esos datos faltantes, pese a ser en columnas diferentes, la gran mayoría coincidía en la misma fila, dando a entender que varios registros están incompletos en diversas columnas. Se ha decidido eliminar todas las filas con *“missings”* en todos los años. Se ha optado por esa opción por que en el caso del año 2021 corresponden al 3,81% de los datos totales de ese año y en 2022 al 3,96%. Se ha considerado que su eliminación no debería tener un impacto significativo en el resultado. En otros casos se han eliminado columnas que se han considerado redundantes como por ejemplo en los tres registros de visitas, cada uno correspondiente a un año, que están compuestos por diez columnas

(Fecha, Día, Hora, Zona básica de Salud, Ámbito de procedencia, Hospital, Área, Provincia, Edad, Sexo y Triaje). El año 2021 cuenta con 80.510 registros, 2022 con 98.406 registros y 2023 con 24.177 registros. Se han eliminado las columnas de hospital, área y provincia ya que en los tres casos era igual, Valladolid. Se ha considerado información redundante.

En los registros meteorológicos se han observado nueve columnas tanto en 2021 como 2022 (date, tavg, tmin, tmax, prcp, snow, wdir, wspd, wpgt, pres y tsun). Estos registros contienen muchos “missings”. No hay registros en ninguna columna de wpgt y tsun por lo que se han eliminado dichas columnas. Las columnas wdir, wspd y pres presentan un solo valor faltante por lo que se decide imputar por la mediana de la columna.

El resto de registros de mortalidad, natalidad, habitantes, contaminantes y partidos no presentan datos faltantes, se han buscado “outliers” que al igual que en los otros registros se ha usado el principio estadístico del rango intercuartílico (IQR), donde los valores menores que $Q1 - 1.5 \cdot IQR$ y mayores que $Q3 + 1.5 \cdot IQR$ son considerados atípicos.

Una vez limpios se han unificado los datos de 2021 y 2022 de cada registro mediante la concatenación, establecido los índices de cada dataframe, normalmente la columna con la fecha, y se ha ordenado por dicho índice. Así se han obtenido y renombrado nuevos dataframes con los datos limpios y ordenados.

4.2- Modelo

Con los datos limpios se empieza a trabajar con el modelo ARIMA. Lo primero que se ha hecho evidente al realizar el análisis y limpieza, es que hay varias variables que no se pueden comparar temporalmente ya que su temporalidad es diferente, como la mortalidad y nacimientos, que no se han registrado en diario como lo hacen los registros de visitas a urgencias, sino

mensualmente, o los partidos que no tienen registros diarios continuos porque no hay partidos cada día.

La literatura consultada sobre el modelo, evidencia que funciona por series temporales simétricas y, sobre todo, no requiere de otras variables para su predicción, si no solo su propio comportamiento para hacer la predicción. En principio podría realizar la predicción sin más variables que las propias del registro de visitas, aun así, se crea el primer DataFrame (df) contando el número de atenciones por día y algunas otras características cuantitativas como los datos meteorológicos y los contaminantes ambientales.

En el primer intento de aplicar ARIMA ha dado error en el código y el siguiente gráfico:

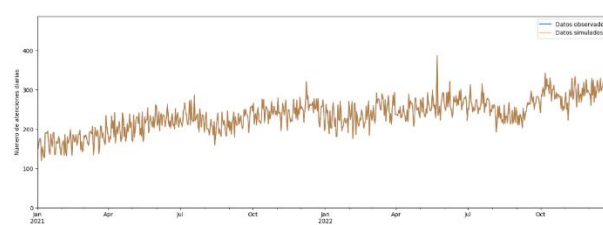


Imagen 4. Primer resultado aplicando ARIMA.

El modelo con la meteorología y climatología funciona para graficar, pero no predice nada.

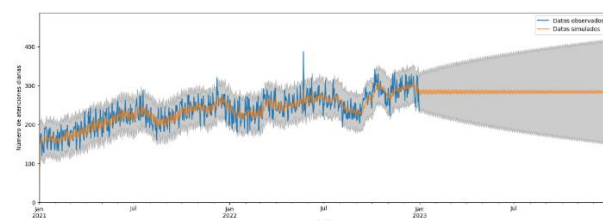


Imagen 5. Uso de ARIMA con variables meteorológicas y contaminantes.

La conclusión extraída es que es debido a que las variables deben estar sincronizadas con la serie temporal a predecir y para hacer predicciones futuras, las variables deben tener valores para el período de pronóstico. El

modelo funciona con su propio comportamiento, no requiere obligatoriamente de otras variables y tampoco se dispone de datos de las variables exógenas al df principal para 2023.

Por todo lo anterior, se ha optado finalmente por usar los datos de visitas de 2021 y 2022, ambos incluyen la fecha de atención, el día de la semana, la hora, el nivel de triaje, la zona básica de salud, el ámbito de procedencia, la edad y el sexo.

Para usar ARIMA se deben establecer los tres componentes principales del modelo: p , d y q .

p : Orden del componente Autoregresivo (AR). Es el número de lags (cada lag representa un desplazamiento temporal hacia atrás en los datos) pasados de la serie temporal que se utilizan para predecir el valor actual. Si $p=1$, significa que el modelo usa el valor inmediatamente anterior para hacer la predicción.

d : Orden de diferenciación (I). Este valor indica el número de veces que la serie temporal debe diferenciarse para hacerla estacionaria. Se utiliza para eliminar tendencias en la serie temporal. Si $d=1$, se diferencia una vez, lo que significa que se toman las diferencias de primer orden entre los valores consecutivos de la serie temporal.

q : Orden del componente de Media Móvil (MA). Determina cuántos errores de predicciones pasadas se utilizan para predecir el valor actual, si $q=1$, el modelo utiliza el error inmediatamente anterior para hacer la predicción.

Para determinar los valores de p y q se generan gráficos de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para la serie diferenciada. La autocorrelación en y la autocorrelación parcial sugirieron que un modelo con $p=1$ y $q=1$ sería óptimo.

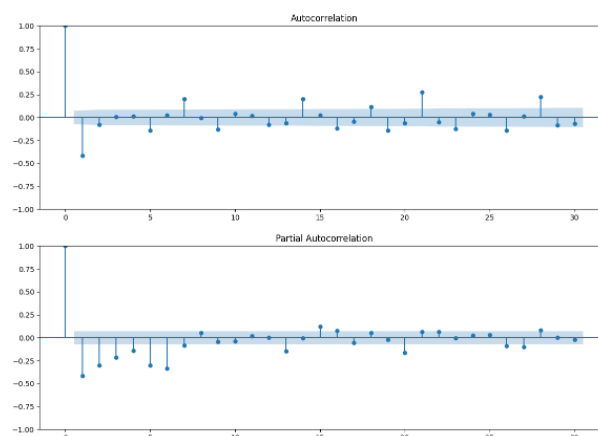


Imagen 5. Resultados de las funciones de ACF y PACF.

La d se determina a partir de la prueba de raíz unitaria, Dickey-Fuller Aumentada (ADF), como la serie temporal original no es estacionaria y una diferenciación la hace estacionaria, entonces $d=1$.

```
ADF Statistic: -1.7692750373649806
p-value: 0.39581890558384314
ADF Statistic (differenced): -8.287106169899078
p-value (differenced): 4.3607381404837443e-13
```

Imagen 6. Resultados de ADF antes y después de la diferenciación

Con estos resultados se establece que el modelo ARIMA será con un orden 1, 1, 1.

Una vez definidos los parámetros, se dividen los datos en conjuntos de entrenamiento y prueba, utilizando el 80% de los datos para el entrenamiento y el 20% restante para la prueba.

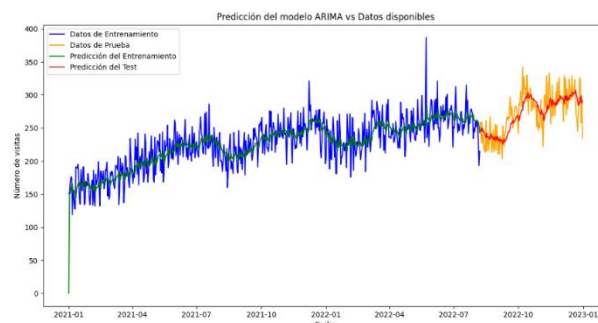


Imagen 7. Resultado de la predicción del modelo ARIMA.

Se ha evaluado calculando el error cuadrático medio (MSE), la raíz del error cuadrático medio (RMSE), y el coeficiente de determinación (R^2) para el conjunto de prueba. Estas métricas permiten evaluar la precisión y el ajuste del modelo a los datos de prueba.

```
R2 Total: 0.6574698557529273
RMSE Total: 24.252326968260814
MSE Total: 588.1753633754307
```

Imagen 8. Resultado de las métricas de evaluación usadas para el modelo ARIMA.

El modelo parece captar la tendencia general en los datos, pero no maneja del todo bien las fluctuaciones más abruptas o los picos. No hay señales claras de overfitting, ya que la predicción en el conjunto de prueba sigue de manera razonable los datos reales. Este resultado es un punto de partida adecuado, capturando de manera razonable la tendencia general de los datos. Sin embargo, el R^2 , junto con los errores de predicción (RMSE y MSE), hace pensar que el modelo podría beneficiarse de ajustes adicionales o del uso de un modelo más complejo.

Dada la importancia de la temporalidad de los datos, la estacionalidad de los mismos y los resultados obtenidos con ARIMA, se ha decidido probar la implementación de un modelo SARIMA. Un modelo ARIMA estándar no puede capturar estos efectos estacionales con la misma precisión, lo que puede llevar a predicciones menos precisas. En cambio, el modelo SARIMA incluye componentes específicamente diseñados para manejar la estacionalidad. Con el uso de SARIMA se obtendría un modelo que considera tanto las relaciones a corto plazo (a través de los términos ARIMA) como las relaciones a largo plazo estacionales.

Empezando con SARIMA, los datos de visitas se han agrupado por fecha para contar el número de visitas diarias. La columna de fecha ha convertido al formato datetime para facilitar el manejo temporal de los datos.

Al igual que para ARIMA, se han determinado los valores de p , d y q , siguiendo el mismo método que antes.

Se ha obtenido que los valores más adecuados son el orden 1, 1, 1.

Para el modelo SARIMA, no sólo necesitamos definir p , q y d , también se deben definir P , D , Q , S .

Para P , D , Q y S , se define 1, 0, 1, 7.

$P=1$ por que el modelo incluirá un término estacional de autoregresión (SAR) de primer orden. $D=0$ por que no se necesita diferenciación estacional adicional (por que ya se ha diferenciado). $Q=1$ por que el modelo incluirá un término estacional de media móvil (SMA) de primer orden, y finalmente $S=7$ por que se ha establecido una periodicidad semanal en los datos.

Una vez definidos los parámetros, se dividen los datos en conjuntos de entrenamiento y prueba, utilizando el 80% de los datos para el entrenamiento y el 20% restante para la prueba.

En el modelo SARIMA también se ha evaluado calculando el error cuadrático medio (MSE), la raíz del error cuadrático medio (RMSE), y el coeficiente de determinación (R^2) para el conjunto de prueba.

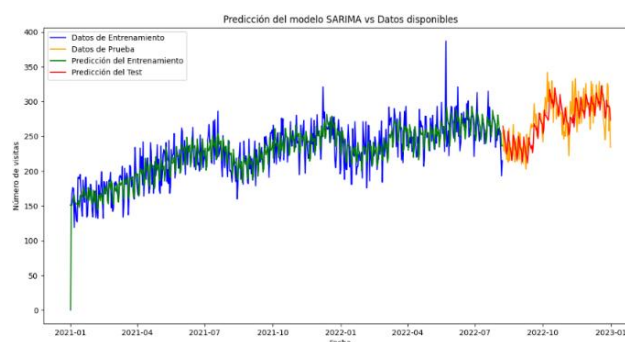


Imagen 8. Resultado de la predicción del modelo SARIMA.

```
R2 Total: 0.7313285533515717
RMSE Total: 21.479035356284637
MSE Total: 461.3489598365255
```

Imagen 9. Resultado de las métricas de evaluación usadas para el modelo.

El modelo SARIMA ha mejorado todas las métricas usadas para su evaluación. Se ha obtenido un R^2 total de 0.7312 y un RMSE total de 21.08, indicando un buen ajuste y capacidad predictiva.

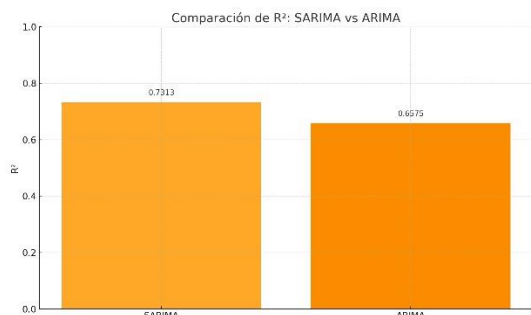


Imagen 10. Diferencia de resultados de R^2 entre modelos.

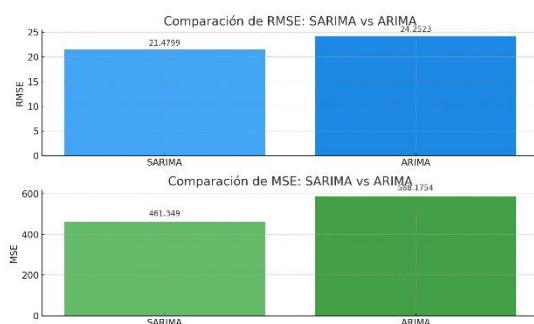


Imagen 11. Diferencia en los resultados de RMSE y MSE entre modelos.

El análisis de residuos y las verificaciones diagnósticas confirmaron la validez del modelo. También se realizan los gráficos de ACF y PACF para confirmar que el modelo funciona bien. No muestran autocorrelaciones significativas, lo que sugiere que los residuos se comportan como un ruido blanco. Esto es indicio de que el modelo SARIMA ajustado es adecuado y ha capturado bien la estructura temporal de los datos.

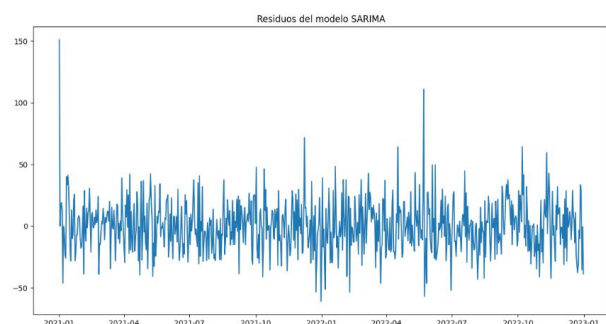


Imagen 12. Gráfico de los residuos del modelo.

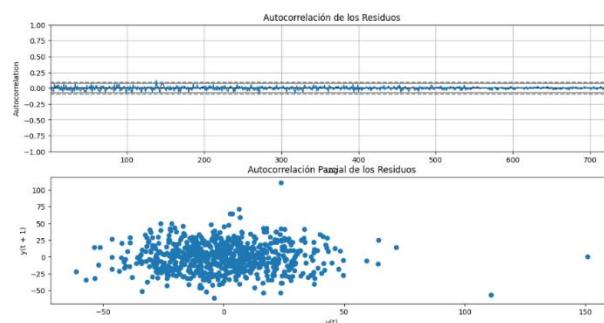


Imagen 13. Gráficos de ACF y PACF.

4.3 Predicción a una semana vista

Con los resultados del ajuste del modelo, se decide realizar un pronóstico más específico, en concreto del número de visitas para la primera semana de enero de 2023 con SARIMA.

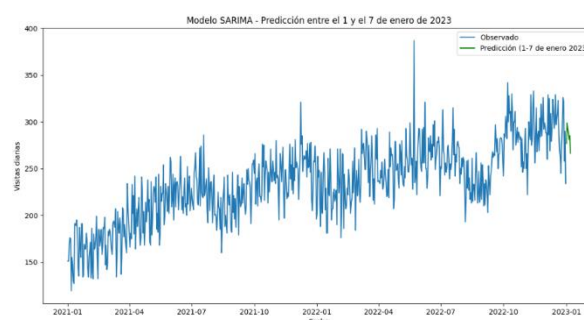


Imagen 14. Gráfico del resultado de la predicción semanal.

```
Predicción entre el 1 y el 7 de enero de 2023:
2023-01-01    276.753773
2023-01-02    298.796030
2023-01-03    292.309665
2023-01-04    288.045237
2023-01-05    281.205784
2023-01-06    285.126463
2023-01-07    266.320137
Freq: D, Name: predicted_mean, dtype: float64
```

Imagen 15. Número de visitas predichas por día.

5- Resultados

Con los resultados obtenidos, se observa que el modelo SARIMA tiene un buen ajuste y buena capacidad predictiva.

No obstante, cuando se ha querido hacer la predicción semanal y se ha comparado con los datos reales de 2023, la diferencia es importante. El modelo ha predicho

1.968,52 visitas durante la primera semana de enero de 2023, en cambio, con los datos reales se observa que hubo 449 visitas en ese período.

6- Conclusiones y líneas abiertas

El cambio de un modelo ARIMA a un modelo SARIMA para poner el foco en la estacionalidad, ha significado dar mejores resultados. Se han obtenido buenos resultados de RMSE y R2 y muestra un gráfico que hace pensar que el modelo es bastante efectivo para hacer predicciones futuras. No obstante, si comparamos la predicción realizada para la primera semana de enero con los datos reales de esas fechas hay una diferencia muy grande en el número de visitas.

Por ello, considero que no es problema del modelo. Pese a que sin duda es mejorable, creo que el problema está en los datos. En el momento de plantear el proyecto, ya se pensó en el posible efecto que podría tener el hecho de ser fechas que coinciden con la pandemia y postpandemia de COVID-19. Es lógico pensar que pueden estar sesgados y/o condicionados por la pandemia. La pandemia parece ser una explicación de que en el año 2021 haya 80.510 visitas, en 2022 98.406 visitas, pero en 2023 se reducen hasta 24.177 visitas.

Se buscaron los registros de visitas a urgencias del Hospital Clínico Universitario de Valladolid con fechas anteriores a 2020, pero no constan datos abiertos con dichos registros, de hecho, empiezan en 2021. Es probable que se crearan los registros a partir de 2020 por necesidad y por el nuevo paradigma que trajo la pandemia, la importancia que tuvieron los datos y su registro y manejo.

SARIMA ha probado ser un buen modelo para realizar este tipo de predicciones. Con datos de calidad y no sesgados seguramente el resultado sería más preciso.

Por otro lado, y viendo que no ha sido posible obtener datos anteriores a 2020, se podría buscar otras maneras

de mejorar la predicción y/o comparar los resultados del modelo que ha dado el mejor resultado. Se podría, por ejemplo, utilizar redes neuronales, que no requieren tanta estacionalidad, y poder así adjuntar todos los datos recopilados en un inicio (demográficos, climáticos, contaminantes, etc.) como variables y observar realmente el peso que tienen en la predicción, más allá de las medidas en el modelo como son las propias de la visita (fecha, día, hora, zona básica de Salud, ámbito de procedencia, edad, sexo y nivel de triaje).

Sin duda, ya sea usando ARIMA, SARIMA o redes neuronales, obtener datos que no tengan un sesgo tan significativo es clave para poder entrenar un modelo que realice las predicciones más precisas.

7- Agradecimientos

En primer lugar, a mi pareja, por su apoyo, comprensión y ánimos.

A mi tutor, Oriol, y a los profesores del máster por sus enseñanzas y consejos.

8- Bibliografía

- [1] Ministerio de Sanidad de España unidad de Urgencias Hospitalarias Estándares Recomendaciones 2010, 145. Disponible online: 11 CalidadSNS/docs/UUH.pdf. <https://sanidad.gob.es/organización/sns/plan> Consultado en Dic 2022.
- [2] Aiken, L. H., Sloane, D. M., Bruyneel, L., Van den Heede, K., Griffiths, P., Busse, R., Diomidous, M., Kinnunen, J., Kózka, M., Lesaffre, E., McHugh, M. D., Moreno-Casbas, M. T., Rafferty, A. M., Schwendimann, R., Scott, P. A., Tishelman, C., van Achterberg, T., & Sermeus, W.; RN4CAST consortium. (2014). Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study.

Lancet, 383(9931), 1824-1830.
[https://doi.org/10.1016/S0140-6736\(13\)62631-8](https://doi.org/10.1016/S0140-6736(13)62631-8)

[3] Cusidó, J., Comalrena, J., Alavi, H., & Llunas, L. (2022). Predicting Hospital Admissions to reduce Crowding in the Emergency Departments. *Applied Sciences*, 12**(21),10764.
<https://doi.org/10.3390/app122110764>

[4] Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *Lancet*. 2020 May 16;395(10236):1579-1586. doi: 10.1016/S0140-6736(20)30226-9. PMID: 32416782; PMCID: PMC7255280.

[5] Sahu M., Gupta R., Ambasta R.K., Kumar P. Artificial intelligence and machine learning in precision medicine: A paradigm shift in big data analysis. *Prog. Mol. Biol. Transl. Sci.* 2022; 190:57–100.

[6] Furian, N., O'Sullivan, M., Walker, C., & Reuter-Oppermann, M. (2022). Machine Learning-Based Patient Selection in an Emergency Department. *arXiv preprint arXiv:2206.03752*.

[7] Hu, Y., Cato, K. D., Chan, C. W., Dong, J., Gavin, N., Rossetti, S. C., & Chang, B. P. (2023). Use of Real-Time Information to Predict Future Arrivals in the Emergency Department. *Annals of Emergency Medicine*, 81**(6), 728-737.
<https://doi.org/10.1016/j.annemergmed.2022.11.005>

[8] Sharafat, A. R., & Bayati, M. (2021). PatientFlowNet: A Deep Learning Approach to Patient Flow Prediction in Emergency Departments. *IEEE Access*, 9**, 45552-45561. <https://doi.org/10.1109/ACCESS.2021.3066164>

[9] Torné Vilagrasa E, Guarga Rojas A, Torras Boatella MG, Pozuelo García A, Paserin Rua M, Borrell Thió C. Análisis de la demanda en los servicios de urgencias de Barcelona. *Elsevier*. 2003. 32: 423-424.

[10] Hospital Clínico Universitario de Valladolid. (n.d.). *Área de influencia*. Portal de Salud de la Junta de Castilla y León. Retrieved July 25, 2024, from <https://www.saludcastillayleon.es/HCUValladolid/es/area-influencia>

[11] Ballesteros Peña, Sendora; Fernández Aedo, Irrintzi; Vallejo de la Hoz, Gorka. Impacto de los partidos de fútbol en la frecuentación de pacientes a un servicio de urgencias. *Emergencias (Sant Vicenç dels Horts)*; 35(5):345-348, oct. 2020.

[12] Anna Ponjoan, et al. Effects of extreme temperatures on cardiovascular emergency hospitalizations in a Mediterranean región: a self-controlled case series study. *Environmental Health*, vol 16, pp 1-32, abril 2017.

[13] Sánchez M, Smally A J. Comportamiento de un servicio de urgencias según el día de la semana y el número de visitas. *emergencias* 2007; 19:319-322.

[14] Idescat. Anuari estadístic de Catalunya. Contaminació atmosfèrica. Contaminants principals. [Internet]. www.idescat.cat. [cited 2022 Dic 12]. Available from: <http://idescat.cat/indicadors/?id=aec&n=15215>.

[15] OMS, Organización Mundial de la Salud. Organización Mundial de la Salud (OMS). [En línea] <http://www.who.int/mediacentre/factsheets/fs313/es/>.

[16] Rouco Alfonso, D. (2020). [Monitorización, análisis y predicción de consumos de calefacción y ACS en edificios]. En J. Díaz Trapiella (Ed.), *Máster Universitario en Ingeniería Industrial* (p. 132). Editorial David Rouco Alfonso. Recuperado de <https://digibuo.uniovi.es/dspace/handle/10651/56297>

[17] de Dios González, Javier; Ramos Terrades, Oriol, dir. Modelos de predicción de series temporales: gripe y COVID-19. 2022. (Enginyeria Informàtica) <<https://ddd.uab.cat/record/264129>> [Consulta: 05 junio 2024].