

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

- Tomamos para cada paciente los 20 diagnósticos y los asociamos a sus respectivas secciones del CIE10. Además indicamos 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 hecho un estudio usando el AIC¹ para valorar qué monomio de edad es mejor. Los resultados fueron:

n	AIC
---	-----

1	16861.92
---	----------

2	16830.71	—> En particular usaremos edad^2 .
----------	-----------------	---

3	16832.77
---	----------

4	16857.53
---	----------

5	16896.22
---	----------

- Hemos añadido $\text{tiempo_estancia_dias}^2$ como variable por ver una relación exponencial desplazada entre esta y *Fallecimiento*.

- Usamos también la variable *SEXO* por haber visto una clara diferencia en las distribuciones de *Fallecimiento* separadas por sexo.

- Por último destacar que hemos eliminado todos los datos de pacientes de menos de 20 años por el escaso número de fallecidos en esas edades.

¹ https://es.wikipedia.org/wiki/Criterio_de_informaci%C3%B3n_de_Akaike

- Una vez hecho el modelo con todas estas variables nos quedamos con las más significativas ($p\text{-valor} > 0.05$) y obtenemos las siguientes:

("S271", "S218", "S200", "S120", "S114", "S102", "S198", "S75", "S159", "S143", ...)

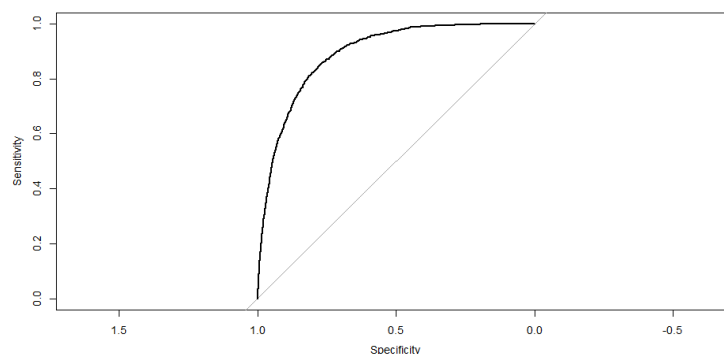
- Ahora vamos a hacer la interacción de todas las variables (incluido *SEXO*) con $edad^2$.

Una vez hecho este modelo procedemos a los tests.

Tests realizados al 30% de los datos (Se deja un enlace a cada término para entender qué es cada cosa)

Área bajo la curva y ROC

0.8901 (Más alto que todos los modelos previos)



Matriz de confusión², accuracy, sensibilidad y especificidad³

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: **0.0484763**, osea si nuestro modelo predice que muere con probabilidad mayor del 4.85% entonces lo etiquetamos como “Fallecido”.

² https://es.wikipedia.org/wiki/Matriz_de_confusi%C3%B3n

³ https://es.wikipedia.org/wiki/Sensibilidad_y_especificidad

Con este umbral podemos construir la matriz de confusión:

Predicho / Real	No muere	Muere
No Muere	18758	185
Muere	5217	998

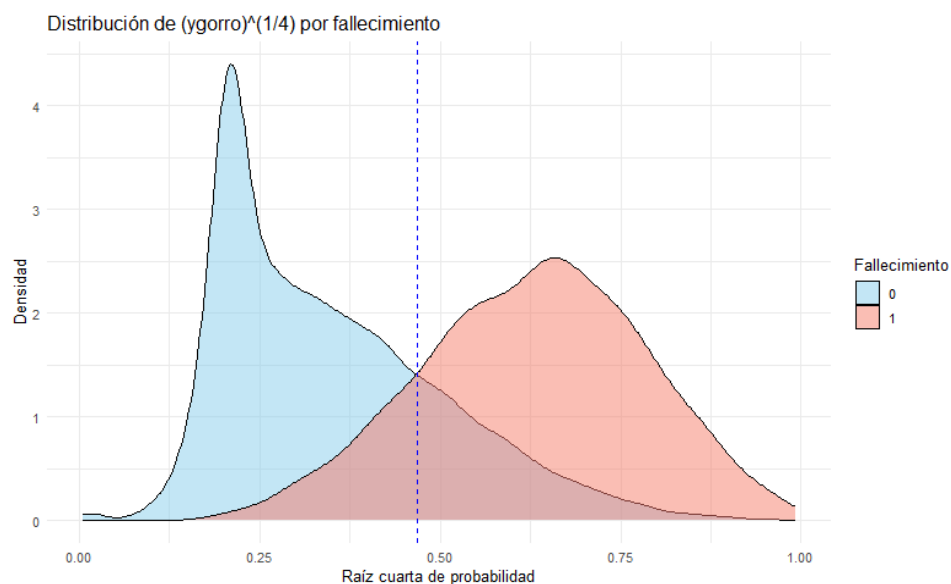
Los valores obtenidos de Accuracy, Sensibilidad (Recall) y Especificidad son:

Accuracy	0.785
Sensibilidad (Recall)	0.844
Especificidad	0.782

Nuestro modelo es bueno etiquetando a aquellos pacientes que acaban muriendo.

Diagrama de densidad, área de solapamiento (OVL)⁴ e índice de Jaccard⁵

La línea azul representa dónde está el umbral de Youden. Todo el diagrama está escalado a la raíz cuarta para verse mejor.

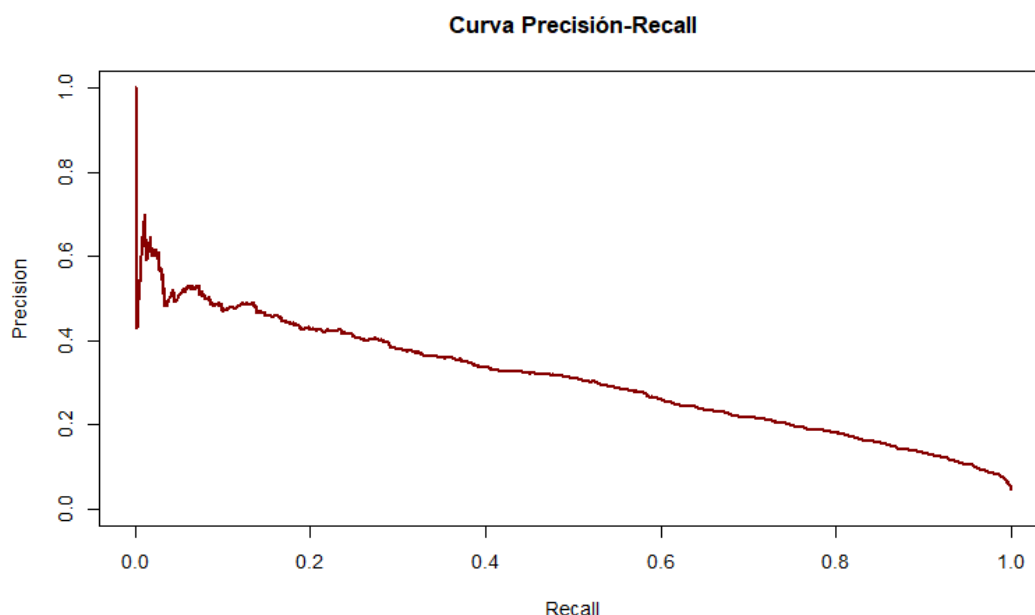


El área de solapamiento obtenido es de **0.335** y el Índice de Jaccard es de **0.222**.

⁴ https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2448-91902021000200128

⁵ https://es.wikipedia.org/wiki/%C3%8Dndice_de_Jaccard

Curva Precisión-Recall y AUC-PR⁶



No es una curva especialmente buena, aun así el **área bajo la curva PR es de 0.31**. En comparación un modelo aleatorio tendría un **AUC-PR de 0.047** puesto que los fallecimientos representan solo el 4.7% de los datos. La diferencia entre el AUC-ROC y el AUC-PR es que este segundo penaliza los falsos positivos mucho más, y muestra que el modelo aún comete muchos errores al predecir el fallecimiento.

Ahora vamos con un par de medidas de **calidad probabilística**⁷.

Log-Loss⁸

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 (o viceversa). Un log-loss bajo indica que el modelo no solo clasifica bien, sino que también **asigna probabilidades bien calibradas y razonables**.

⁶ <https://glassboxmedicine.com/2019/03/02/measuring-performance-auprc/>

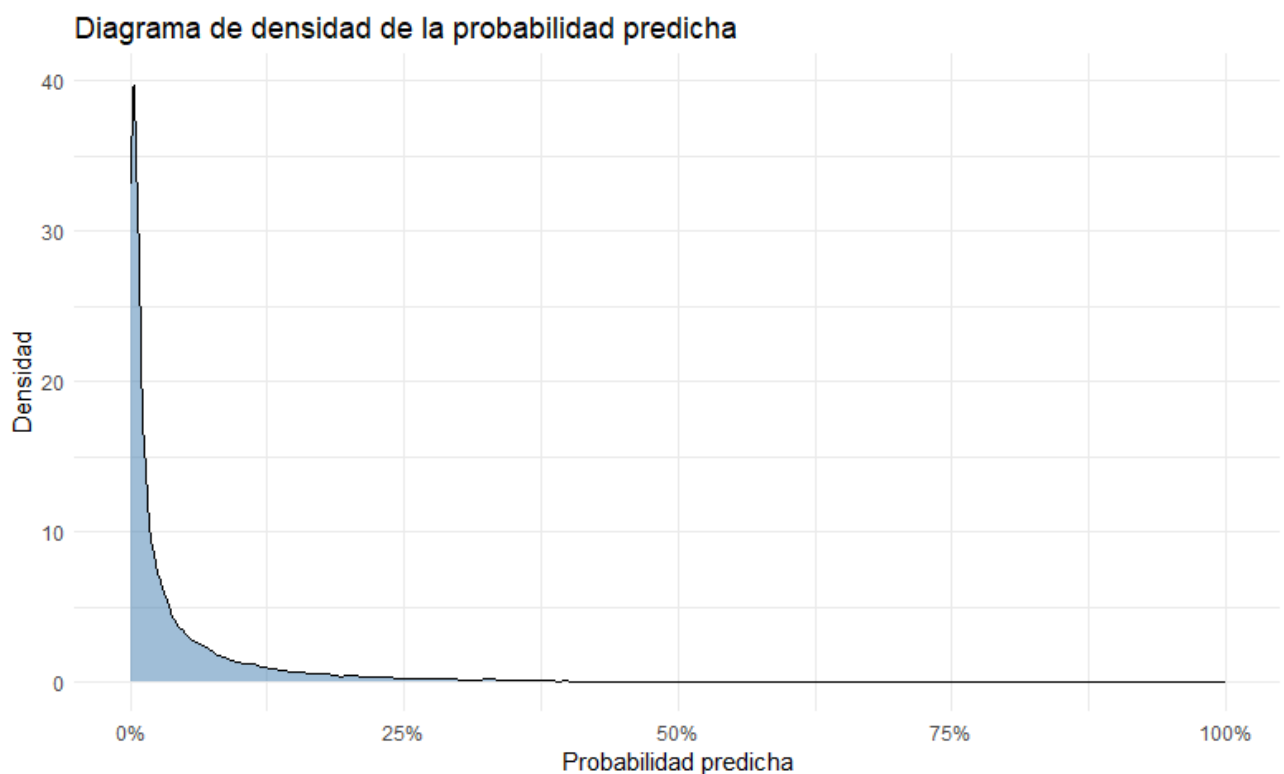
⁷ La **calidad probabilística** de un modelo se refiere a qué tan bien las probabilidades predichas por dicho modelo representan la verdadera incertidumbre del evento que se intenta predecir.

⁸ <https://www.kaggle.com/code/dansbecker/what-is-log-loss>

A pesar del fuerte desbalance de clases (~5% de fallecimientos), el modelo presenta un log-loss de 0.1354, 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.

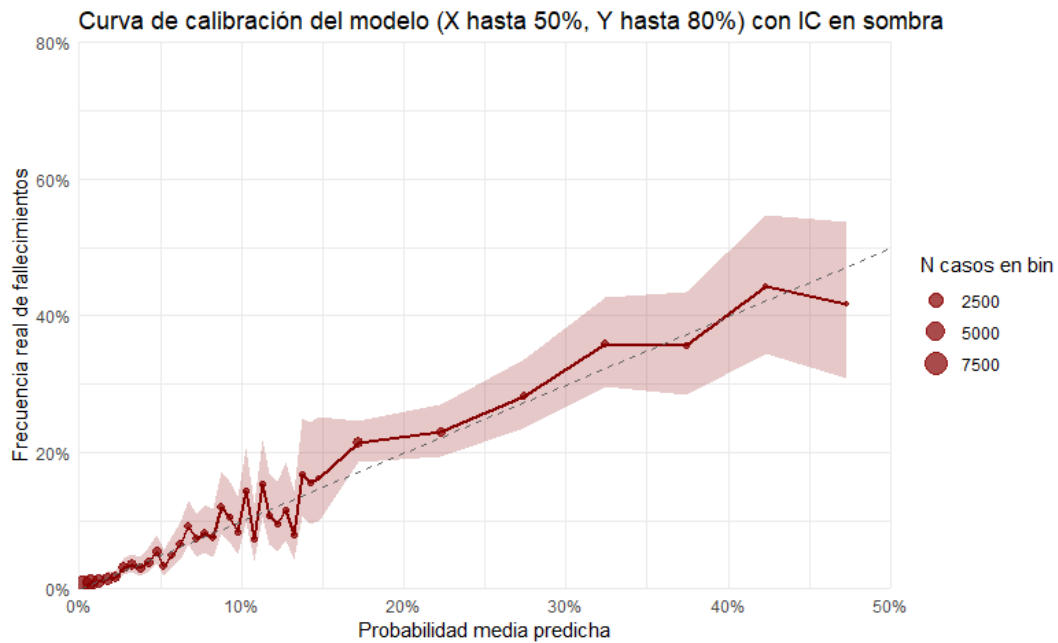
Calibration curve⁹

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. Teniendo en cuenta que nuestra densidad de valores predichos se ve así:

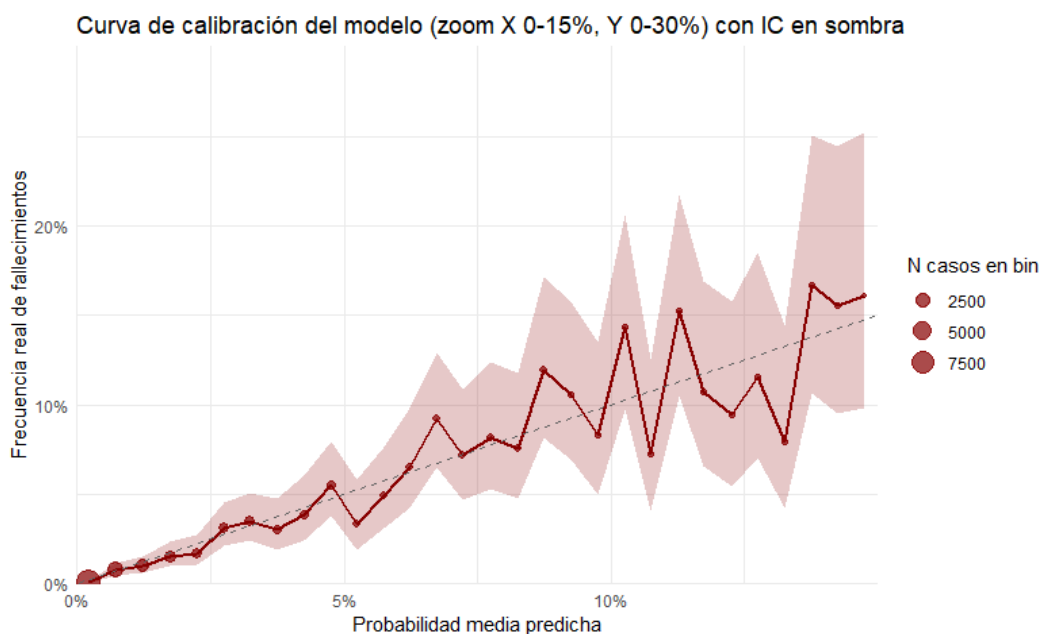


Vamos a hacer los grupos solo hasta el 50% ya que ahí es donde tenemos la mayoría de nuestros datos:

⁹ <https://www.geeksforgeeks.org/machine-learning/calibration-curves/>



La curva de calibración obtenida muestra un buen ajuste a la línea de referencia ideal (recta discontinua gris), lo que indica que las probabilidades predichas por el modelo se corresponden adecuadamente con las frecuencias reales observadas. Esto sugiere que el modelo está bien calibrado y sus predicciones probabilísticas son fiables para reflejar la verdadera probabilidad de fallecimiento. Como la media de predicciones es de 0.04 vamos a acercarnos al intervalo 0%-15% para verlo más de cerca:



La línea de referencia ideal (línea punteada) se encuentra siempre dentro del intervalo de confianza de las estimaciones observadas, lo que indica que no hay

evidencia estadística significativa de descalibración en el modelo. Esto refuerza la conclusión de que las probabilidades predichas son coherentes con los resultados reales dentro de la variabilidad esperada, mostrando una calibración sólida.

Conclusiones

Habría que probar a hacer este estudio con un mayor conjunto de datos, por ejemplo coger los últimos 5 años de datos médicos de Aragón y ver qué tal funciona el modelo. En cuanto a posibles mejoras, se probó un modelo que contaba el número de procedimientos que se le hacían a cada paciente pero no se obtuvo una significancia relevante así que se descartó. Hacer una buena clasificación en grupos de procedimientos podría dar muy buenos resultados.

Otra posibilidad de mejora es tomar aquellas secciones del CIE10 que se consideren especialmente generales (e.g. Enfermedades infecciosas intestinales) y hacer unas subsecciones sin aumentar excesivamente el número de estas (ya que al hacer la interacción con edad se duplican).

Aun así, el modelo final ha visto una mejora considerable en cuanto a rendimiento predictivo y las evaluaciones realizadas a este lo corroboran.