

Modelado predictivo de materiales: EDA, evaluación, simulación y explicación física

Harold Sebastián Rodríguez
UAX

November 11, 2025

Contents

1	EDA (Exploratory Data Analysis)	2
1.1	Semiconductor (Clasificación)	2
1.2	Estabilidad (Clasificación)	4
1.3	Fotovoltaico (Clasificación)	6
1.4	Band gap (Regresión)	8
2	Evaluación de modelos	10
2.1	Semiconductor (Clasificación)	10
2.2	Estabilidad (Clasificación)	11
2.3	Fotovoltaico (Clasificación)	13
2.4	Band gap (Regresión)	14
3	Simulación con datos nunca vistos	14
3.1	Estabilidad (Clasificación)	15
3.2	Fotovoltaico (Clasificación)	15
3.3	Band gap (Regresión)	15
4	Explicación física e interpretabilidad	16

1 EDA (Exploratory Data Analysis)

1.1 Semiconductor (Clasificación)

Usamos el rango de la masa que es una variable que hemos creado para a partir del máximo y mínimo de la masa.

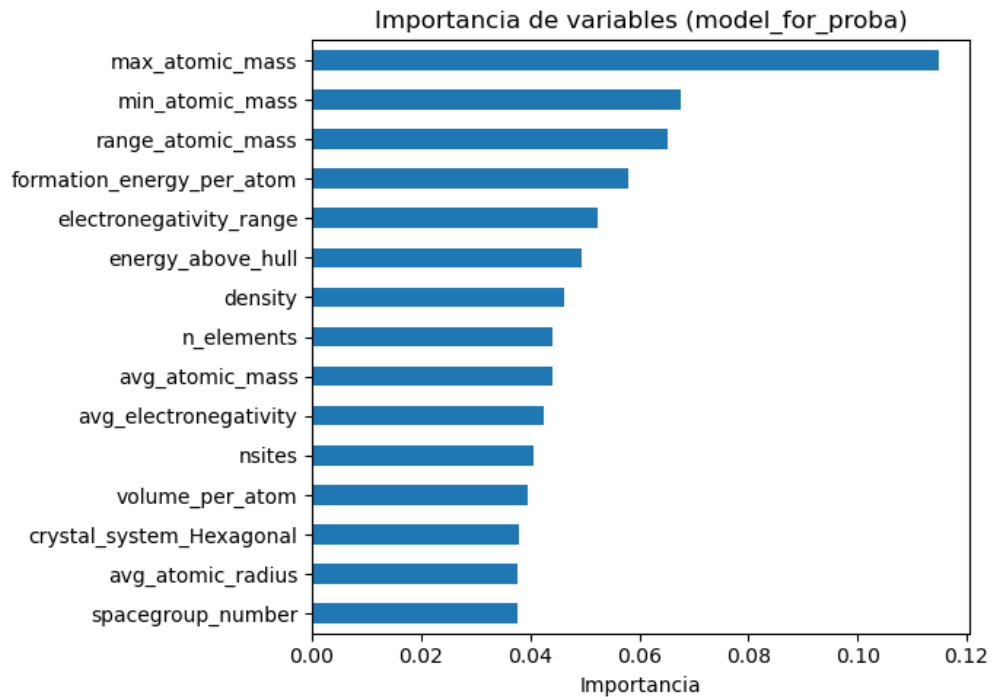


Figure 1: Importancias del modelo

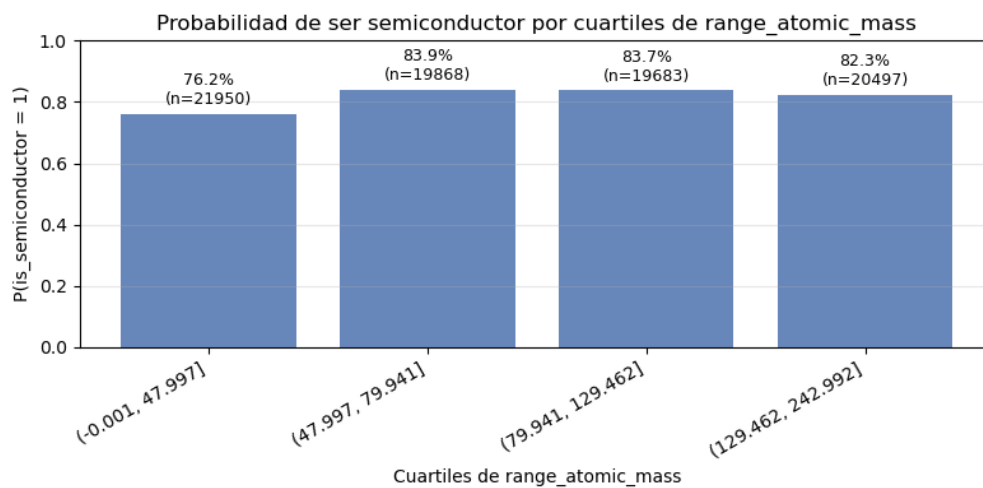


Figure 2: Masa vs semiconductor

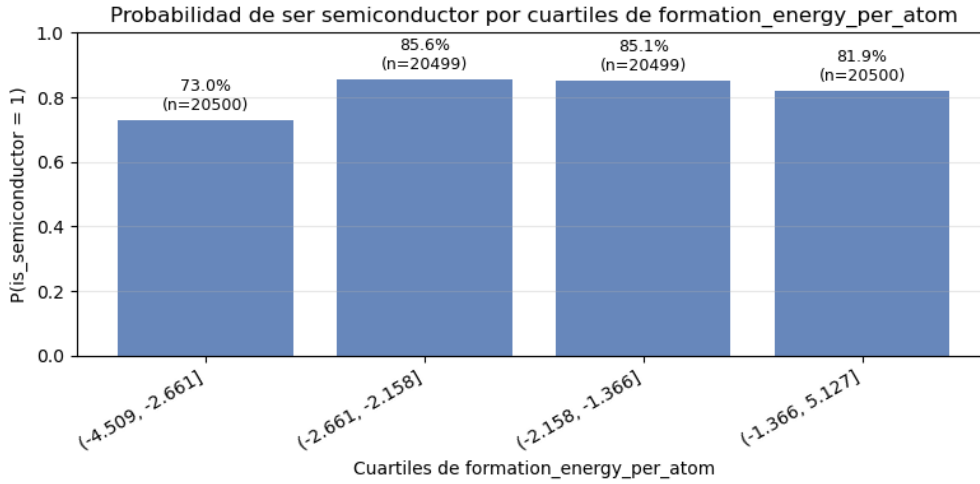


Figure 3: Energía vs semiconductor

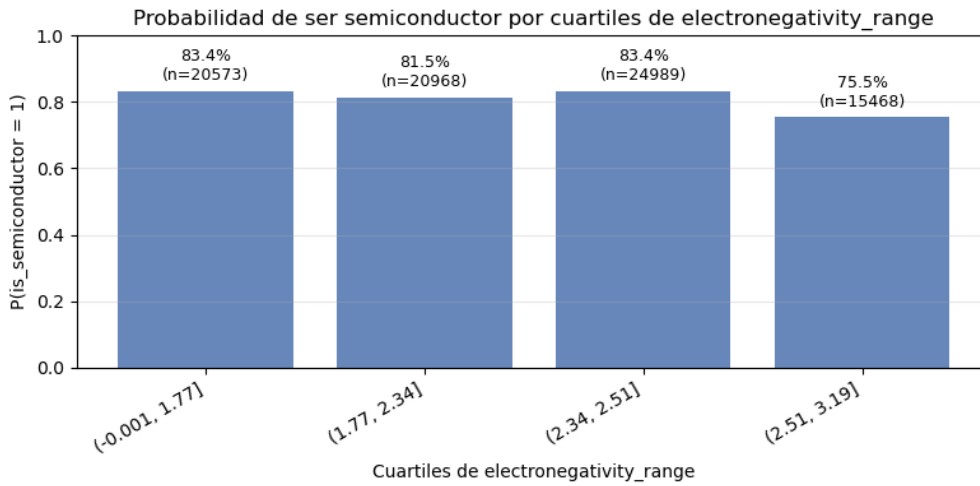


Figure 4: Electronegatividad vs semiconductor

La probabilidad de que un material sea semiconductor puede explicarse a partir de la interacción entre tres propiedades fundamentales: la **electronegatividad**, la **masa atómica** y la **energía de formación por átomo**.

El **rango de electronegatividad** entre los elementos de un compuesto determina el tipo de enlace químico predominante. Cuando este rango es muy alto, los enlaces tienden a ser iónicos, generando una gran separación entre las bandas de valencia y conducción, lo que da lugar a materiales aislantes. Por el contrario, un rango muy bajo favorece enlaces metálicos, con solapamiento de bandas y ausencia de gap. Los materiales con un rango *intermedio* presentan enlaces parcialmente covalentes, abriendo un *band gap* finito característico de los semiconductores.

La **masa atómica** influye en la anchura de las bandas electrónicas y en el acoplamiento espín-órbita. Elementos más pesados tienden a estrechar o escindir las bandas, facilitando la aparición de un gap, mientras que los más ligeros tienden a mantenerlas anchas, promoviendo el comportamiento metálico. Una mezcla de masas moderadas o elevadas puede así favorecer la semiconductividad.

Finalmente, la **energía de formación por átomo** condiciona la viabilidad termodinámica del material. Sólo los compuestos con energías de formación suficientemente negativas pueden estabilizar la estructura cristalina que da lugar al gap.

En conjunto, se plantea la hipótesis de que la combinación de un *rango de electronegatividades intermedio*, *masas atómicas moderadas o altas* y una *energía de formación suficientemente negativa*

define una región de estabilidad donde el material presenta un band gap adecuado, es decir, un comportamiento **semiconductor**.

1.2 Estabilidad (Clasificación)

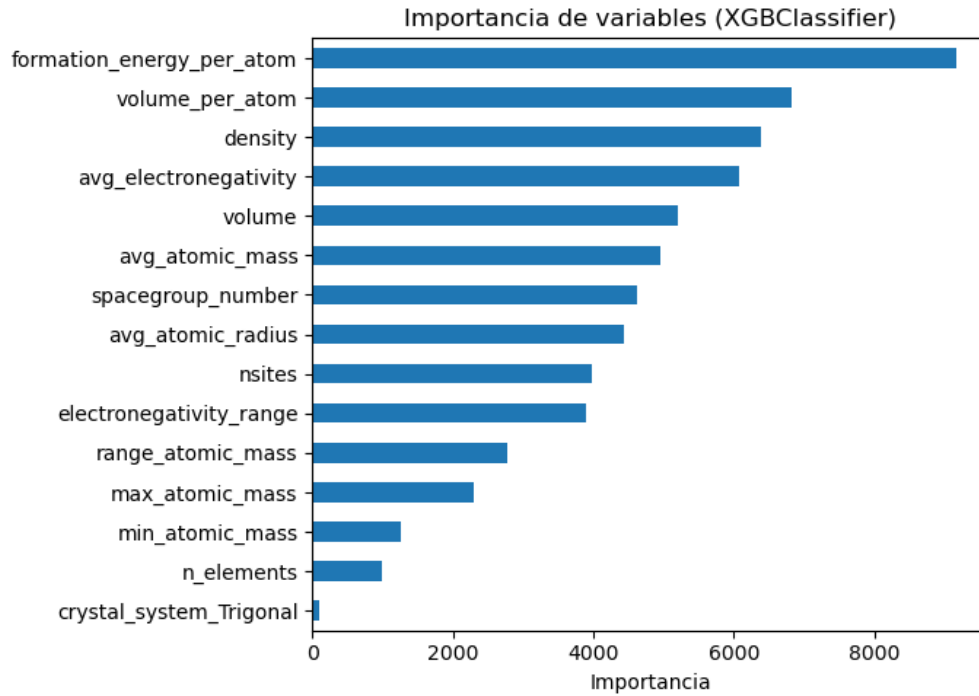


Figure 5: Importancias del modelo

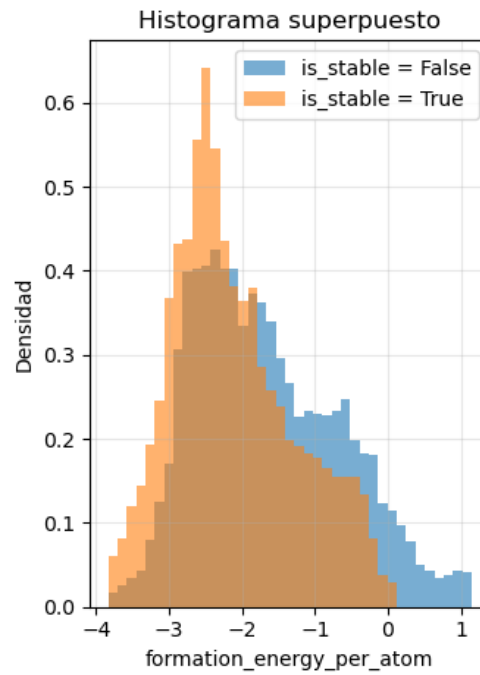


Figure 6: Energía vs Estabilidad

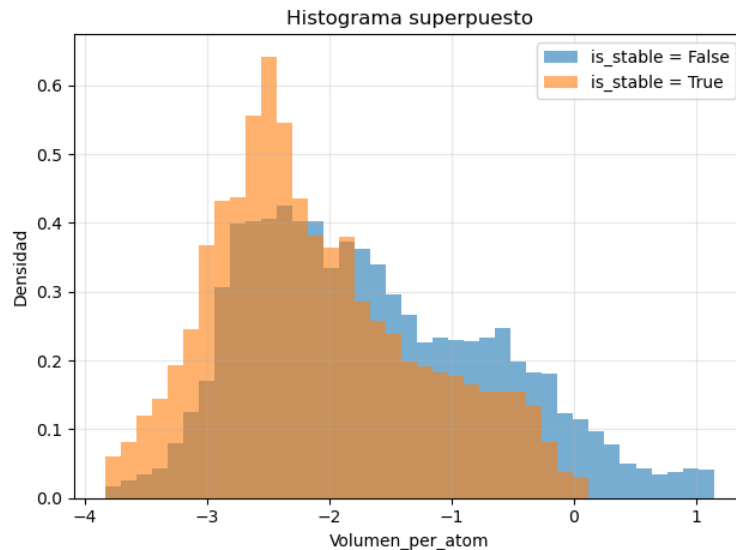


Figure 7: Volumen vs Estabilidad

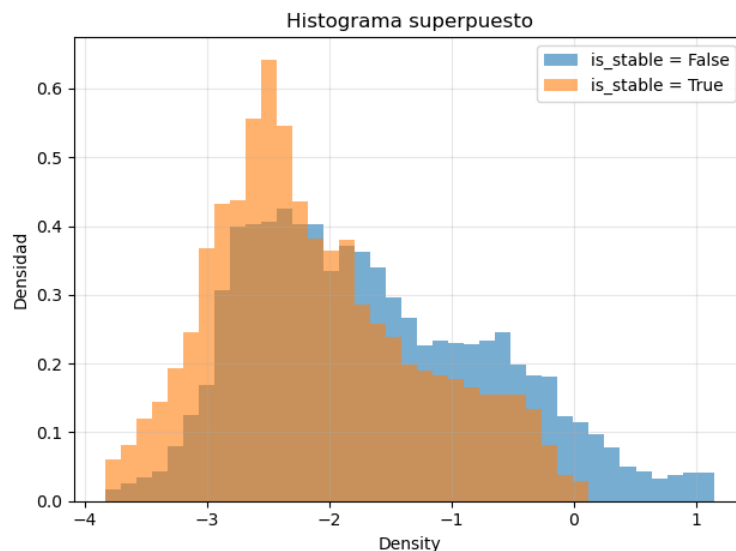


Figure 8: Densidad vs Estabilidad

Energía de formación Indicador más directo de estabilidad intrínseca: valores bajos señalan fases más favorables y resistentes a descomponerse.

volumen Mide cuán “apretada” está la estructura. A presiones altas, las fases que ocupan menos volumen tienden a ser preferidas; útil para comparar polimorfos.

Densidad Refleja el empaquetamiento de la red cristalina. Materiales más densos suelen ser favorecidos cuando aumenta la presión; aporta información similar a la del volumen.

Lectura conjunta La energía de formación marca la base de cuán estable es una fase; el volumen (y, de forma correlacionada, la densidad) indica cómo cambia esa estabilidad con la presión y qué estructura es más probable que domine bajo diferentes condiciones.

1.3 Fotovoltaico (Clasificación)

En los gráficos presentados luego comparamos la simetría del átomo ya que parece ser que es un variable importante en el modelo, luego veremos si tiene sentido físico en conjunto con las demás.

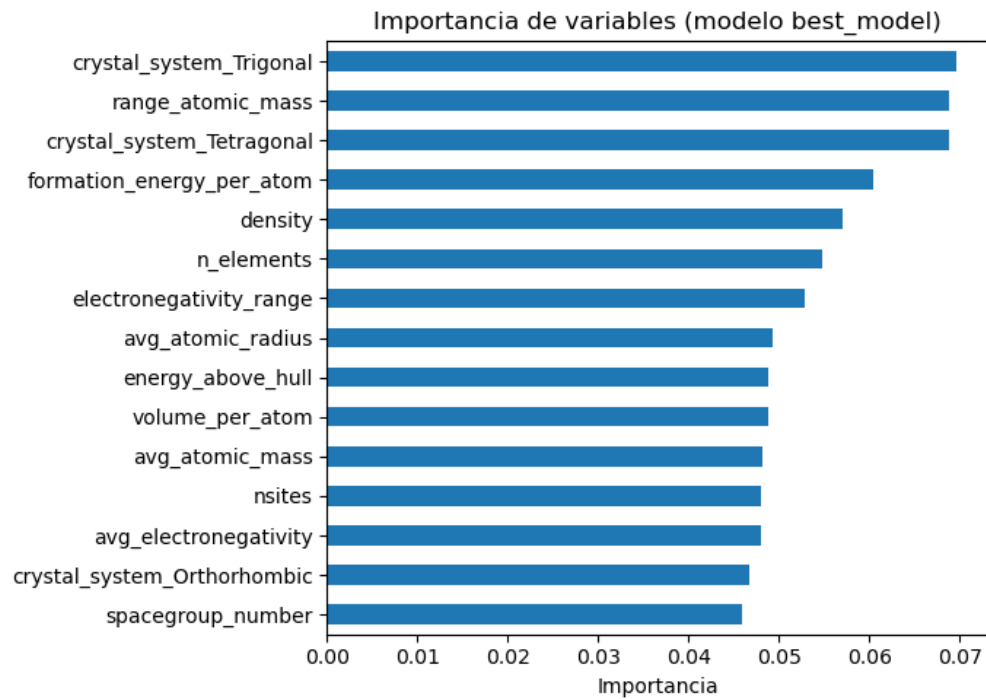


Figure 9: Importancias del modelo

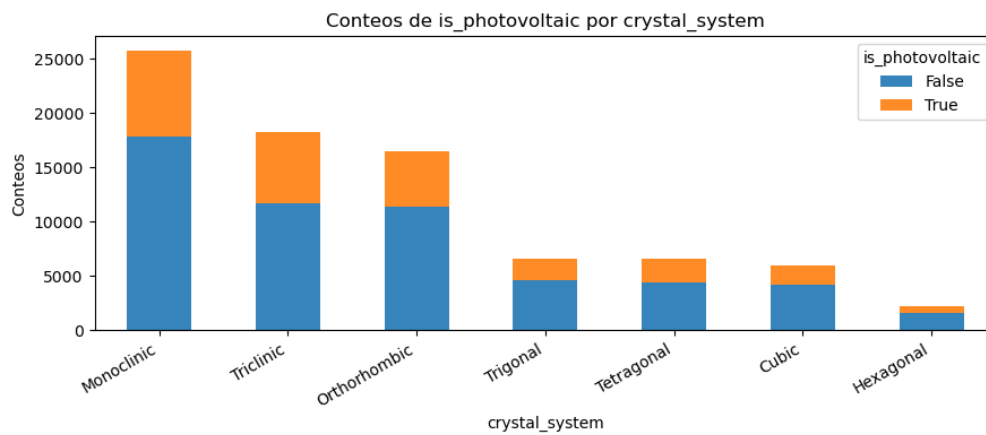


Figure 10: Simetría vs Fotovoltaico

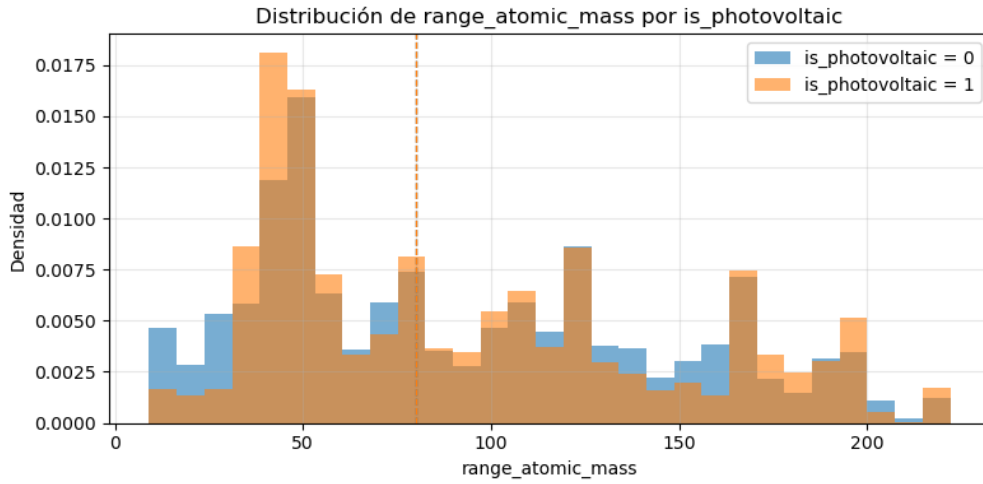


Figure 11: Masa vs Fotovoltaico

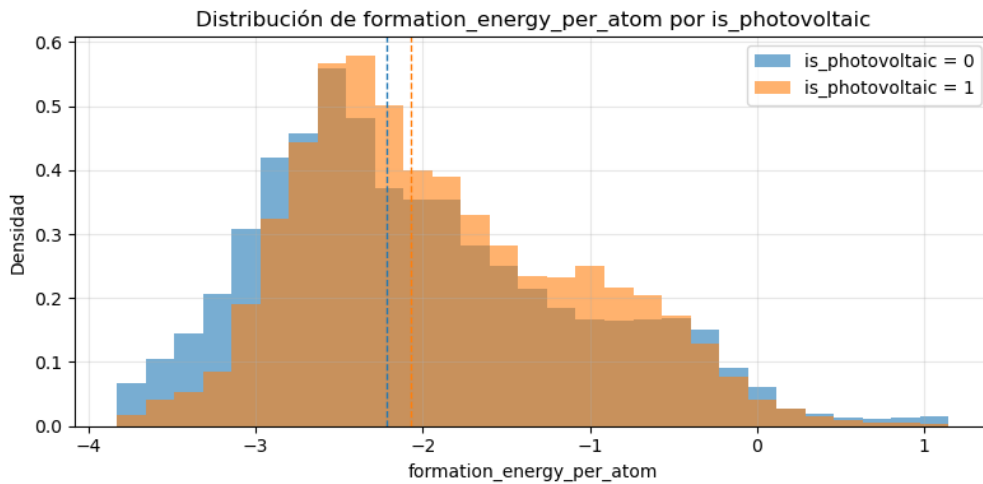


Figure 12: Energía vs Fotovoltaico

Simetría. Determina si el material presenta **brecha directa** y qué transiciones ópticas son **permitidas**

Rango de masas. Sirve como indicador de **elementos pesados** y, por tanto, de **acoplamiento espín-órbita (SOC)**. Un SOC alto puede ajustar el **band gap** a la ventana FV.

Energía de formación. Cuantifica la **estabilidad termodinámica** y la sintetizabilidad de la fase. Se correlaciona con las **energías de formación de defectos**: si las trampas profundas son energéticamente costosas, el material muestra **tolerancia a defectos**, lo que favorece mayor V_{oc} y vidas medias largas.

Relación. Un buen material FV combina: (i) **simetría** que permita **absorción eficiente** (gap directo y transiciones ópticas activas), (ii) **presencia de elementos pesados** (alto SOC, cribado dieléctrico y fonones favorables) reflejada en un *Rango de masas* elevado, y (iii) **estabilidad** suficiente (baja *Energía de formación*) que limite la formación de **defectos profundos**. Estas tres características en conjunto dan una información coherente para poder predecir si un material es FV.

1.4 Band gap (Regresión)

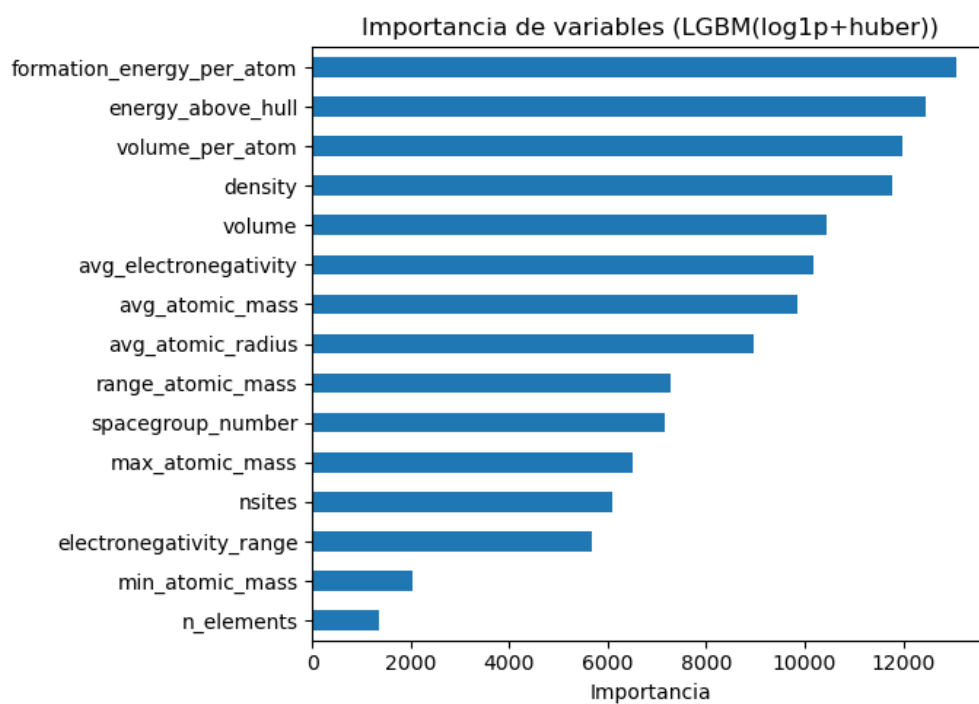


Figure 13: Importancias del modelo

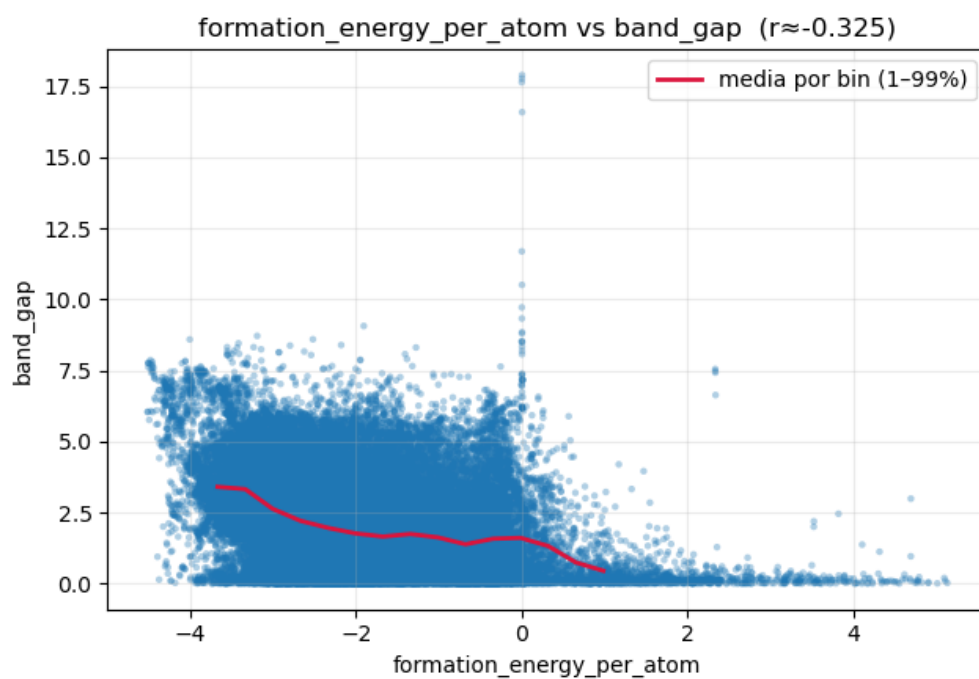


Figure 14: Energía de formación VS Band Gap

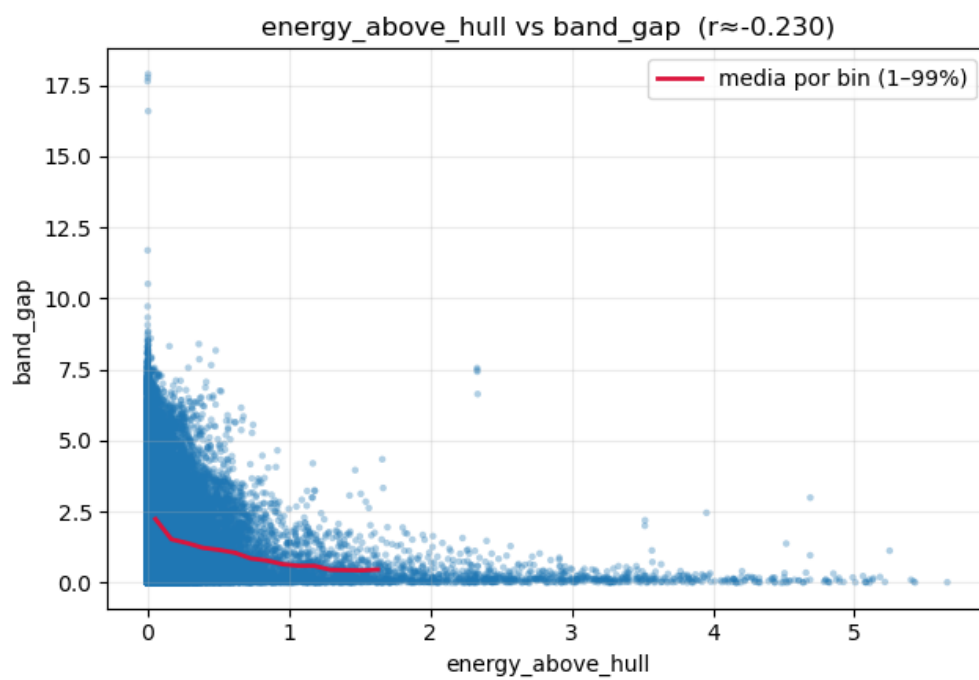


Figure 15: Energía por encima del casco VS Band Gap

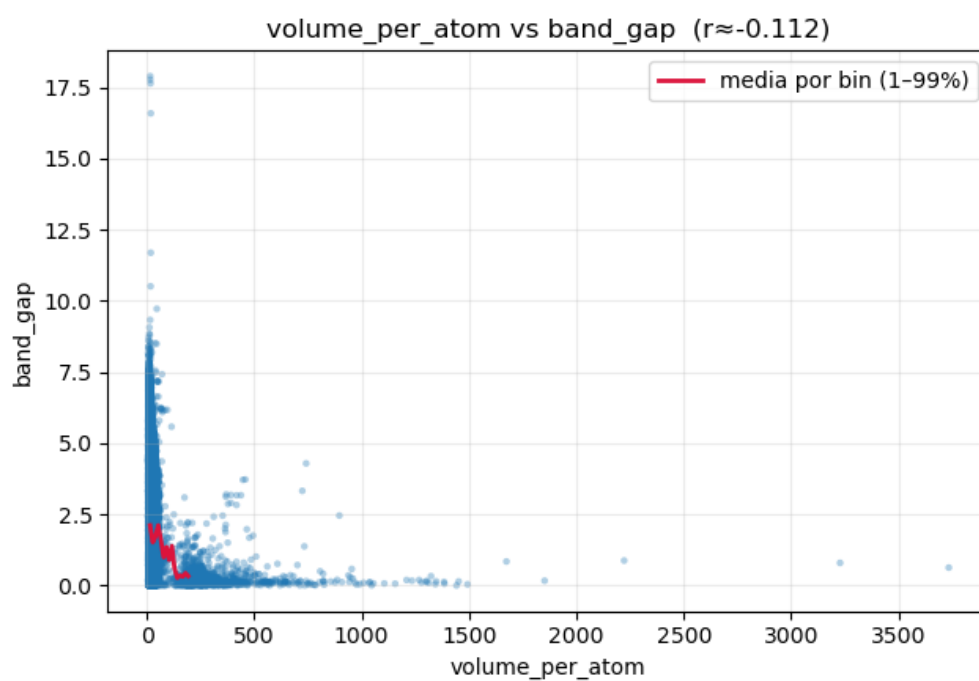


Figure 16: Volumen VS Band Gap

2 Evaluación de modelos

2.1 Semiconductor (Clasificación)

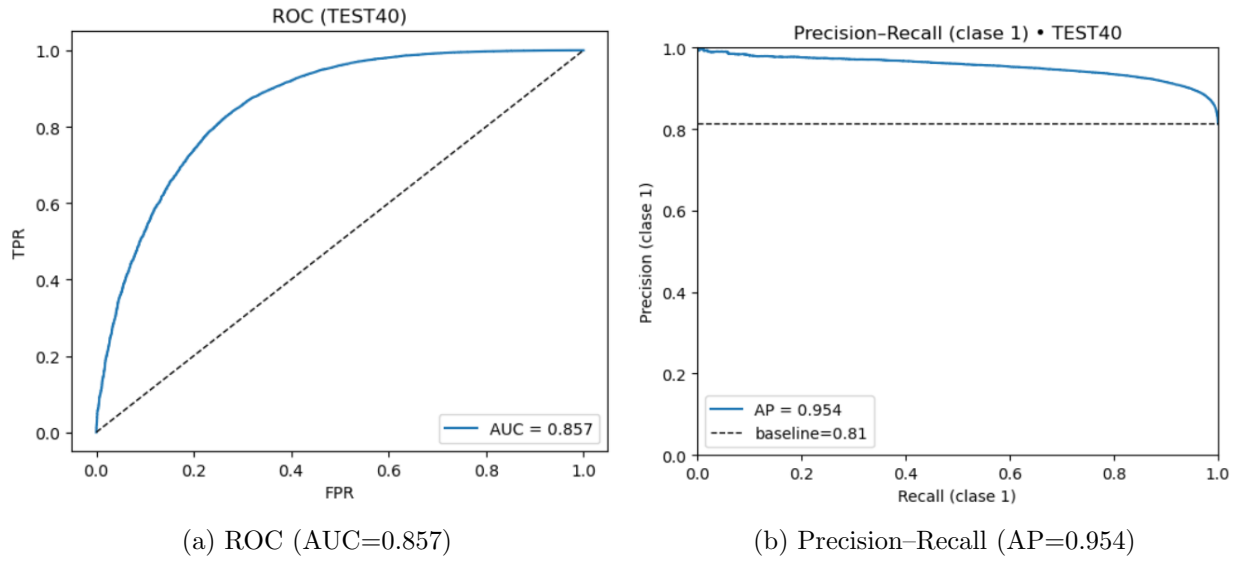


Figure 17: Semiconductor (Clasificación): curvas en TEST.

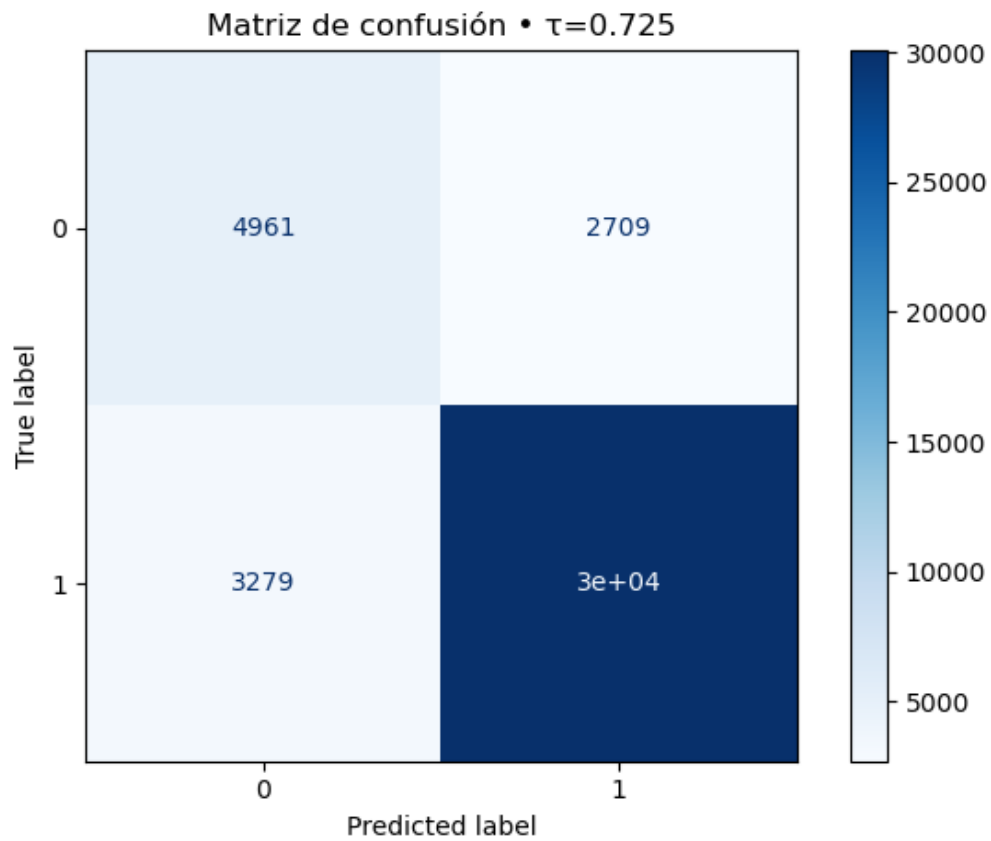
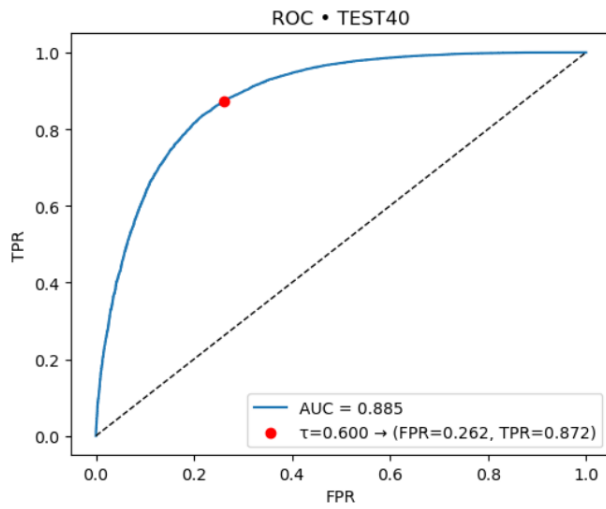


Figure 18: Matriz de confusión

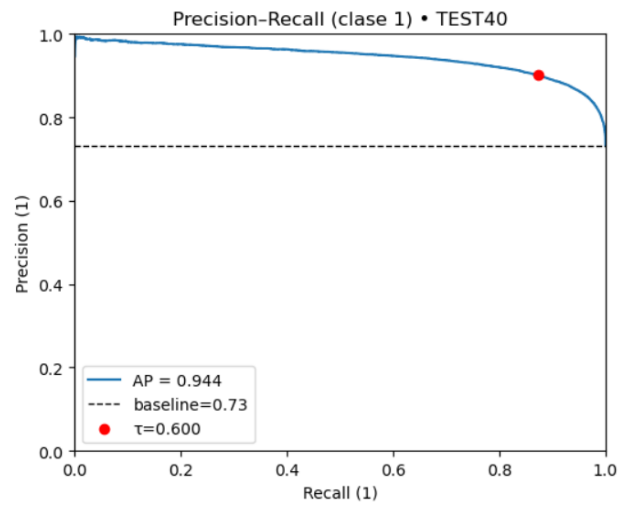
Reporte de clasificación:				
	precision	recall	f1-score	support
0	0.602	0.647	0.624	7670
1	0.917	0.902	0.909	33329
accuracy			0.854	40999
macro avg	0.760	0.774	0.767	40999
weighted avg	0.858	0.854	0.856	40999

Figure 19: métricas (VALIDACIÓN y TEST)

2.2 Estabilidad (Clasificación)



(a) ROC (AUC=0.872)



(b) Precision-Recall (AP=0.944)

Figure 20: Semiconductor (Clasificación): curvas en TEST.

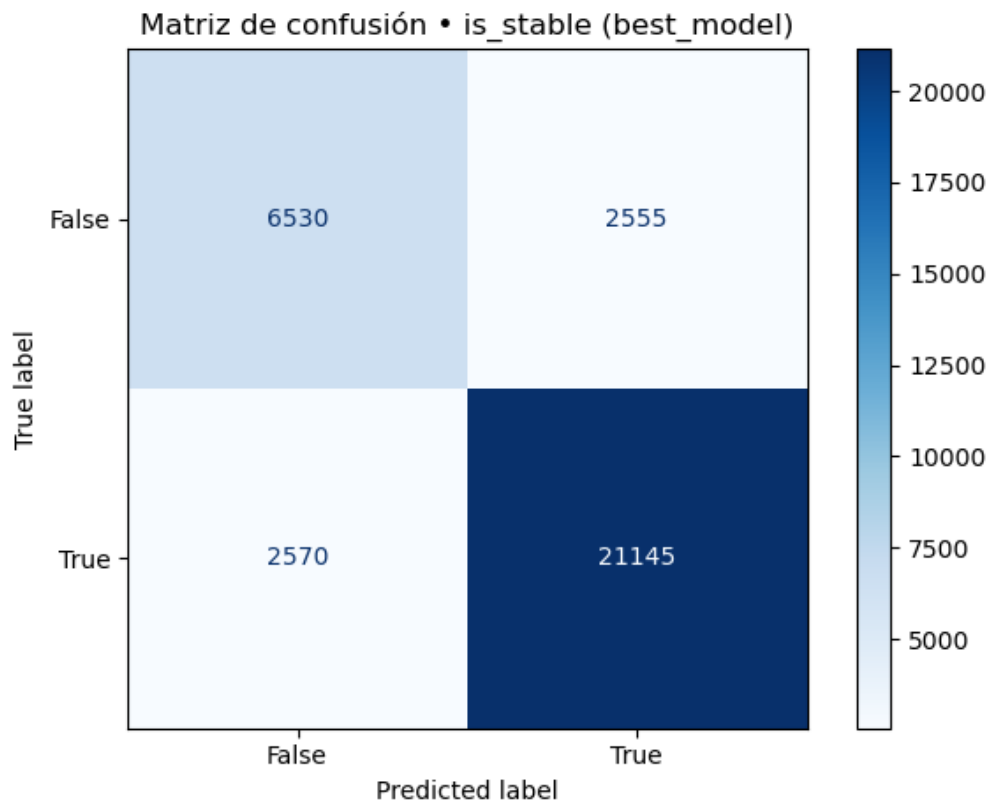


Figure 21: Matriz de confusión

	precision	recall	f1-score	support
False	0.718	0.719	0.718	9085
True	0.892	0.892	0.892	23715
accuracy			0.844	32800
macro avg	0.805	0.805	0.805	32800
weighted avg	0.844	0.844	0.844	32800

Figure 22: métricas (VALIDACIÓN y TEST)

2.3 Fotovoltaico (Clasificación)

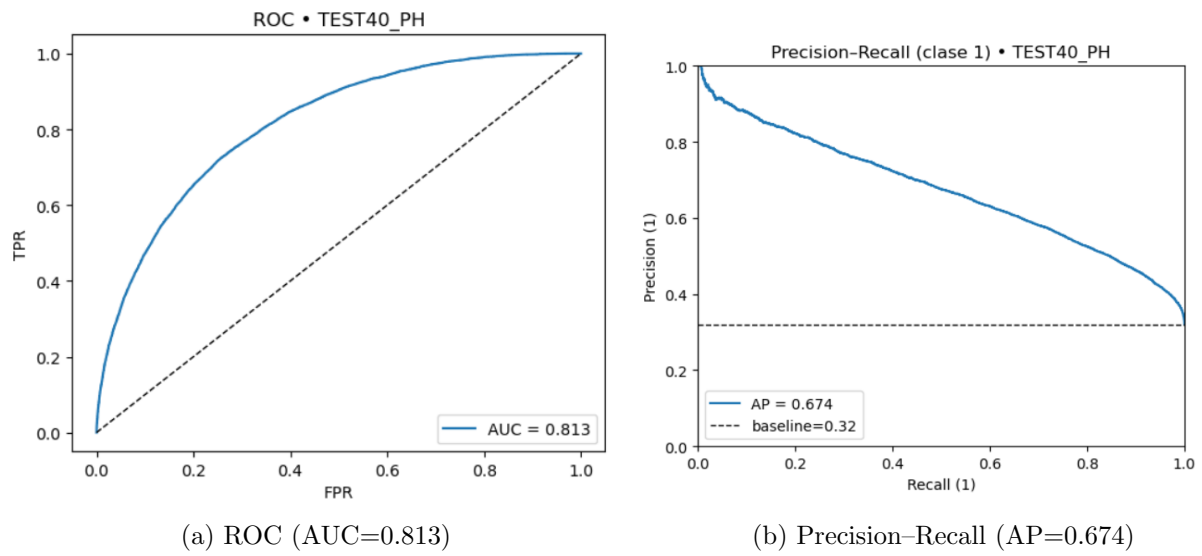


Figure 23: Semiconductor (Clasificación): curvas en TEST.

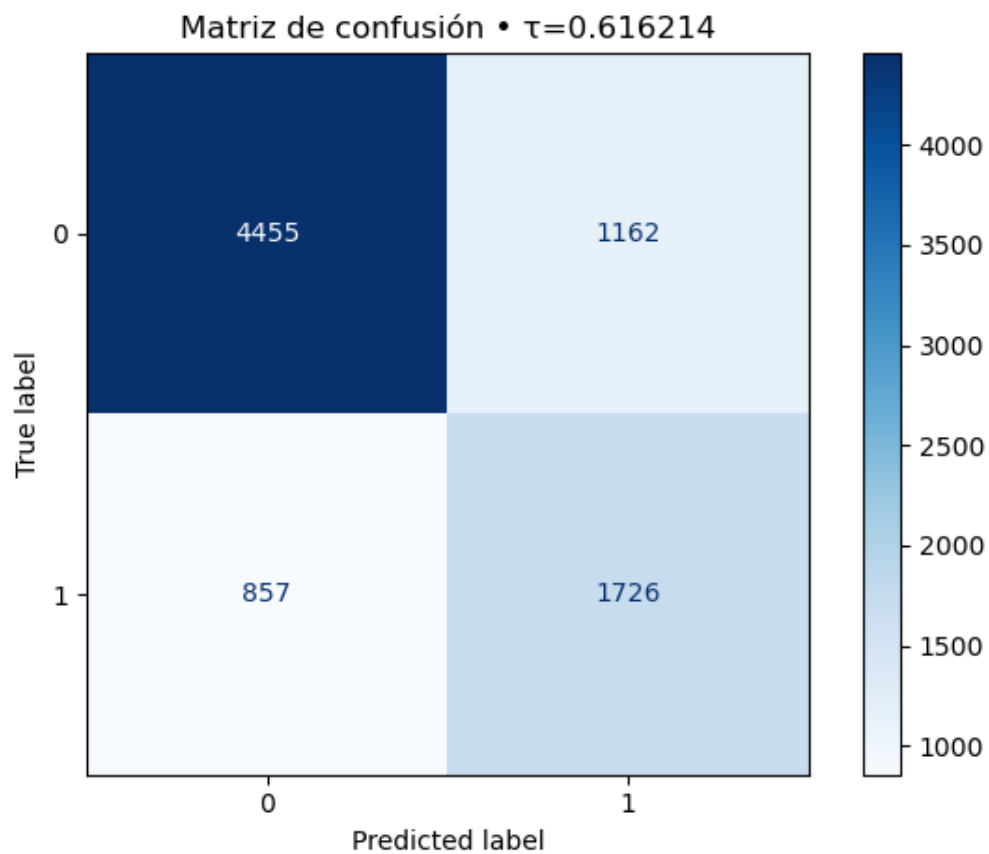


Figure 24: Matriz de confusión

	precision	recall	f1-score	support
0	0.829	0.801	0.815	27862
1	0.607	0.650	0.628	13137
accuracy			0.753	40999
macro avg	0.718	0.726	0.721	40999
weighted avg	0.758	0.753	0.755	40999

Figure 25: métricas (VALIDACIÓN y TEST)

2.4 Band gap (Regresión)

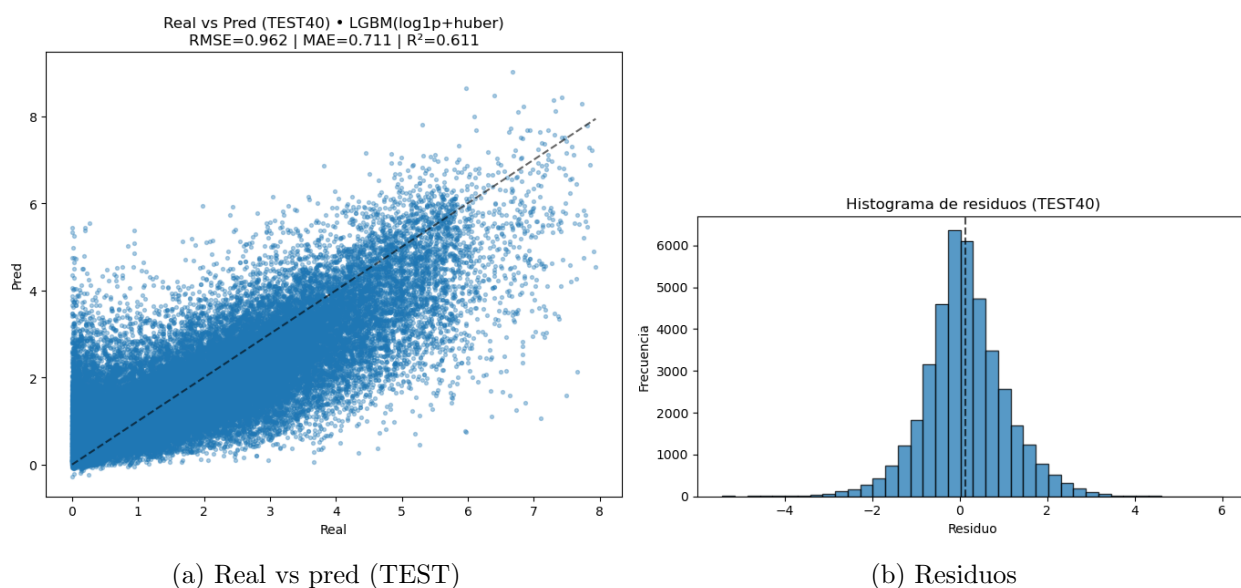


Figure 26: Band gap (Regresión): diagnósticos en TEST.

3 Simulación con datos nunca vistos

En este apartado evaluaremos nuestro modelo con 40K nunca vistos.

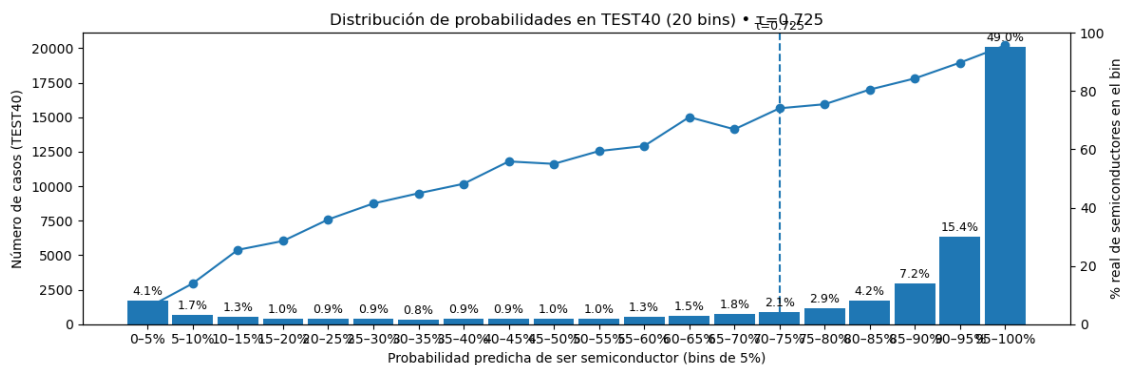


Figure 27: Calibración / Reliability (TEST40, Semiconductor (Clasificación)).

3.1 Estabilidad (Clasificación)

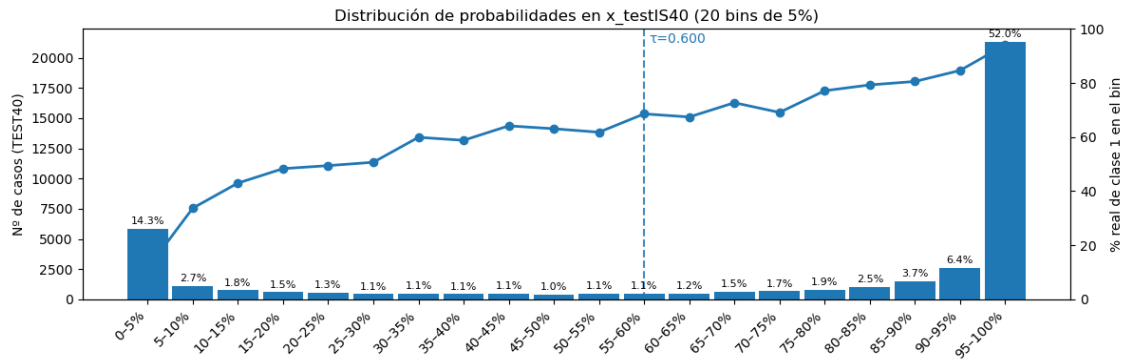


Figure 28: Calibración / Reliability (TEST40, Semiconductor (Clasificación)).

3.2 Fotovoltaico (Clasificación)

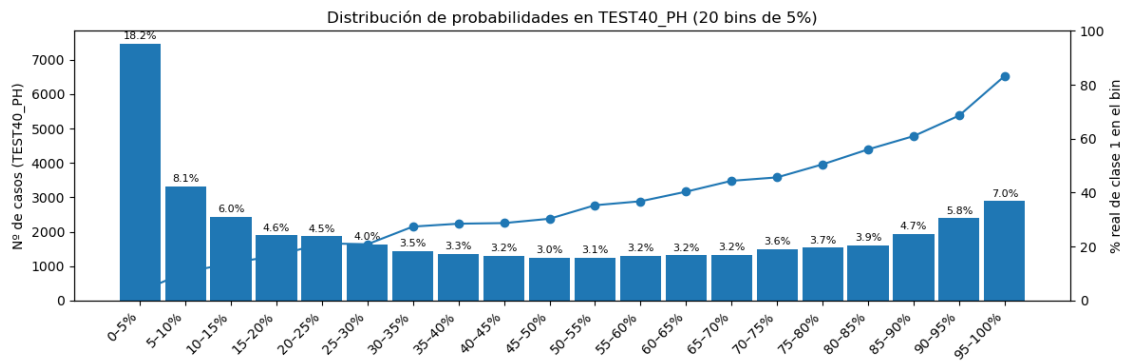


Figure 29: Calibración / Reliability (TEST40, Semiconductor (Clasificación)).

3.3 Band gap (Regresión)

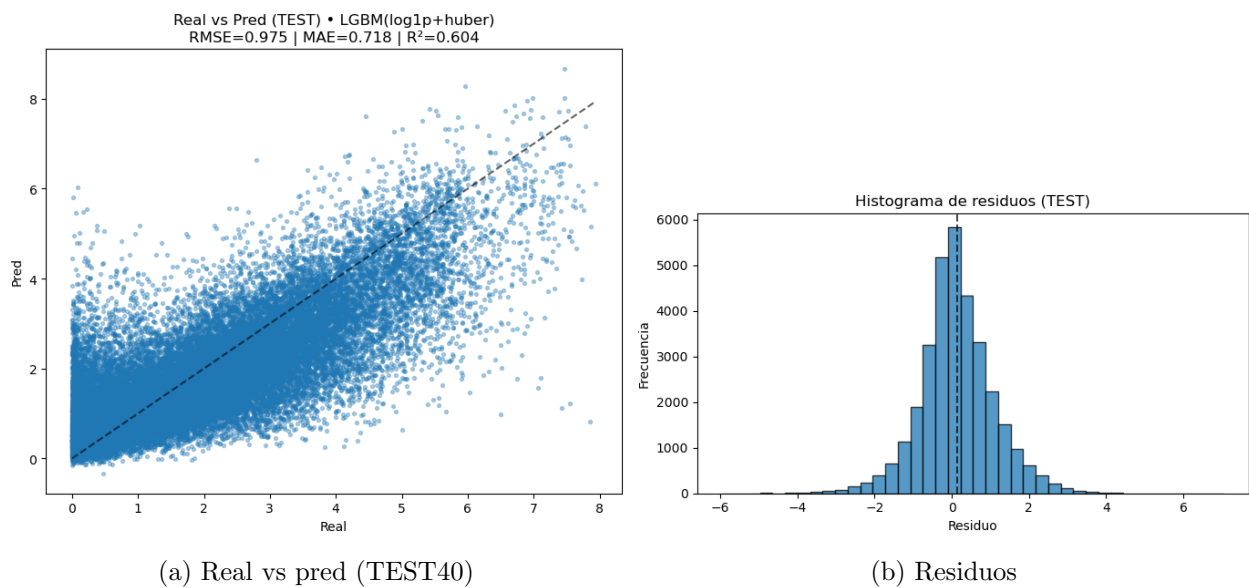


Figure 30: Band gap (Regresión) en datos nunca vistos.

4 Explicación física e interpretabilidad

Explicación física de materiales semiconductores. Un sólido es semiconductor si $E_g > 0$ y la densidad intrínseca es pequeña:

$$n_i = \sqrt{N_c N_v} e^{-\frac{E_g}{2k_B T}} \Rightarrow \text{si } E_g \gg k_B T \text{ (a 300 K), no es metálico.}$$

Rango de electronegatividad $\Delta\chi$ (tipo de enlace $\rightarrow E_g$). La fracción de ionicidad (Pauling/Phillips–Van Vechten) crece con $\Delta\chi$:

$$f_i \approx 1 - \exp\left[-\frac{(\Delta\chi)^2}{4}\right],$$

y el *band gap* aumenta con la ionicidad y disminuye con la covalencia (más *hopping*). En un modelo de *tight-binding*:

$$E_g \approx \underbrace{(\varepsilon_A - \varepsilon_B)}_{\text{polaridad} \propto \Delta\chi} - 2|t(d)|, \quad t(d) \simeq t_0 e^{-\alpha d}.$$

Mayor $\Delta\chi \Rightarrow$ mayor polaridad y, típicamente, mayor E_g .

Masa atómica / número atómico (SOC y ancho de banda). El acoplamiento espín-órbita escala como

$$\Delta_{\text{SO}} \propto \frac{Z^4}{n^3},$$

y puede abrir o modificar el gap cerca del VBM/CBM:

$$E'_g \simeq E_g \pm \Delta_{\text{SO}}.$$

Además, átomos pesados suelen dar mayor apantallamiento y bandas más estrechas, lo que (en ausencia de gran *hopping*) favorece $E_g > 0$.

Energía de formación por átomo E_f (estabilidad y defectos). Materiales con E_f más negativo son termodinámicamente accesibles y tienden a presentar mayores energías de formación de defectos:

$$E_f^{\text{def}} = E_{\text{def}} - E_{\text{bulk}} - \sum_i n_i \mu_i + q(E_F + E_V) + \Delta,$$

reduciendo la densidad de estados profundos que cerrarían el gap efectivo. Así, E_f no fija E_g , pero filtra fases “buenas” (gap limpio, no metálicas).

Conclusión.

$\Delta\chi$ (polaridad) $\uparrow \Rightarrow E_g \uparrow$; Z (SOC) $\uparrow \Rightarrow E'_g$ ajustado ; $E_f \downarrow \Rightarrow$ fase estable con gap realizable
--

Explicación física de materiales estables

1) Energía de formación. Es el indicador *directo* de estabilidad química. Una energía de formación más negativa implica que la fase, al formarse desde sus referencias, libera más energía y por tanto *no le conviene* descomponerse. En equilibrio, las fases observadas son las de menor energía libre; a presión moderada y temperatura no muy alta,

$$G \simeq E_{\text{form}} \Rightarrow E_{\text{form}} \text{ bajo} \Rightarrow \text{fase estable.}$$

2) Volumen. La estabilidad también depende de la presión a través de

$$G = U + PV - TS, \quad \left(\frac{\partial G}{\partial P}\right)_T = V.$$

A mayor presión, las fases con *menor volumen por átomo* disminuyen más su G y resultan favorecidas. Esta variable distingue bien entre *polimorfos* de igual composición.

3) densidad. Para una composición dada, densidad alta \leftrightarrow volumen molar bajo. Por ello, las fases *más densas* suelen ser preferidas cuando aumenta la presión, y reflejan un empaquetamiento más compacto (mayor cohesión).

Explicación física de materiales fotovoltaico

1) Simetría. La simetría del cristal influye en el tipo de *band gap*. En materiales con gap **directo**, la absorción óptica es más eficiente:

$$\alpha_{\text{directo}} \propto \sqrt{h\nu - E_g},$$

mientras que en los **indirectos** requiere fonones y es más débil. Por eso, ciertos sistemas cristalinos (como los no centrosimétricos) favorecen transiciones directas y mejor absorción solar.

2) Masa atómica . Un rango amplio de masas implica presencia de elementos pesados, asociados a un acoplamiento espín-órbita (SOC) más fuerte:

$$\Delta_{\text{SO}} \propto Z^4.$$

Este efecto puede ajustar el tamaño del gap dentro de la ventana ideal para la conversión fotovoltaica (1–2 eV) y mejorar el cribado de defectos, reduciendo recombinación.

3) Energía de formación. Indica estabilidad termodinámica. Si la energía de formación es baja (más negativa), la fase es estable y menos propensa a generar defectos:

$$c_{\text{def}} \propto e^{-E_f^{\text{def}}/k_B T}.$$

Menos defectos profundos implican menor recombinación y mayor voltaje de salida.

Explicación física de materiales con band gap

1) Energía de formación por átomo. Refleja la estabilidad y el tipo de enlace entre los átomos. Enlaces más fuertes (energía de formación más negativa) suelen corresponder a mayor separación entre los niveles de enlace y antienlace, lo que produce un gap más amplio:

$$E_g \propto |\varepsilon_{\text{antienlazante}} - \varepsilon_{\text{enlazante}}|.$$

Por tanto, materiales muy estables tienden a tener un band gap mayor.

2) Energía por encima del casco . Indica la estabilidad relativa frente a otras fases posibles. Fases cercanas al casco ($E_{\text{above hull}} \approx 0$) son las que pueden existir realmente. Aunque no determina directamente el valor del gap, garantiza que el gap calculado corresponda a una estructura físicamente realizable y no a una fase inestable.

3) Volumen por átomo. El volumen afecta a las distancias interatómicas y, por tanto, al solapamiento orbital. Cuando el volumen disminuye (átomos más cercanos), aumenta el solapamiento y el gap tiende a reducirse:

$$t(d) \propto e^{-\alpha d}, \quad E_g \approx (\varepsilon_A - \varepsilon_B) - 2|t(d)|.$$

Por el contrario, estructuras más abiertas (mayor volumen) suelen presentar un band gap más grande.