

# Predicción de propiedades ópticas no-lineales a partir de la DFT y el Machine Learning

Prediction of non-linear optical properties from DFT and Machine Learning

C. A. Pacheco-González<sup>1</sup>, J. Hernández-Varela<sup>1</sup>, C. A. Medina-Jiménez<sup>1</sup>, M. Malagón-Alvarez<sup>1</sup>, H. Salazar-Pérez<sup>1</sup>, X. Jiménez-Gutiérrez<sup>1</sup>, M. Á. Vázquez-Olguín<sup>1</sup>, J.E Castellanos-Águila<sup>1</sup>

<sup>1</sup>División de ingenierías Campus Irapuato-Salamanca, Universidad de Guanajuato, Carr. Salamanca-Valle Km 3.5+1.8, Comunidad de Palo Blanco, 36700 Salamanca, Gto., México.  
ca.pachecogonzalez@ugto.mx<sup>\*</sup>

## Resumen

El diseño de las condiciones experimentales requiere considerar diversos parámetros relacionados con el campo de interés. Los cálculos a primeros principios o ab-initio suelen ser computacionalmente costosos, aumentando exponencialmente conforme incrementa el número de átomos, variables de cálculo y precisión de los resultados. Por este motivo, mediante el uso del Machine Learning y de las propiedades estructurales, electrónicas y ópticas de un líquido iónico de referencia, se propusieron tres modelos que permiten determinar la energía de interacción en función de la distancia en que interactúan sus iones constituyentes. Los resultados obtenidos, se pudo comprobar que los modelos basados en regresiones polinómicas son efectivos para predecir de manera coherentemente la propiedad bajo estudio, pero a un menor costo computacional.

**Palabras clave:** Machine Learning; Regresiones Lineales; DFT; Energía de Interacción.

## Introducción

Cuando se diseñan las condiciones experimentales, se deben tener en cuenta varios parámetros relacionados con el área de estudio. Desde un punto de vista práctico, las propiedades a evaluar suelen resultar de la simple adición de sus componentes. Este enfoque ayuda a tomar decisiones sobre la viabilidad del experimento a desarrollar. Sin embargo, se omiten las variaciones provocadas por efectos fácilmente controlables en el experimento, como los son los diferentes tipos de solventes o campos externos aplicados, lo que, dependiendo de la especificidad de la propiedad, se requiere para tener un mayor control de la respuesta a las condiciones del experimento.

Los cálculos teóricos desde primeros principios (o ab initio) han mejorado la predicción de las variables a considerar a la hora de diseñar el experimento. En este contexto, en el campo de la óptica no lineal, Rodríguez et al. presentaron un mapa de susceptibilidad no lineal de segundo y tercer orden de casi 1500 combinaciones de iones para diseñar líquidos iónicos [1]. Sin embargo, la precisión de los resultados teóricos con respecto a los datos experimentales es computacionalmente costosa, aumentando exponencialmente a medida que aumenta el número de átomos o variables de cálculo.

En los últimos años, el desarrollo y uso del Machine Learning (ML) ha apoyado a los cálculos ab initio mediante la aproximación de funciones de densidad [2]. En otros casos, ha predicho, optimizado y acelerado las mismas propiedades buscadas, pero a un menor coste computacional [3, 4]. Por lo tanto, en la búsqueda de incursionar en el campo del ML, partimos de un trabajo teórico previo [5] en el que se estudió la dependencia intermolecular de las propiedades ópticas no lineales del líquido molecular Tetrafluoroborato de 1-Butil-3-Metilimidazolio (BMIM+...BF4-).

A partir de las propiedades estructurales, electrónicas y ópticas, en condiciones de cálculo específicas, se obtuvieron modelos predictivos altamente eficientes para determinar el comportamiento de estas propiedades en condiciones no calculadas.

## Metodología

Para llevar a cabo nuestro estudio, partimos del modelo basado en el potencial de Lennard-Jones (VLJ), el cual se usa principalmente para comprender la interacción molecular mediante la relación entre la distancia y energía de interacción. Desde primeros principios, su obtención es computacionalmente costosa, limitando su generalización en diferentes condiciones experimentales. Por ello, se busca encontrar un método más general, basado en un modelo matemático, que pueda ajustarse con otros parámetros y sea aplicable en el campo del ML.

Por lo tanto, en este trabajo de investigación se propusieron tres métodos para determinar la energía de interacción del sistema molecular bajo estudio, el primer método se basó en la expresión del VLJ; el segundo de una interpolación polinómica de grado  $n$ , y el tercer modelo mediante la interpolación polinómica mejorada. Dado a que uno de los parámetros considerados en un trabajo previo de referencia [5], fue el solvente (representado por su valor de constante dieléctrica,  $\epsilon$ ), la energía de interacción se aproximó para los diferentes solventes considerados.

A continuación, se presentan los tres métodos evaluados en este trabajo de investigación:

### 1. Modelo de Basado en el Potencial de Lennard-Jones.

A partir de los datos de referencia, se obtuvieron los parámetros que definen el potencial de Lennard-Jones,  $\epsilon$ ,  $\sigma$  y  $r_0$ , los cuales son dependientes de la energía y la distancia de interacción, tal y como se puede observar en la ecuación 1.

$$V_{LJ} = 4e \left[ \left( \frac{s}{x} \right)^{12} - \left( \frac{s}{x} \right)^6 \right]. \quad (1)$$

Mediante la representación gráfica se puede observar que  $\epsilon$  es el punto más bajo, y  $r_0$  es el valor de la distancia de interacción donde se tiene  $\epsilon$ , esto permite obtener  $\sigma$  a partir de un despeje de la ecuación de Lennard Jones.

$$s = \frac{r_0}{\sqrt[6]{4e}}. \quad (2)$$

Para todas las constantes dieléctricas, se obtuvieron las constantes  $\epsilon$  y  $\sigma$ , para posteriormente generar una gráfica y a partir de ellas se realizó una regresión a cada una de las constantes para obtener una expresión que describa el comportamiento de estas, permitiendo obtener su comportamiento en diferentes valores de constante dieléctrica sin haber realizado los cálculos desde primeros principios.

### 2. Modelo Basado en una regresión polinómica

El segundo modelo propone un método basado en regresiones polinómicas para cada solvente, es decir, queremos aproximar un polinomio de la forma:

$$f(x) = c_0 + c_1x + c_2x^2 + \dots + c_nx^n, \quad (3)$$

para cada solvente, donde  $n$  es el grado del polinomio y  $c_i$  son los coeficientes de este [6].

Como primer paso, se debe determinar el grado del polinomio que mejor aproxime los valores para cada solvente. Esto se logra calculando el error cuadrático medio:

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (\hat{f}_i(x) - f_i(x))^2}{m}} \quad (4)$$

para cada polinomio desde grado 1 hasta grado  $n$ , siendo el mejor aquel que minimice el error y evite un desajuste excesivo, donde  $m$  representa el número de observaciones hechas.

Después, considerando todos los solventes, se elige el mismo grado del polinomio para todos. Esto permite aproximar cada constante  $c_i$  de los polinomios asociados a cada solvente realizando un modelo de optimización para cada constante  $c_1, \dots, c_n$  teniendo  $n + 1$  modelos, donde  $n$  representa el grado del polinomio elegido.

Por último, para validar el modelo creado se aplica la validación cruzada k-fold calculando el error cuadrático medio (4) sobre los datos originales con los datos predichos con el modelo creado. Este método repite el proceso de aprendizaje k veces con diferentes conjuntos de datos de entrenamiento y validación.

### 3. Modelo Basado en una Interpolación Polinómica Mejorado

Se diseñó un modelo que utiliza como base la interpolación polinómica, además de añadir un nuevo algoritmo de aprendizaje automático conocido como Gradient Boosting (GBM, por sus siglas en inglés), buscando un mejor desempeño en las predicciones. GBM permite obtener una mayor adaptación del Modelo Polinómico debido a que es un modelo no paramétrico, es decir, no sigue lo impuesto por la regresión polinómica, sino que está construido directamente a partir de los datos iniciales.

El aumento de Gradiente o Gradient Boosting es un algoritmo de aprendizaje automático en conjunto. Es decir, en lugar de contar con un solo modelo, podemos trabajar con distintos de manera secuencial, corrigiendo los errores de modelos anteriores, a esta técnica se le conoce como Boosting. El principal diferenciador de Gradient Boosting radica en que cada modelo se entrena con el error residual del modelo anterior. El error residual es la diferencia entre los datos observados y los datos predichos.

GBM utiliza como principal técnica los bosques aleatorios (random forest), el cual combina la salida de múltiples árboles de decisiones para alcanzar un solo resultado, esta técnica fue registrada por Leo Breiman en 2001 [7].

De manera común gradient Boosting inicia por el modelo más simple, el cual es una constante, esta constante es elegida de manera que minimice la función de pérdida, la cual en este caso es representada por la suma de los cuadrados de los residuos (error cuadrático medio). Para el modelo se implementó la formulación basada en gradientes descendientes de métodos de refuerzo.

Para nuestro modelo, los datos de  $x$  (observaciones) correspondientes a la distancia de interacción, e  $y$ , (características objetivo) corresponden a la energía de interacción.

Cuando el modelo inicializa con una constante:

$$F_0(x) = \operatorname{argmin}_{\gamma} \sum_{i=1}^n L(y_i, \gamma). \quad (5)$$

La ecuación (5) indica el argumento mínimo de la sumatoria de los residuos  $y$ , como resultado, nuestra nueva predicción será un promedio de todos los valores  $y$  y del primer ciclo. Para minimizar la suma tenemos que encontrar el valor de  $\gamma$  para cada dato, teniendo en cuenta la predicción anterior y las nuevas muestras propuestas, por lo que se calcula el valor de  $\gamma$  para cada relación  $x$  e  $y$  por separado,

$$\gamma_{jm} = \operatorname{argmin}_{\gamma} \sum_{x_i \in R_{ij}} L(y_i, F_{m-1}(x_i) + \gamma). \quad (6)$$

La actualización de predicciones mediante un modelo nuevo se realiza mediante la siguiente ecuación:

$$F_m(x) = F_{m-1}(x) + v \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm}). \quad (7)$$

En la ecuación anterior, para cada muestra, sumamos el valor de predicción anterior y los valores de  $\gamma$  que determinamos para cada muestra. El valor de  $\gamma$  es esencial para evitar un sobreajuste, de esta manera podemos repetir los pasos M veces.

La predicción de solventes requiere conocer los coeficientes que determinan a su respectivo polinomio de grado 8. Aquí determinamos que para la predicción de estos coeficientes utilizaremos dos técnicas de regresión distintas una de ellas aplicará a los coeficientes pares y otra a los impares.

Para determinar los coeficientes pares, observamos que el comportamiento de ellos tiene una similitud al comportamiento de descarga perteneciente a un capacitor, por lo cual definimos su función como:

$$\beta_{par} = V_0 * e^{-(\alpha * constante\ dielectrica)}. \quad (9)$$

De la ecuación podemos determinar una nueva para saber el valor de  $\alpha$ :

$$\alpha = - \frac{\log\left(\frac{\beta_{par}}{V_0}\right)}{Constante\ dielectrica}. \quad (10)$$

Por otro lado, se aplicó una función exponencial para la determinación de coeficientes impares, donde  $c_1$ ,  $c_2$  y  $c_3$  corresponden a valores ajustados al solvente.

$$\beta_{impar} = c_1(1 - \exp(-c_2)) + c_3. \quad (11)$$

## Resultados y Discusión

Modelo basado en el potencial de Lennard-Jones

A partir del modelo del potencial de Lennard-Jones, se representó el cambio de las constantes que definen el VLJ dependiendo de la constante dieléctrica (Figura 1). Debido a que cada constante dieléctrica tiene su ecuación correspondiente que está definida por una energía y distancia de interacción, se realizó una regresión a cada una de ellas para obtener una ecuación general que determine  $\epsilon$  y  $\sigma$ , a partir de  $C_{die}$ . Esta ecuación permitió predecir estos mismos parámetros mediante el uso de ML.

La ecuación obtenida es de forma exponencial y se presenta a continuación:

$$y = A - (A - B) * e^{B_0 * (C_{die} - C_{die0})}, \quad (12)$$

en donde A es el valor más alto de  $\epsilon$ , B es el primer valor de los datos en  $\epsilon$ ,  $C_{die0}$  es el primer valor de los datos de la constante dieléctrica,  $C_{die}$  es la variable independiente,  $B_0$  es la constante la cual se le aplicarán los métodos de regresiones (una para  $\epsilon$  y otra para  $\sigma$ ) e Y es la variable dependiente, la cual en nuestro caso corresponde a la energía de interacción.

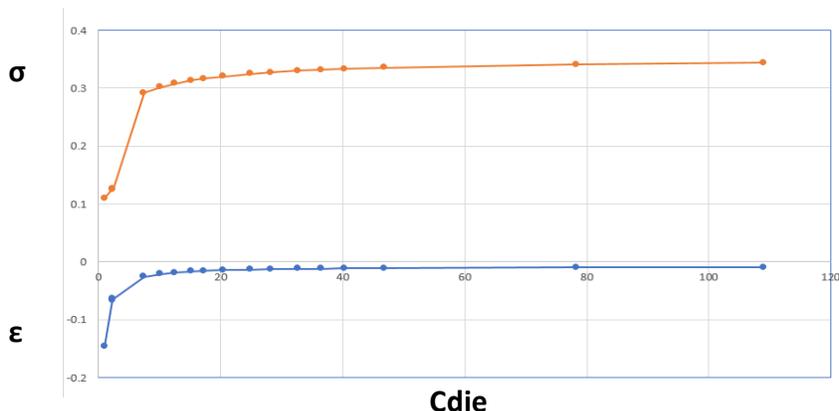


Figura 1. Cambios de  $\epsilon$  y  $\sigma$  derivados de la ecuación del potencial de Lennard-Jones a partir de los datos obtenidos para cada solvente estudiado.

A partir de la ecuación 12, se hizo un ajuste con una regresión exponencial para determinar  $B_0$ , dependiente del valor de la constante dieléctrica (Fig. 2a). Es importante remarcar que los puntos rojos corresponden a los datos empíricos. En la Figura 2b) se presenta el correspondiente error cuadrático para cada  $B_0$ .

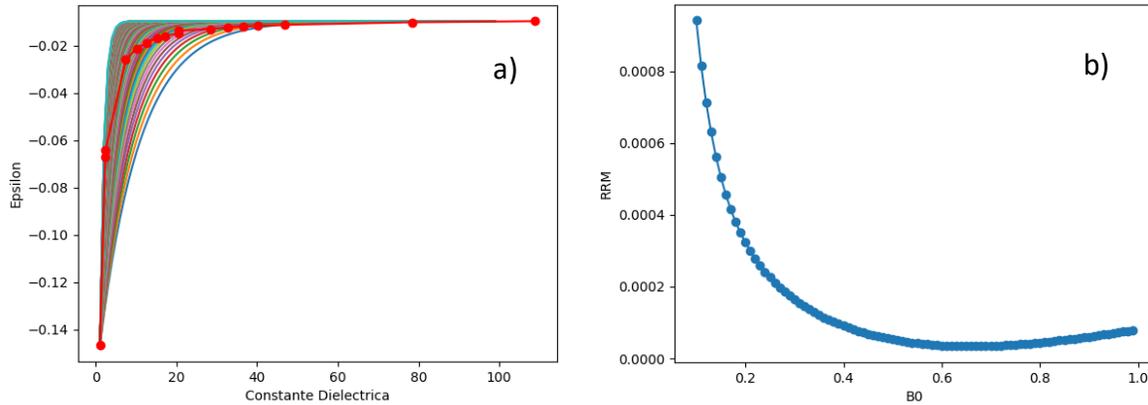


Figura 2. a) Ajuste exponencial para  $\epsilon$  y b) Error cuadrático (RRM) para cada  $B_0$  propuesto.

Posteriormente, a partir del método de paro por gradiente se optimizaron los valores obtenidos. Esto permitirá compararlos con los datos empíricos. De acuerdo con nuestros resultados el valor de  $B_0$  dependiente de  $\epsilon$ , que menor error cuadrático presenta es de 0.66. De igual forma se determinó  $B_0$  dependiente de  $\sigma$ , siendo en este caso igual a 0.16. Estos valores se usarán para predecir la energía de interacción a diferentes distancias y solventes ( $C_{die}$ ).

Para ello se desarrolló un código basado en dos funciones a partir de la ecuación 12, obteniéndose la siguiente forma, la cual describe adecuadamente los comportamientos obtenidos desde primeros principios, pero a un coste computacional mucho menor.

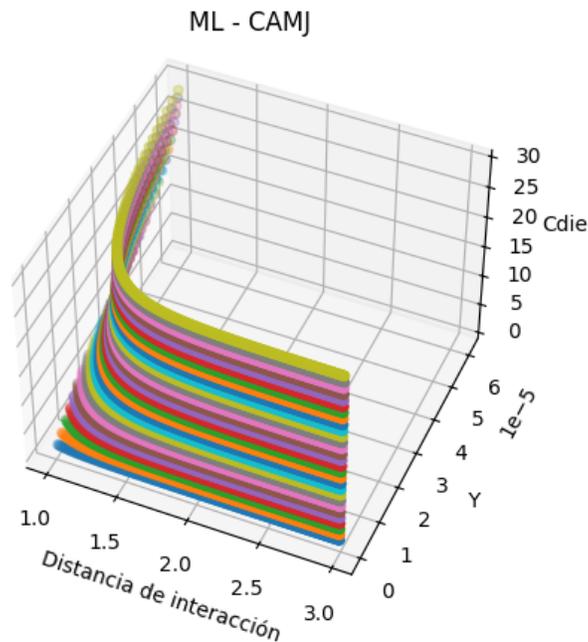


Figura 3. Representación de modelo desarrollado a partir de las ecuaciones de distancia de interacción contra energía de interacción, pero con un diferente valor de constante dieléctrica (solventes), las cuales van de 0 a 30 en intervalos de 1.

### Modelo Basado en una interpolación polinómica

Para implementar el modelo basado en una interpolación polinómica, se requiere definir el grado del polinomio de (1), el cual fue de grado 8 ya que, mediante una evaluación visual, se determina que se ajusta mejor a los datos (Fig. 4). Si bien los polinomios de grado superior presentan menor error cuadrático medio, se observa un desajuste excesivo en los últimos datos. En la Tabla 1 se muestran los coeficientes obtenidos para el polinomio de grado 8.

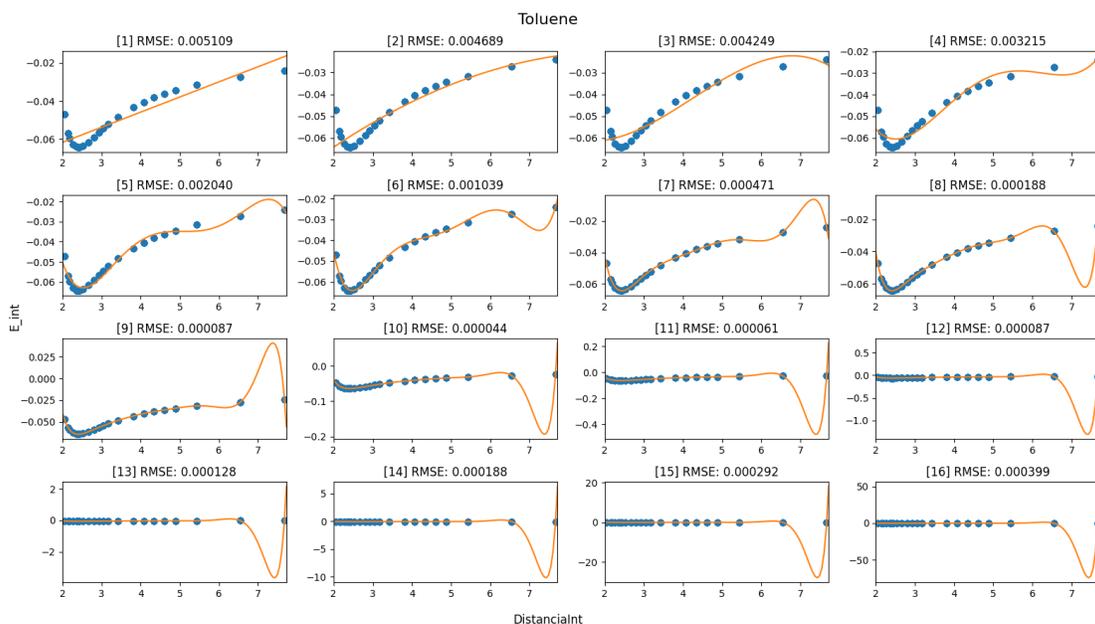


Figura 4. Ajustes de curvas para cada uno de los coeficientes de los polinomios de grado 8.

Tabla 1. Coeficientes de los polinomios de grado 8 para cada solvente.

Solvente	c0	c1	c2	c3	c4	c5	c6	c7	c8
Benzene	13.059296	-25.809479	21.754791	-10.281328	2.981916	-0.543436	0.060758	-0.003809	0.000103
Toluene	12.847191	-25.367803	21.365445	-10.088922	2.923602	-0.532346	0.059467	-0.003725	0.000100
THF	10.326816	-20.126842	16.766645	-7.825846	2.240006	-0.402664	0.044391	-0.002744	0.000073
DCE	10.163761	-19.754480	16.419050	-7.646180	2.183537	-0.391602	0.043072	-0.002656	0.000070
1-Hexanol	10.119579	-19.634639	16.296082	-7.578215	2.161080	-0.387032	0.042510	-0.002618	0.000069
1-Pentanol	10.038184	-19.456420	16.133673	-7.495672	2.135454	-0.382058	0.041921	-0.002579	0.000068
MEt	9.969674	-19.309775	16.003019	-7.431052	2.115979	-0.378391	0.041499	-0.002552	0.000067
Acetone	9.948565	-19.253599	15.945130	-7.398601	2.105069	-0.376131	0.041217	-0.002533	0.000067
Ethanol	9.722381	-18.789971	15.542615	-7.203883	2.047573	-0.365511	0.040017	-0.002457	0.000065
NitroEthane	9.917713	-19.169913	15.859500	-7.351454	2.089557	-0.372988	0.040832	-0.002506	0.000066
Methanol	9.981811	-19.301064	15.973585	-7.405980	2.105221	-0.375761	0.041128	-0.002524	0.000066
NM	9.862436	-19.050700	15.752396	-7.297877	2.073200	-0.369866	0.040468	-0.002483	0.000065
EthaneDiol	9.935449	-19.193011	15.869256	-7.350397	2.087327	-0.372195	0.040698	-0.002495	0.000066
DMSO	9.861203	-19.038516	15.735717	-7.287097	2.069252	-0.369001	0.040356	-0.002475	0.000065
Water	9.810662	-18.931843	15.641950	-7.241064	2.055415	-0.366393	0.040055	-0.002455	0.000065
Formamide	9.776029	-18.858250	15.576125	-7.208187	2.045374	-0.364473	0.039831	-0.002441	0.000064

Teniendo todos los coeficientes se crearon 9 modelos que ajustan a estos. Los coeficientes pares se modelaron con la ecuación:

$$f_1(solv) = c_1 \exp(c_2 solv) + c_3, \quad (13)$$

mientras que los coeficientes impares se modelaron con la ecuación

$$f_2(solv) = c_1 \ln(c_2 solv) + c_3, \quad (14)$$

resultando en los ajustes de curvas mostrados en la Figura 4.

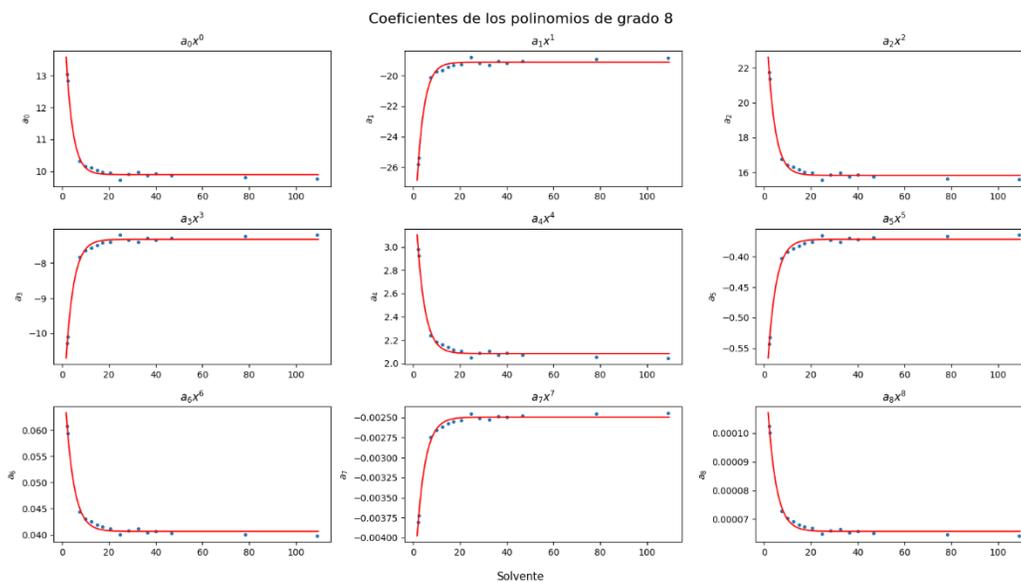


Figura 5. Ajustes de curvas para cada uno de los coeficientes de los polinomios de grado 8.

Por último, queda validar el modelo, para ello se aplicó k-fold con 5 pliegos sobre todo el conjunto de datos. Esto permite tener un aproximado del 80% de los datos para entrenamiento y 20% para validación en cada pliegue.

Aplicando el error cuadrático medio (Eq. 2) se obtuvieron los siguientes resultados:

**Tabla 2.** Errores medios de los pliegos obtenidos.

	Pliego 1	Pliego 2	Pliego 3	Pliego 4	Pliego 5	Media
RMSE	0.00128553	0.00222387	0.00149843	0.00140991	0.00145029	0.00157361

Se puede observar un error bajo congruente con la escala de la energía de interacción sobre cada uno de los pliegos. De modo que solo queda entrenar el modelo sobre todo el conjunto de datos, dando el resultado mostrado en la Figura 5.

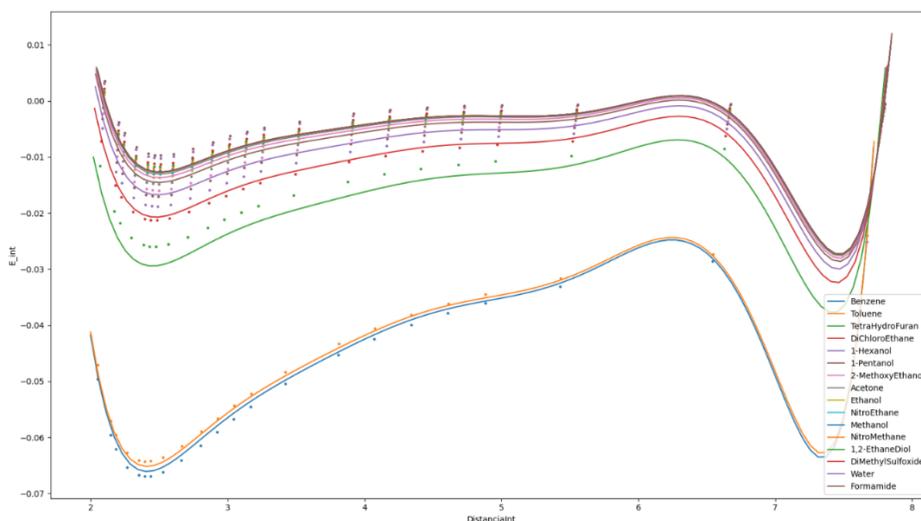


Figura 6. Predicciones del modelo creado sobre todo el conjunto de datos.

#### Modelo Basado en una Interpolación Polinómica Mejorada

De la misma manera que el modelo anterior, el enfoque de la predicción contiene en conjunto la energía de interacción y la distancia de interacción en relación con la constante dieléctrica asociada a los solventes.

Tomando como base, los resultados obtenidos en el modelo anterior, incorporamos un polinomio de grado 8 a nuestro modelo, además de agregar GBM al modelo, de manera que, según la metodología propuesta, la predicción final tome en cuenta no solo al polinomio de grado 8, sino que tome en cuenta la aproximación obtenida mediante GBM, para obtener una predicción más cercana a lo esperado por la teoría.

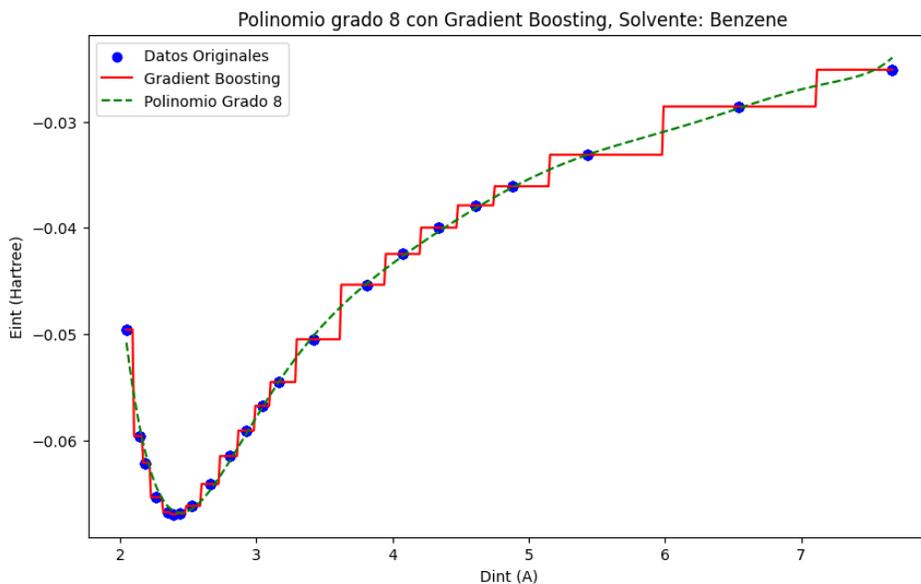


Figura 7. Modelo de Interpolación Polinómica con GBM, Benceno.

Como se puede observar en la imagen anterior, los resultados obtenidos indica una mejora en las aproximaciones que puedan ser predichas, ya que al tener una referencia adicional como GBM, el polinomio de grado 8 tiene un comportamiento con mayor aproximación a la tendencia prevista por los datos originales.

Posterior a la obtención del resultado que parece ser positivo, procedemos a calcular los coeficientes expresados por la formula (1), para cada solvente. De esta manera tenemos una idea más concreta de los que podemos esperar en la predicción de resultados.

Dado que podemos determinar que el comportamiento de la relación existente entre la energía de interacción y la distancia de interacción en relación con la constante dieléctrica asociada a los solventes puede ser expresada mediante un polinomio de grado 8, para la obtención de predicciones asociadas a nuevos solventes podemos, como opción más viable, determinar mediante la predicción de nuevos coeficientes para un polinomio de grado 8, el comportamiento de un nuevo solvente.

De esta manera, el siguiente paso este dado por la creación de un algoritmo que determine nuevos coeficientes dependiendo únicamente del valor del solvente. Esto nos fue posible determinarlo mediante las dos técnicas anteriormente descritas, las cuales responden a un comportamiento similar a la función de descarga para un capacitor y una función exponencial.

Además, en este nuevo algoritmo también cuenta con la inclusión de GBM para minimizar el error en la aproximación de cada coeficiente.

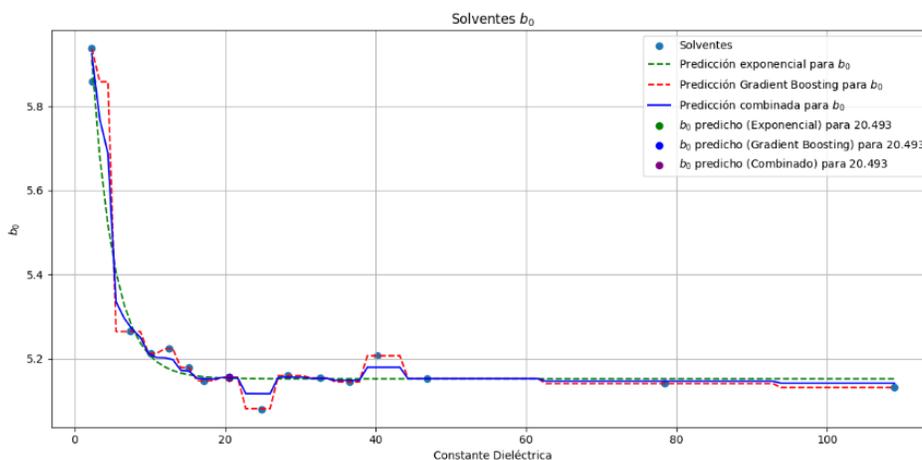


Figura 8. Coeficiente  $B_0$  en relación con la constante dieléctrica de cada solvente.

Como podemos observar, dentro de nuestra gráfica, la inclusión de ambas técnicas de aprendizaje automático entrega un modelo de mayor precisión que la obtenida en ambas por separado.

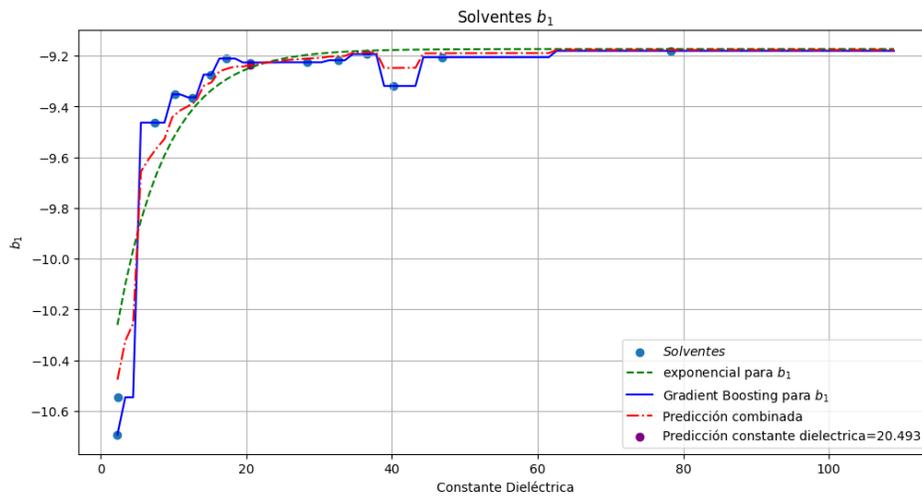


Figura 9. Coeficiente B1 en relación con la constante dieléctrica de cada solvente.

De la misma manera, ahora aplicando una exponencial, podemos observar que la combinación de ambos métodos podría otorgarnos resultados más coherentes en base a los datos originales.

Ambos métodos funcionan de manera separada para cada uno de los coeficientes, debido a la naturaleza del comportamiento en los solventes, el modelo aplicado para  $\beta_0$  tendrá un comportamiento similar a los coeficiente pares del solvente y de igual manera, el modelo aplicado en  $\beta_1$  tendrá un comportamiento similar a los coeficiente impares del solvente, por lo que solo intercalando funciones, podemos obtener una predicción de coeficientes para un nuevo solvente.

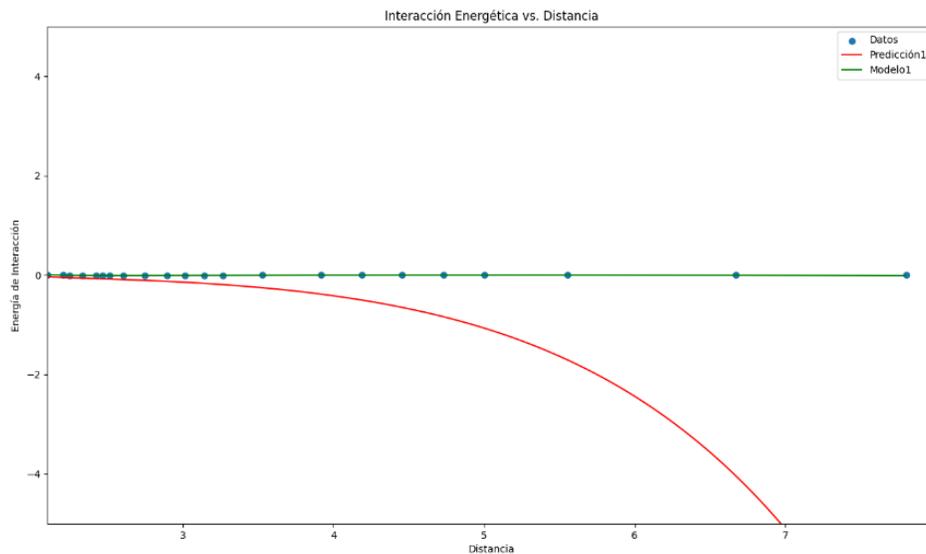


Figura 10. Comparación de Modelo de Datos Originales y Modelo Predictivo.

Como podemos observar en la figura anterior, el modelo 1, corresponde a la representación del polinomio de grado 8 para un solvente específico, en este caso "Formamida", Ahora utilizando nuestro modelo predictivo, determinamos los coeficientes para el mismo solvente, esto con la intención de conocer el error existente en el modelo, dicho polinomio está representado por "Predicción 1".

Los resultados obtenidos son inesperados y comparando los polinomios obtenidos, el comportamiento del polinomio predicho está completamente alejado de lo esperado, aunque al analizar los coeficientes, estos no

muestran cambios significativos entre sí, la representación si refleja un comportamiento erróneo de nuestro modelo predictivo.

Uno de los posibles motivos por los cuales el modelo predictivo con GBM muestra este comportamiento, puede ser por la misma inclusión de este algoritmo a nuestro modelo. Mas apreciable, en la obtención de coeficientes, ya que, al observar los coeficientes de cada solvente, podemos observar que ciertos datos en específico muestran una dispersión con respecto a la tendencia general de los solventes. Esta dispersión proviene del modelo matemático con el que se obtuvieron los datos originales y que podría no ser correcto tomarla en cuenta, algo que, si realiza nuestro modelo con GBM, de esta manera, al estar tomando en cuenta la dispersión de datos concretos, podemos tener predicciones erróneas basadas en la dispersión.

## Conclusiones

De acuerdo con los resultados obtenidos del Machine Learning, los modelos propuestos a partir de resultados ab-initio, predecen eficientemente y reduciendo el costo computacional. Los modelos propuestos, Lennard-Jones e interpolación polinómicas, demuestran ser efectivos en la estimación de la energía de interacción en función de la constante dieléctrica del solvente. A pesar de los buenos resultados, se observó que la técnica del Gradient Boosting puede introducir dispersión a mayores distancias de interacción. Finalmente, el modelo basado en la definición del potencial de Lennard-Jones, presenta los mejores ajustes con respecto a los datos de referencia. Esto refuerza la necesidad de corregir las dispersiones presentadas en los modelos de interpolaciones ya sea mediante mayor cantidad de datos o con las formas funcionales de ajuste.

## Bibliografía/Referencias

- [1] P. Díaz-Rodríguez, J. C. Cancilla, N. V. Plechkova, G. Matute, K. R. Seddon, and J. S. Torrecilla, "Estimation of the refractive indices of imidazolium-based ionic liquids using their polarisability values," *Phys. Chem. Chem. Phys.*, vol. 16, pp. 128–134, 1 2014.
- [2] John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke, Finding Density Functionals with Machine Learning, *Phys. Rev. Lett.* 108, 253002 (2012).
- [3] R. Nagai, R. Akashi, O. Sugino, Completing density functional theory by machine learning hidden messages from molecules. *npj Comput Mater* 6, 43 (2020).
- [4] V. L. Deringer, M. A. Caro, G. Csányi, Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Adv. Mater.* 2019, 31, 1902765.
- [5] H. Romero-Torres, M. Malagón-Álvarez, C. A. Pacheco-González, Gutiérrez-García, I. M., Balanzar-Urzua, I. E., Martínez-Vega, J., ... Castellanos-Águila, J. E. (2023). Estudio de la dependencia intermolecular en las propiedades ópticas en líquidos moleculares. *JÓVENES EN LA CIENCIA*, 21, 1–6.
- [6] Natekin A y Knoll A (2013) Gradient boosting machines, un tutorial. *Frente. Neurobot.* Isaías 7:21. doi: 10.3389/fnbot.2013.00021\
- [7] Muthukrishnan. (2018, June 17). Maths behind Polynomial regression (Muthukrishnan, Ed.) [Review of Maths behind Polynomial regression]. <https://muthu.co/maths-behind-polynomial-regression/>
- [8] Gaussian 09, Revision D.01, M. J. Frisch, et. al., Gaussian, Inc., Wallingford CT, 2016.