Predicción de banda prohibida en materiales bidimensionales con aprendizaje máquina.

Jair Othoniel Dominguez Godinez[1]

[1]Centro de Investigación Científica y de Educación Superior de Ensenada

15 de abril de 2024

Resumen

La predicción de la banda de energía prohibida en materiales emiconductores es un problema que se resuelve a partir de simulaciones de primeros principios pero estas son muy costosas computacionalmente hablando por lo que se requieren recursos como el supercómputo y la paralelización de cálculos. Los materiales bidimensionales han atraído la atención de la comunidad científica por sus amplias aplicaciones en catálisis, nanoelectrónica, almacenamiento de energía, espintrónica y producción de hidrógeno. Especialmente los semiconductores 2D han mostrado aplicaciones en muchos de estos campos y se consideran como la siguiente generación de nanomateriales. En este trabajo se abordó dicho problema utilizando técnicas de aprendizaje máquina supervisado. Se minó una base de datos de materiales bidimensionales de la cual se extrajo el valor de banda prohibida para seiscientos materiales, se filtraron aquellos valores que estuvieran entre 0.5 eV y 3 eV. Se generó un espacio de características de 38 dimensiones y a través de técnicas de extracción de características se redujo el espacio para entrenar un conjunto de modelos de aprendizaje máquina. Se probaron modelos lineales, máquinas de soporte vectorial, arboles de decisión, metaensambles (Boosting) y una red neuronal artificial. Mis resultados indican que a pesar de un bajo rendimiento en la predicción los modelos de AM son una herramienta poderosa y computacionalmente menos costosa para predecir propiedades de nanoestructuras.

Palabras clave. Materiales bidimensionales, banda prohibida, aprendizaje máquina y aprendizaje supervisado.

1. Introducción.

Los nanomateriales bidimensionales son la nueva generación de nanotecnologías que nos permiten enfrentar problemas como la contaminación a través de la degradación de contaminantes y la síntesis de nuevos nanomateriales con

una mejor eficiencia en el almacenamiento o generación de energía que permitan una transición hacia energías más limpias. Los materiales 2D resultan de gran interés ya que muestran un rango muy amplio de propiedades que pueden ser aplicadas en catálisis, dispositivos electrónicos, optoelectrónicos, trampas y detectores de moléculas tóxicas, producción de hidrógeno y espintrónica.
Para estas aplicaciones la búsqueda y predicción de nuevos materiales resulta crucial porque existe una rica familia de materiales bidimensionales como los Dicalcogenuros de Metales de Transición (TMD), MXenes, Compuestos Metal-Orgánicos (MOF's), estructuras 2D monoatómicas como el graféno o el siliceno, entre mechas más. Así que debido a la gran variedad y número de monocapas que se pueden obtener es preciso predecir y buscar nuevos materiales 2D.
Tradicionalmente las simulaciones computacionales de primeros principios que están basadas en la Teoría Funcional de la Densidad desarrollada por Pierre Hohenberg, Walter Kohn y Lu Sham, en la aproximación de Teoría de Perturbaciónes de Muchos Cuerpos (GW) resulta dar muy buenos resultados al comparar con los valores experimentales de la banda prohibida de semiconductores, sin embargo resulta computacionalmente costosa en los cálculos. Por su parte las aproximaciones de Densidad Local (LDA) y Perdew-Burke-Ernzerhof (PBE) a la interacción de correlación e intercambio de los electrones son menos costosas sin embargo suelen subestimar el valor de la banda prohibida siendo una opción menos confiable para predecir dicho valor [1]. La DFT nos permite predecir distintas propiedades electrónicas de nanoestructuras, sin embargo su implementación resulta computacionalmente costosa, estamos hablando de cálculos que pueden tardar desde días hasta meses. El interés en predecir la banda prohibida (E_g) radica en el hecho de que esta es la energía necesaria para hacer que un electrón en la banda de valencia brinque a la banda de conducción. En este sentido una aproximación computacional diferente y menos costosa en tiempo y recursos resulta necesaria para estudiar desde otro enfoque la predicción de propiedades en nanomateriales. Predecir la E_g resulta determinante ya que esto nos permite predecir en que rango de energías del espectro de radiación electromagnético el material va a necesitar energía para que el electrón de la banda de valencia pase a la banda de conducción.
Actualmente existen una gran cantidad de bases de datos de propiedades físicas y químicas de materiales que se obtienen tanto de experimentos como de simulaciones, un ejemplo de estos son Open Quantum Materials Database y Materials Project que almacenan una gran cantidad de datos de estructuras cristalinas ofreciendo información de los parámetros de red, propiedades electrónicas, térmicas, estructurales, etc. Aprovechando estas bases de datos se puede trabajar sobre ellas haciendo uso de metodologías basadas en la ciencias de datos para obtener, filtrar y analizar información mediante el uso de técnicas de aprendizaje máquina (AM) que aprende de patrones dados para mapear un espacio de características a un espacio de menor dimensión que definimos como nuestra variable objetivo. El AM es una subárea de la inteligencia artificial (IA) que se encarga de estudiar y aplicar modelos estadísticos para determinar patrones en los datos. El uso de técnicas basadas en aprendizaje estadístico puede permitir una toma de decisiones basada en argumentos cuantitativos para un diseño y

síntesis más inteligente, la incorporación de estas técnicas no pretende sustituir las simulaciones *ab-initio* sino más bien hacer una buena amalgama que nos permita un mayor poder de predicción y diseño de materiales.

2. Antecedentes.

2.1. Aprendizaje máquina e Inteligencia Artificial

Alguna vez hemos escuchado hablar de la industria 4.0 que busca explotar las tecnologías más recientes como los sistemas interactivos (Realidad Virtual y robótica), las nanociencias toman su papel en las industrias farmaceúticas, energéticas y electrónicas, la salud y la agricultura, el internet de las cosas (sistemas interconectados) y el Big Data que básicamente es el aprovechamiento de las enormes bases de datos para la toma de decisiones y estrategias.
Dentro de este paradigma la IA es un área del conocimiento que tiene muchas aplicaciones como la conducción autónoma, la visión por computadora, el reconocimiento de patrones, el procesamiento del lenguaje natural, sistemas autónomos y optimización de procesos. Se entiende por IA como el estudio de cómo las máquinas aprenden. Entendemos por agente inteligente como todo aquel programa de computadora, ginoide, androide máquina o robot que mejora su experiencia E respecto a una tarea T de acuerdo a una medida de rendimiento P [18].
El aprendizaje máquina forma parte de la IA y se define como el conjunto de métodos matemáticos y computacionales que son capaces de detectar patrones en la información y hacer predicciones futuras. Existen diferentes paradígmas como las redes neuronales, los algoritmos genéticos, el aprendizaje basado ejemplos, el probabilístico y el analítico. El AM basado en ejemplos es de tres tipos supervisado, no supervisado y por refuerzo. El aprendizaje supervisado es capaz de resolver problemas de clasificación y regresión numérica. La información está representada en forma de bases de datos etiquetadas como espacio de características y objetivo. Esto implica que se conocen ejemplos de las predicciones en función de las características del fenómeno.
Un flujo de trabajo muy general en este campo es el siguiente. Conozco un conjunto de datos que vienen de la realidad, de esos datos puedo hacer una división en datos de entrenamiento de los cuales el modelo de AM es capaz de aprender, datos de evaluación que forman parte del conjunto de datos que permiten evaluar la capacidad de la predicción, estos datos se dan al aprendiz que será capaz de detectar patrones en los datos y posteriormente nos permitirá validar con nuevos datos nunca antes vistos por el modelo. Otras partes internas del trabajo consisten el minado de las bases de datos es decir recolección mediante técnicas computacionales para formar bases datos, la limpieza y filtrado ya que nuestros datos pueden no estar estructurados, ingeniería de características que consiste en la selección de características conforme a la importancia o peso que tienen según ciertas aproximaciones, la extracción de carcaterísticas o las reducciones de dimensionalidad que nos permitan pasar de una dimensión n a

una de menor dimensión mediante transformaciones lineales que conserven la información de los datos de mayor dimensionalidad en uno de menor (Análisis de Componentes Principales). Muchas veces es necesario hacer uso de una técnica llamada hiperparametrización que consiste en buscar aquellos valores para los parámetros de los moedelos que maximicen los rendimientos las mejores puntuaciones en las métricas de rendimiento como el error cuadrático medio o el coeficiente de determinación cuadrático. Estas métricas se aplican para un problema de regresión en el cual se busca una función $f(x)$ mediante un función de $h(x)$ llamada hipótesis que haga la mejor aproximación minimizando el error entre la predicción y el valor real dada cualquier x_i [19].

2.2. Informática de Materiales

Fruto de la generación de datos experimentales y teóricos de materiales provenientes de la investigación y la ingeniería se han formado bases de datos que recopilan o almacenan información que puede ser de utilidad para muestreos estadísticos, análisis de información y diseño de experimentos. La informática de materiales busca mejorar y predecir y mejorar el diseño materiales basado en técnicas computacionales como pueden ser simulaciones de computacionales de sistemas físicos, la aplicación de la ciencia de datos a la ciencia e ingeniería de materiales y el diseño inverso o directo de nanomateriales [20].

2.3. Aplicaciones en nanociencias

Actualmente existe una comunidad amplia y diversa de investigadores que adoptan metodologías de la IA para investigar, predecir y diseñar estructuras orgánicas e inorgánicas. Por ejemplo en el área de mecánica molecular específicamente en el docking molecular se utilizan algoritmos genéticos para la optimización de la energía de enlace entre proteínas y ligandos usando algoritmos genéticos Lamarkianos a este tipo de implementaciones se le conocen como algoritmos bioinspirados puesto que del paradigma que parten es una hipótesis de como funcionan la evolución, el cerebro y la conciencia en los organismnos vivos. Existe una IA generativa llamada AlphaFold que es un programa desarrollado por DeepMind de Alphabet, realiza predicciones de la estructura de las proteínas mediante el sistema de redes neuronales artificiales profundas.
Se ha aplicado el AM para predecir la banda prohibida en materiales semiconductores tomando diferentes bases de datos para estructuras cristalinas inorgánicas en bulto [7], para materiales 2D de la familia de MXenes funcionalizados [5], para la familia de los TMD's [8], para materiales bidimensionales en conjunto [9]. De lo más reciente se ha usado AM para la identificación de materiales topológicos [10]. En física teórica se han invertido las ecuaciones de Kohn-Sham para construir la contribución de la energía de correlación e intercambio informando redes neuronales [11]. Estas investigaciones aplican el flujo de trabajo del AM en la que básicamente se recolecta información fenomenológica del modelo del cual tenemos ejemplos o información, esta información viene en forma de estructuras de datos que tienen que ser filtrada y adaptada para que el mo-

delo pueda entender y trabajar los datos en una representación que permitan informar al modelo de AM y finalmente entrenar al modelo extraer para hacer predicciones del futuro.
En los artículos [9], [8] y [5] para poder predecir la banda prohibida se considera un espacio de características $\vec{X}$ que codifica información de propiedades estructurales, térmicas, electrónicas y cantidades estadísticas de propiedades químicas de cada elemento que constituye a la estructura. Este espacio de características considera que aquellas variables que mayor aportan a la predicción son aquellas de origen de simulaciones de DFT y otras características se extraen de modelos que codifcan información física y química en vectores numéricos mediante propiedades como la composición elemental en el material y las posiciones atómicas. A través del aprendiz se mapea espacio $\vec{X}$ a la variable objetivo y que es el valor a predecir y del cual tenemos ejemplos (regresión), que en este caso al ser un problema en el que el espacio de características es explicito se trata de un algoritmo de AM supervisado y un problema de regresión. Los modelos de AM que se pueden usar son lineales, máquinas de soporte vectorial, arboles aleatorios de regresión, metaensambles como el Bagging, y redes neuronales artificales (perceptrón multicapa).
El AM es una herramienta muy poderosa para el campo de las nanociencias mostrando increíbles aplicaciones tanto para la predicción, síntesis, análisis y diseño de nuevos nanomateriales. En este tipo de trabajos el espacio de característica depende de variables que tienen que ser calculadas a partir de primeros principios, es decir se utilizan como características de los modelos la energía de formación, los parámetros de red, la energía total del sistema, el tensor elástico y la energía de la banda prohibida utilizando la aproximación de gradiente generalizado de Perdew-Burke-Ernzerhof para describir la interacción electrónica. Esto resulta inconsistente con el propósito de dismuinir el tiempo en cálculos para la predicción de nanomateriales, esto implica que para cualquier nueva predicción es necesario dar esta información al modelo lo que resulta inconsistente, ya que queremos una predicción que sólo dependa de información ya disponible en las bases de datos y que no sea necesario hacer ningún tipo de experimento o simulación para obtener los datos de una nueva predicción de información no vista por el modelo. Mi propuesta es generar un espacio de características que sólo dependa de cantidades fáciles de obtener y que no tenga que hacer ninguna simulación o experimento extra para una nueva predicción.

3. Metodología computacional

Se ha adoptado un enfoque inspirado en la ciencia de datos y en las referencias [5] y [6]. Que permite un flujo de trabajo sobre las bases de datos para predecir la energía de la banda prohibida con un costo computacional y de recursos menor que los métodos *ab-initio* ya que no requiere del uso de supercómputo y todas las simulaciones se hicieron en mi máquina local.

3.1. Herramientas computacionales

El código de programación que se usó fue Python el cual resulta tener una gran cantidad de librerías y API's (interfaz de programación de aplicaciones) para obtener las bases de datos, manipularlas, aplicar los modelos de AM y visualizar los datos. La API usada es Matminer [4] que es una paquetería de informática de materiales que tiene incorporadas diferentes funciones como la generación de descriptores y el acceso a diferentes bases de datos de nanomateriales [4]. Se usó la paquetería de sklearn [12] que tiene implementadas funciones de AM tanto para regresión como para clasificación, preprocesamiento de datos, extracción y selección de características, reducciones de dimensionalidad y métricas de desempeño. Otra paquetería muy importante que se usó para extraer datos de propiedades químicas de los elementos y las estructuras fue Pymatgen [13] y Materials Project [14].

3.2. Flujo de trabajo.

Primero se hizo el minado de la base de datos que para este estudio se usó Jarvis-2D-DFT [3] que contiene información predicha de materiales bidimensionales como la energía de la banda prohibida, la energía de formación, la función dieléctrica y la energía de exfoliación de un conjunto de materiales bidimensionales. Mostrando ser una base datos coherente y completa ya que las predicciones fueron hechas utilizando DFT implementado en VASP, los materiales 2D tienen diversas estructuras geométricas con celdas unitarias tetragonales, hexagonales o monoclinicas, los funcionales usados fueron vdW-DF no locales (optB88-vdW), funcionales híbridos HSE06 (Hartree-Fock/DFT), radio de convergencia de 0.0001 eV y radio de corte de energía 500 eV.
Posterior a la recolección de datos se hizo el filtrado de los datos buscando valores nulos o incompletos en la variable objetivo que es la banda prohibida. Se filtraron los valores de banda de entre (0.5 eV, 3.0 eV). Definida la variable objetivo E_g y cada una de las instancias, en este casa cada monocapa.
Se procedió a aplicar diferentes descriptores (generación de características) para obtener el espacio de características $\vec{X}$, estos descriptores están basados en composición química del material ([16], [17]), estructura [15] y cantidades estadísticas de propiedades químicas como la masa atómica media en la celda unitaria, el número atómico medios o el radio covalente medio. Estos descriptores lo que hacen es transformar información física en vectores numéricos que pueden ser introducidos al modelo de AM. Una vez definido el espacio de características $\vec{X}$ y la variable objetivo se hizo una división de los datos en datos de entrenamiento que son los datos de los que el modelo de AM extrae el conocimiento tomando el 70 % para entrenamiento y el 30 % para evaluación, probando a su vez con una división de validación cruzada a 20 pliegues, que consiste en una selección aleatoria del porcentaje de datos de entrenamiento y de evaluación. Entonces la división de la base de datos es $(\vec{X}_{train}, y_{train})$ para datos de entrenamiento y $(\vec{X}_{test}, y_{test})$ para el conjunto de evaluación, Usando métricas de comparación específicas se compararon cada uno de los modelos utilizados

usando el coeficiente de determinación R^2 que se define como

$$R^2 = 1 - \frac{\sum_i (y_i - \acute{y_i})^2}{\sum_i (y_i - \acute{y_{mean}})^2} \tag{1}$$

que se mide entre 0 y 1, donde 1 indica un buen ajuste de la variable predicha en función de las características y 0 un mal ajuste.

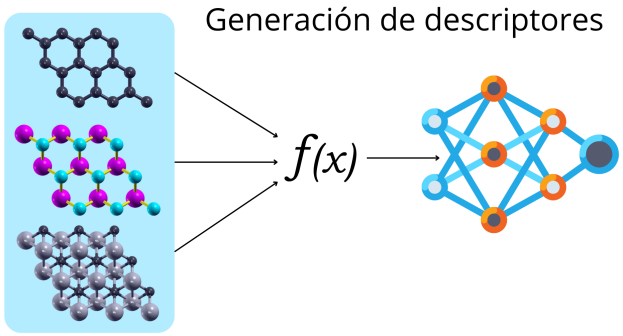

Figura 1: Representación esquemática de la generación de descriptores para cada material, el propósito es que dada alguna propiedad física se genere un vector numérico que describa propiedades estructurales y de composición.

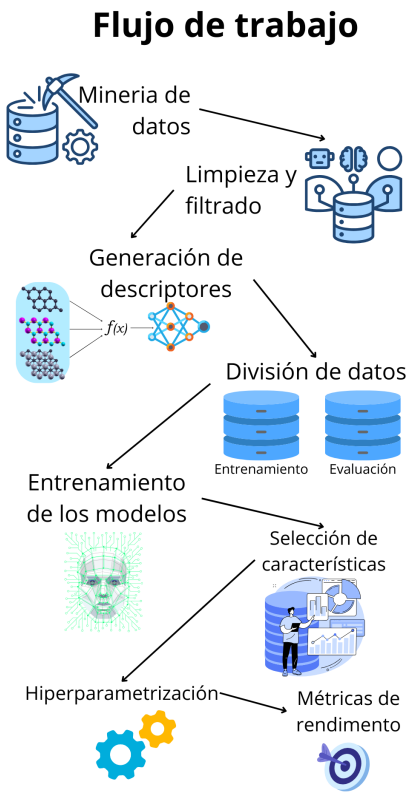

Figura 2: Descripción gráfica del flujo de trabajo.

Cuadro 1: Espacio de características

Variable	**Descripción**
E_{HOMO}	Energía del orbital molecular ocupado más alto
$\sigma^{number}_{spacegroup}$	Desviación media del número de grupo espacial
$\sigma^{position}_{columns}$	Desviación media de la posición en la columna de la tabla periódica
$n_{electron}$	número de electrones
ΔH	Entalpía de formación
$T^{Melting}_{promedio}$	Temperatura promedio del cambio de fase sólido a liquido
Ω_{Yang}	Describe el equilibrio entre la entropía de mezcla y la entalpía de mezcla de la fase líquida.

4. Resultados

4.1. Selección de características

Se generó un espacio de características de 31 dimensiones el cual se redujo a partir de la selección de características reduciendo el espacio a 7 como se muestra en el cuadro 1, a través de la técnica de información mutua entre variables 3. El mapa de correlación entre variables nos muestra que en general el espacio puede ser reducido para aquellas variables que muestran una alta correlación, valores con tendencia hacia el rojo en la figura 4. Como parte del preprocesamiento de los datos se hizo una normalización estándar de los datos de entrenamiento y prueba para el espacio de características esto quiere decir que cada a cada rato se le resto la media y se dividió entre la varianza, esto porque los modelos suelen ser sensibles a cambios abruptos en los valores de cada característica 5.

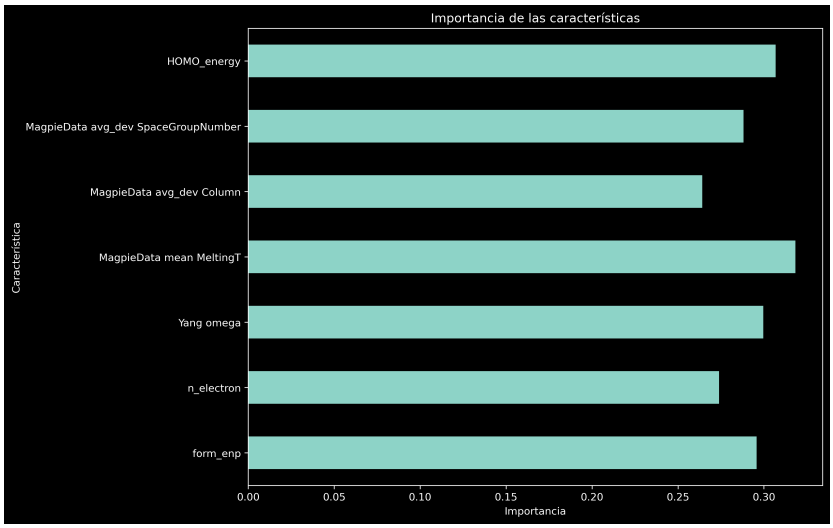

Figura 3: Importancia de las características

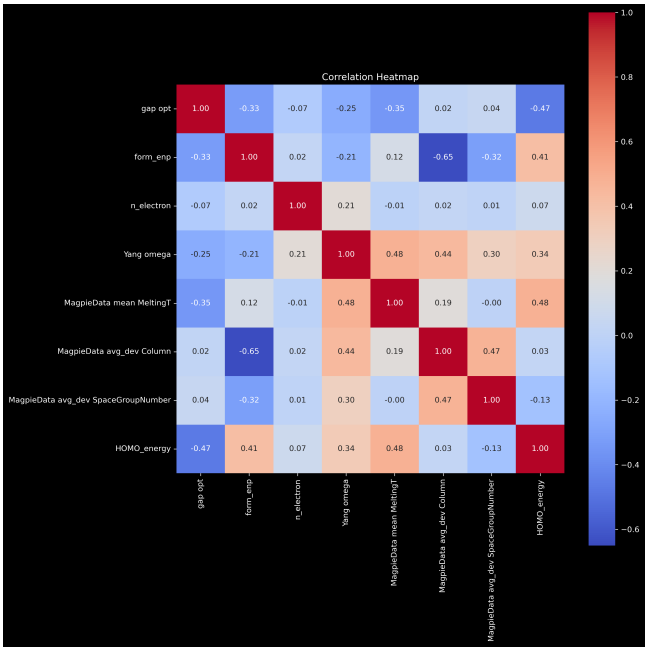

Figura 4: Mapa de correlación entre variables

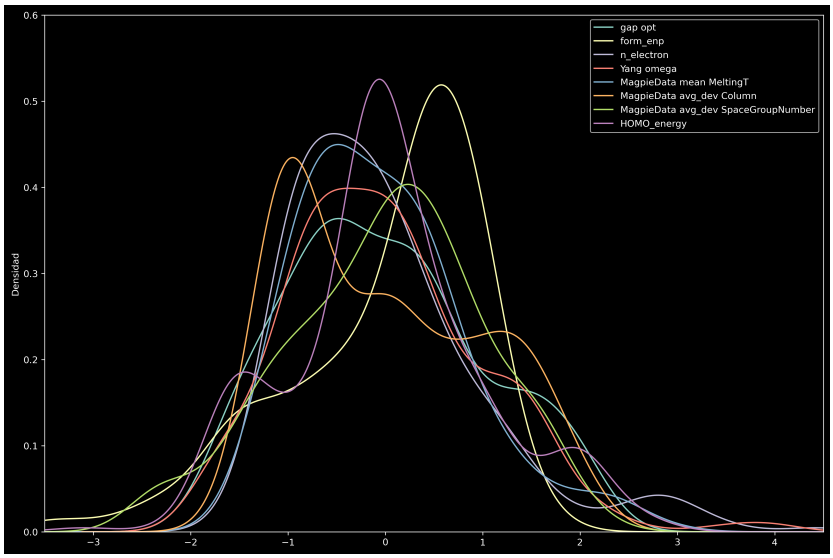

Figura 5: Gráfica de densidad para los datos normalizados]

4.2. Métricas de rendimiento

Como se mencionó se utilizó la métrica R^2 como métrica de comparación mostrando que el modelo con mejor rendimiento es el regresor de bosque aleatorio 2.

Cuadro 2: Métricas de comparación entre modelos

Modelo de AM	R^2
Lineal simple	0.36
Ridge	0.37
Lasso	0.40
Elastic Net	0.43
Regresión de bosque aleatorio	0.60
Máquina de soporte vectorial	0.40
Gradiente Boosting	0.38
Red neuronal multicapa perceptrón	0.45

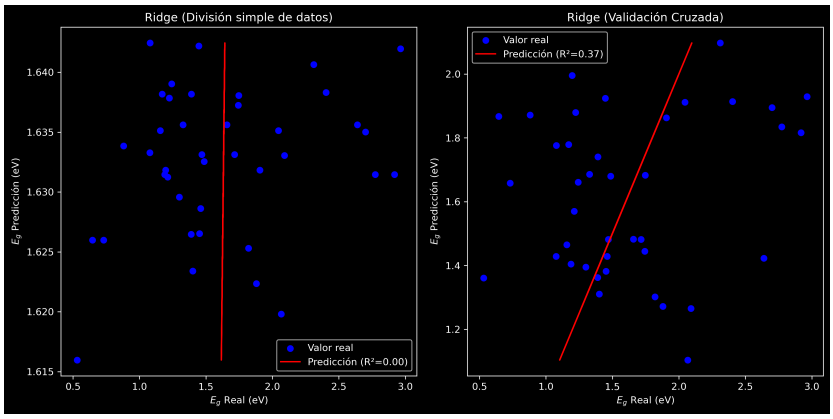

Figura 6: Gráfica de comparación del modelo lineal ridge

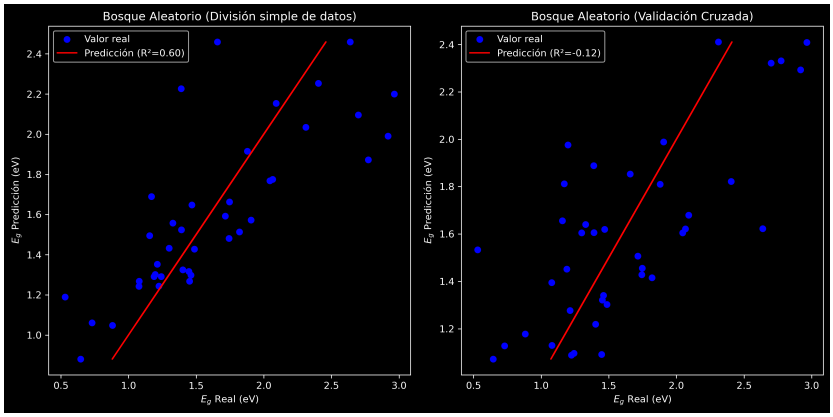

Figura 7: Gráfica de comparación del modelo regresor de bosque aleatorio

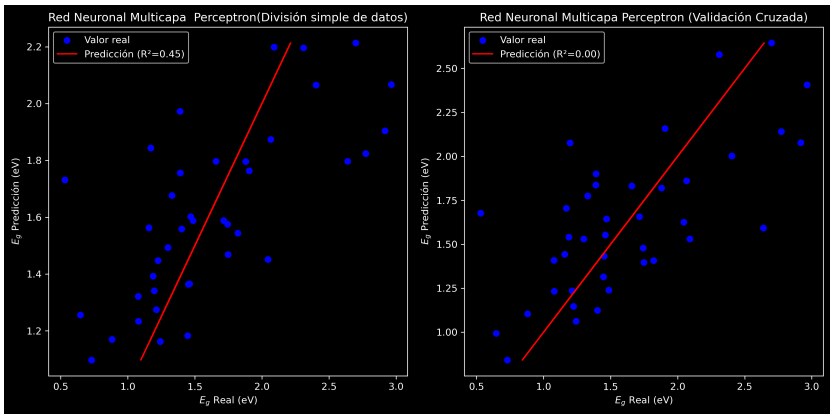

Figura 8: Gráfica de comparación del modelo red neuronal artifical (perceptron)

5. Conclusiones

Mis resultados muestran que trabajar con un espacio de características que no dependa de variables extraídas de cálculos de primeros principios muestra un bajo rendimiento de acuerdo al coeficiente de determinación R^2 mostrando que de los modelos lineales el Ridge tiene un rendimiento de 0.48 y el modelo que mejor resultados muestra es el regresor de bosque aleatorio con un R^2 de 0.5. Es necesario aplicar una reducción de dimensionalidad o como PCA o TSnet para transformar los datos en otro espacio que maiximice el rendimiento de los modelos. Los modelos de AM muestran ser eficientes en el tiempo de cálculo pero aún se tiene que buscar un espacio de características que no dependa de cálculo de primeros principios. A pesar del bajo rendimiento mostrado hasta ahora aún no se ha exprimido todo el potencial de los modelos de AM quedando como trabajo a futuro buscar las mejores combinaciones del espacio de características para predecir la banda prohibida. Además los modelos de AM pueden considerarse cajas negras ya que no dan una explicación fenomenológica de los problemas a tratar.

Referencias

[1] Ohno, K., Esfarjani, K., and Kawazoe, Y. (2018). Computational Materials Science. En Springer eBooks.

[2] Malakar, P., Thakur, S. H., Nahid, S. M., and Islam, M. M. (2022). Data-Driven Machine Learning to Predict Mechanical Properties of Monolayer Transition-Metal Dichalcogenides for Applications in Flexible Electronics. ACS Applied Nano Materials, 5(11), 16489-16499. https://doi.org/10.1021/acsanm.2c03564

[3] Choudhary, K., Kalish, I., Beams, R. et al., High-throughput Identification and Characterization of Two-dimensional Materials using Density functional theory. Sci Rep 7, 5179 (2017).

[4] Ward et al., Matminer: An open source toolkit for materials data mining. Comput. Mater. Sci. 152, 60-69 (2018)

[5] Rajan, A. C., Mishra, A., Satsangi, S., Vaish, R., Mizuseki, H., Lee, K., and Singh, A. K. (2018). Machine-Learning-Assisted accurate band gap predictions of functionalized MXENE. Chemistry of Materials, 30(12), 4031–4038.

[6] VanderPlas, J. (2023). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.

[7] Olsthoorn, B., Geilhufe, R. M., Borysov, S. S., and Balatsky, A. V. (2019). Band Gap Prediction for Large Organic Crystal Structures with Machine Learning. Advanced Quantum Technologies, 2(7-8). https://doi.org/10.1002/qute.201900023

[8] Malakar, P., Thakur, S. H., Nahid, S. M., and Islam, M. M. (2022). Data-Driven Machine Learning to Predict Mechanical Properties of Monolayer Transition-Metal Dichalcogenides for Applications in Flexible Electronics. ACS Applied Nano Materials, 5(11), 16489-16499. https://doi.org/10.1021/acsanm.2c03564

[9] Zhang, Y., Xu, W., Liu, G., Zhang, Z., Zhu, J., and Li, M. (2021). Bandgap prediction of two-dimensional materials using machine learning. PLOS ONE, 16(8), e0255637. https://doi.org/10.1371/journal.pone.0255637

[10] Claussen, N., Bernevig, B. A., and Regnault, N. (2020). Detection of topological materials with machine learning. Physical Review. B, 101(24). https://doi.org/10.1103/physrevb.101.245117

[11] Martinetto, V., Shah, K., Cangi, A., and Pribram Jones, A. (2024). Inverting the Kohn-Sham equations with physics-informed machine learning. Machine Learning: Science And Technology. https://doi.org/10.1088/2632-2153/ad3159

[12] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay; 12(85):28252830, 2011.

[13] Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Mater., 2013, 1, 11002, doi:10.1063/1.4812323.

[14] Ong, S. P.; Cholia, S.; Jain, A.; Brafman, M.; Gunter, D.; Ceder, G.; Persson, K. a. The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles, Comput. Mater. Sci., 2015, 97, 209–215, doi:10.1016/j.commatsci.2014.10.037.

[15] Ward, L., Agrawal, A., Choudhary, A. et al. A general-purpose machine learning framework for predicting properties of inorganic materials. npj Comput Mater 2, 16028 (2016). https://doi.org/10.1038/npjcompumats.2016.28

[16] M. A. Butler and D. S. Ginley 1978 J. Electrochem. Soc. 125 228 DOI 10.1149/1.2131419

[17] Butler, M. A., and Ginley, D. S. (1978). Prediction of Flatband Potentials at Semiconductor-Electrolyte Interfaces from Atomic Electronegativities. Journal Of The Electrochemical Society, 125(2), 228-232. https://doi.org/10.1149/1.2131419

[18] Russell, S., and Norvig, P. (2016). Artificial intelligence: A Modern Approach, Global Edition.

[19] Murphy, K. P. (2012). Machine learning: A Probabilistic Perspective. MIT Press.

[20] Rajan, K. (2005). Materials informatics. Materials Today, 8(10), 38-45. https://doi.org/10.1016/s1369-7021(05)71123-8