

# Informe de modelo de predicción de Muerte Hospitalaria en función del CMBD - Mejora

Esto es una continuación de la presentación del 17 de Junio donde se han aplicado unas mejoras del modelo hablado y se han obtenido unos mejores resultados.

## Explicación de las variables del modelo

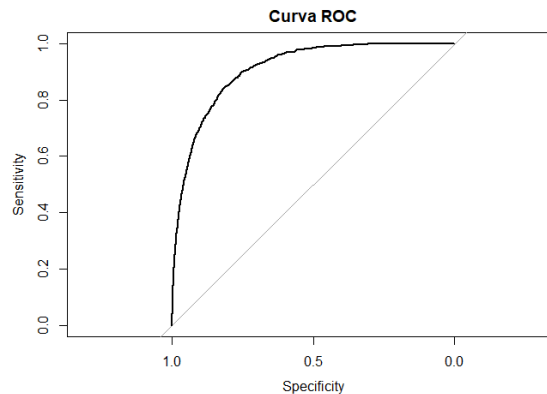
- Se han tomado para cada paciente los 20 diagnósticos y se asocian a sus respectivas secciones del CIE10. Además se indica usando el POA si ese diagnóstico se encontraba en el momento de la admisión o no. Es decir, si un paciente tiene la sección 17 y el POA asociado a ese diagnóstico es “Sí”, ponemos un 1 en la variable dummy “S17”.
- Se ha usado  $edad^2$  y se ha añadido  $tiempo\_estancia\_dias^2$  como variable por ver una relación exponencial desplazada entre ésta y *Fallecimiento*.
- Se ha usado también la variable *SEXO* por haber visto una clara diferencia en las distribuciones de *Fallecimiento* separadas por sexo.
- Por último destacar que se han eliminado todos los datos de pacientes de menos de 20 años por el escaso número de fallecidos en esas edades. También se han eliminado todos los contactos con el sistema por temas de embarazo.
- No se ha hecho la interacción de edad con los diagnósticos por no encontrar una mejora en el modelo.
- Se han eliminado las secciones del CIE10 relacionadas factores que influyen en el estado de salud y contacto con los servicios sanitarios (Z00-Z99).

Una vez hecho este modelo procedemos a los tests.

## Tests realizados al 30% de los datos

### Área bajo la curva ROC

Obtenemos un *AUC* de 0.909 (Significativamente más alto que los modelos previos)



Aun así tenemos que tener en cuenta que como nuestro modelo está sesgado a los no fallecidos, este área bajo la curva puede estar incentivada por una predisposición al modelo de predecir muchos no fallecidos.

### Matriz de confusión<sup>1</sup>, accuracy, sensibilidad y especificidad<sup>2</sup>

Lo primero que se tiene que hacer es marcar un valor de la probabilidad de predicción donde podamos clasificar a unos como que mueren y a otros como que no mueren. Para encontrar este umbral, usamos el **criterio de Youden** que busca maximizar la diferencia entre la tasa de verdaderos positivos y la de falsos positivos. El umbral obtenido es: **0.051**, osea si nuestro modelo predice que muere con probabilidad mayor del **5.1%** entonces lo etiquetamos como “Fallecido”.

Con este umbral podemos construir la *matriz de confusión*:

Predicho / Real	No muere	Muere
No Muere	18014 (81.6%)	185 (15.4%)
Muere	4066 (18.4%)	998 (84.6%)

---

<sup>1</sup> [https://es.wikipedia.org/wiki/Matriz\\_de\\_confusi%C3%B3n](https://es.wikipedia.org/wiki/Matriz_de_confusi%C3%B3n)

<sup>2</sup> [https://es.wikipedia.org/wiki/Sensibilidad\\_y\\_especificidad](https://es.wikipedia.org/wiki/Sensibilidad_y_especificidad)

Por lo tanto los valores obtenidos de Accuracy, Sensibilidad (Recall) y Especificidad son:

Accuracy	0.817
Sensibilidad (Recall)	0.846
Especificidad	0.816

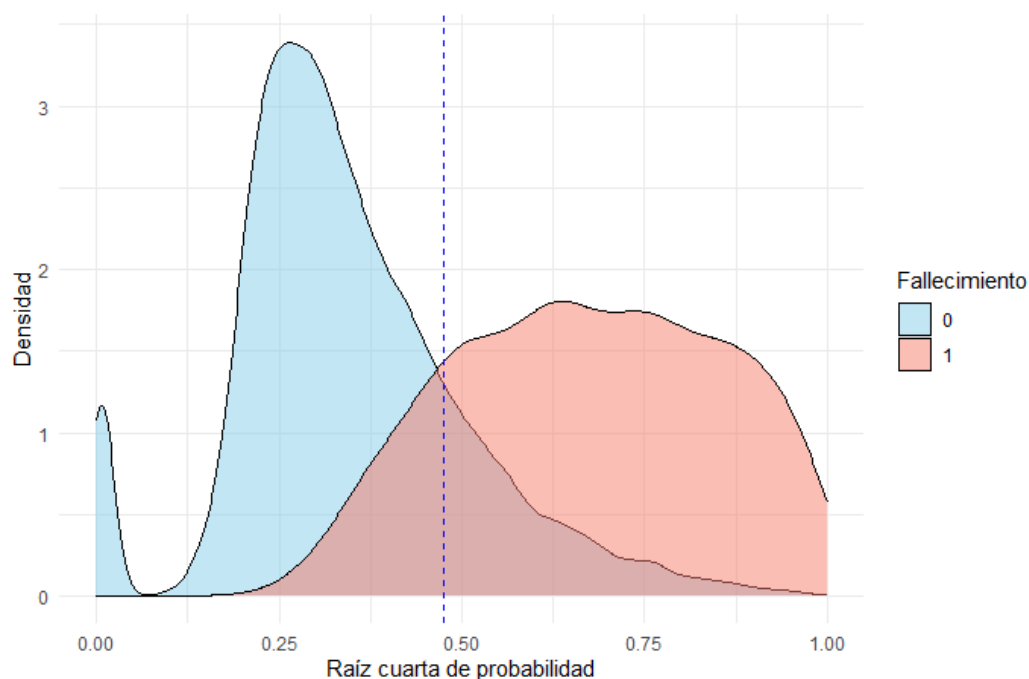
*Accuracy* mide todos los que están bien etiquetados. Un **82%** es un valor

*Recall* mide los positivos que han sido correctamente etiquetados. Un **85%** es un buen valor, de cada 100 fallecidos el modelo predice correctamente a 85.

*Especificidad* mide los negativos que han sido correctamente etiquetados como negativos. Un valor de **82%** es un valor decente para lo que estamos trabajando.

### Diagrama de densidad, área de solapamiento (OVL)<sup>3</sup>

El *diagrama de densidad* representa los valores estimados de probabilidad de muerte separados por aquellos que realmente fallecen y los que no. La línea azul representa dónde está el umbral de Youden. Todo el diagrama está escalado a la raíz cuarta para verse mejor:



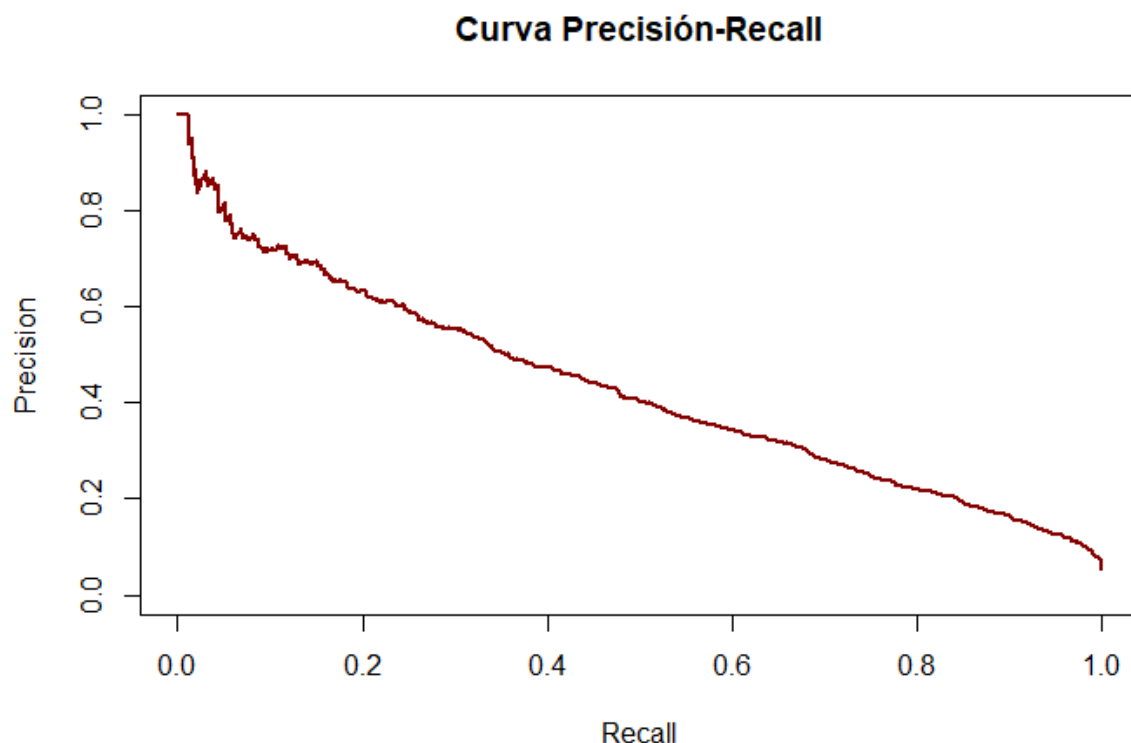
<sup>3</sup> [https://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S2448-91902021000200128](https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2448-91902021000200128)

A primera vista ya se puede observar como hay una separación significativa de las dos densidades, esto nos incita a pensar que efectivamente el modelo es bueno separando los casos clínicos.

El *área de solapamiento (OVL)* es un valor entre 0 y 1 que representa la proporción de la población para la cual el modelo no logra separar correctamente las clases, es decir, donde las distribuciones de ambas clases se superponen. Se ha obtenido un valor de **0.28**, este se considera bajo y por lo tanto un buen modelo discriminativo.

### Curva Precisión-Recall y AUC-PR<sup>4</sup>

La *curva Precision-Recall (PR)* muestra la relación entre la *precisión* y el *recall* del modelo al variar el umbral de decisión. Es especialmente útil en problemas con clases desbalanceadas, ya que se centra en el rendimiento sobre la clase positiva. Una curva que se mantenga alta indica un buen equilibrio entre aciertos y errores en las predicciones positivas.



---

<sup>4</sup> <https://glassboxmedicine.com/2019/03/02/measuring-performance-auprc/>

La *curva Precision-Recall* obtenida muestra una caída progresiva de la precisión a medida que aumenta el recall, lo cual es esperable en modelos que deben equilibrar ambos criterios. Se observa que, aunque el modelo alcanza altos valores de precisión en niveles bajos de recall, su *rendimiento se degrada conforme intenta capturar más positivos*.

El *área bajo la curva PR (AUC-PR)* es de **0.43**, lo que indica un rendimiento moderado, superior al azar (que estaría en torno al **0.04**). Este valor sugiere que el modelo logra detectar una parte razonable de los positivos con una precisión aceptable, aunque comete errores cuando trata de abarcar más casos.

Ahora vamos con un par de medidas de **calidad probabilística**<sup>5</sup>.

### Log-Loss<sup>6</sup>

El *log-loss* (o *cross-entropy loss*) mide la calidad de las probabilidades predichas por un modelo de clasificación binaria. Evalúa no solo si una predicción es correcta, sino también *cuán confiado estaba el modelo al hacerla*. Penaliza fuertemente las predicciones erróneas con alta confianza, es decir, si el modelo asigna una probabilidad cercana a 1 a un resultado que finalmente es 0. Un log-loss bajo indica que el modelo no solo clasifica bien, sino que también *asigna probabilidades bien calibradas y razonables*.

A pesar del fuerte desbalance de clases (~4% de fallecimientos), el modelo presenta un log-loss de **0.1345**, indicando una muy buena calibración probabilística. Esto significa que las probabilidades asignadas por el modelo son razonablemente fiables en promedio, sin mostrar sobreconfianza en predicciones erróneas.

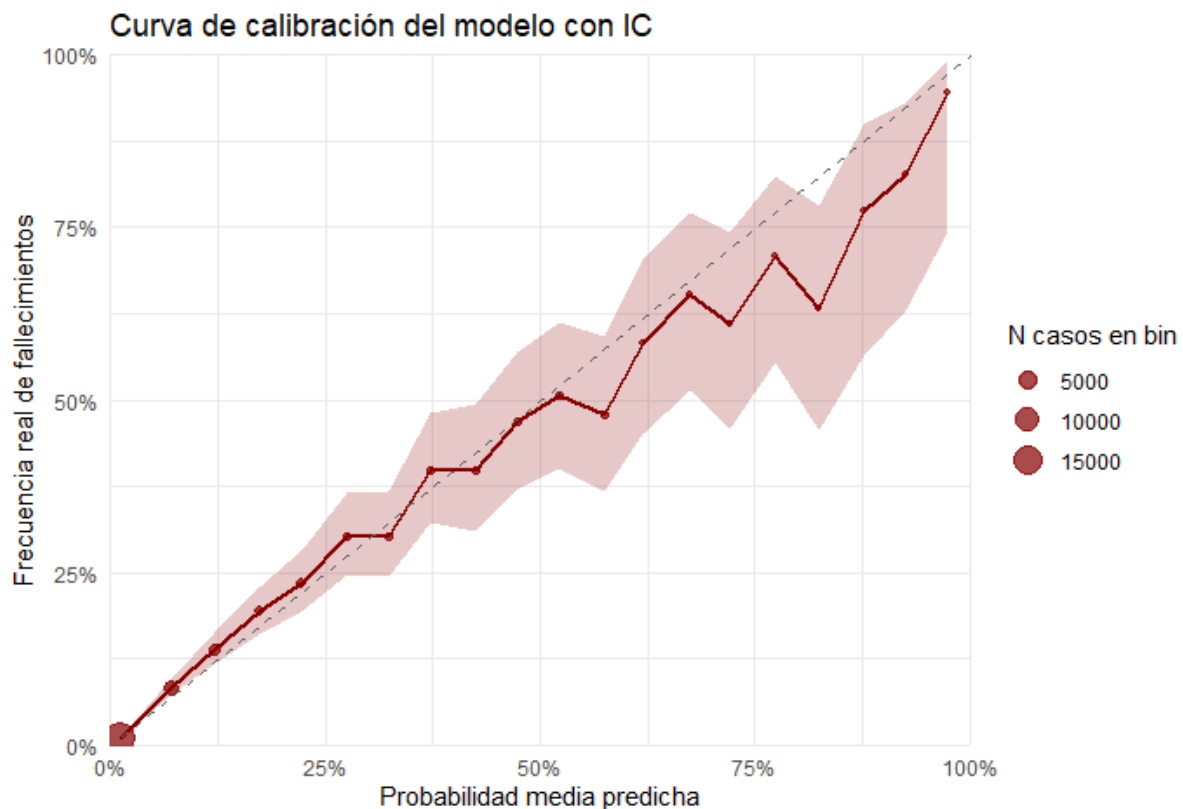
---

<sup>5</sup> La calidad probabilística de un modelo se refiere a qué tan bien las probabilidades predichas por dicho modelo representan la verdadera incertidumbre del evento que se intenta predecir.

<sup>6</sup> <https://www.kaggle.com/code/dansbecker/what-is-log-loss>

## Calibration curve<sup>7</sup>

La *curva de calibración* muestra si las probabilidades predichas están bien ajustadas a la realidad: un modelo *bien calibrado* es aquel cuyas probabilidades predichas reflejan correctamente las frecuencias observadas en los datos. Por ejemplo, para todos los casos a los que el modelo asigna una probabilidad cercana al 0.2, aproximadamente el 20% deberían haber experimentado el evento:



Se puede ver como la tendencia en valores bajos es buena. Para valores bajos la probabilidad del modelo se ajusta a la probabilidad real. En general se observa una tendencia global estable: el modelo predice mayores riesgos para pacientes con mayor probabilidad de fallecimiento, lo que indica una buena calibración general. Sin embargo, hay cierta variabilidad en los tramos medios-altos, con intervalos de confianza más amplios (en gran parte por menor cantidad de datos en esos rangos).

<sup>7</sup> <https://www.geeksforgeeks.org/machine-learning/calibration-curves/>

## Conclusiones

Este modelo final ha resultado explicar bastante bien la probabilidad de muerte hospitalaria en función de las secciones en los diagnósticos del CIE10, la edad, el sexo y el tiempo de hospitalización. Además el resultado más claro es el diagrama de densidad donde se ve una separación clara y un *área de solapamiento* de solo **0.28**. Con unos valores de *accuracy*, *recall* y *especificidad* bastante altos para un estudio en el que la probabilidad real de muerte hospitalaria es a priori dependiente de innumerables factores. Se puede dar (y se da) el caso en el que dos personas con los mismos diagnósticos uno fallece y el otro no. Luego con esta restricción intrínseca del problema se da más valor a los resultados obtenidos en este estudio.

Ahora sería probar el modelo con más cantidad de datos y de diferentes años. Se propone tomar los datos clínicos de Aragón de 2020 a 2024 y hacer un estudio similar para poder tener unos resultados más sólidos.

## Mejoras posibles al modelo

Para cerrar con el informe se dejan por escrito algunas mejoras que se le podrían hacer. Estas no garantizan un mejor modelo, pero en teoría debería darse.

- Hacer unas subsecciones en las secciones del CIE10 que se consideren que abarcan demasiadas enfermedades de diferente gravedad. Intentando no hacer demasiadas subsecciones para no caer en sobreajuste del modelo.
- Hacer unas secciones de los procedimientos que se hacen a los pacientes, estas pueden ser en función del criterio de un especialista que identifique los grupos posibles que se podrían hacer.