

Universidad Peruana Cayetano Heredia



FACULTAD DE CIENCIAS Y FILOSOFÍA

**Comparación de modelos de aprendizaje automático  
químicamente interpretativos en la clasificación de  
moléculas de colorantes naturales de celdas solares  
sensibilizadas por tintes según relaciones  
estructura-propiedad**

*Proyecto de tesis*

Autor:

Bach. Dan Santivañez Gutarra

Asesora:

Dra. María Quintana Caceda

Lima - Perú

2022

# Índice

<b>1. Introducción</b>	<b>3</b>
1.1. Problemática . . . . .	3
1.2. Justificación . . . . .	5
1.3. Antecedentes . . . . .	5
<b>2. Pregunta de investigación y objetivos</b>	<b>6</b>
2.1. Pregunta de investigación . . . . .	6
2.2. Objetivos . . . . .	6
<b>3. Marco Teórico</b>	<b>6</b>
3.1. Celdas solares sensibilizadas por tintes (CSPT) . . . . .	6
3.1.1. Componentes de las CSPT . . . . .	6
3.1.2. Funcionamiento de las CSPT . . . . .	6
3.1.3. Tipos de sensibilizantes . . . . .	6
3.2. Colorantes naturales . . . . .	6
3.2.1. Características físicas y químicas . . . . .	6
3.2.2. Familias de colorantes naturales usados como sensibilizantes . . . . .	6
3.3. Enfoque de investigación direccionado por datos . . . . .	6
3.3.1. Tipos de enfoques de investigación en química . . . . .	6
3.3.2. Tipos de estructuras de datos en química . . . . .	6
3.3.3. Extracción y preprocesamiento de datos . . . . .	6
3.3.4. Tratamiento de datos . . . . .	6
3.3.5. Descubrimiento y análisis de características . . . . .	6
3.3.6. Entrenamiento y evaluación de modelos . . . . .	6
3.3.7. Validadores estadísticos de los modelos . . . . .	6
3.4. Modelos de aprendizaje automático . . . . .	6
3.4.1. Naturaleza de los modelos de aprendizaje automático . . . . .	6
3.4.2. Interpretatividad de los modelos de aprendizaje automático . . . . .	6
3.4.3. Clasificadores clásicos . . . . .	6
3.4.4. Clasificadores ensamblados . . . . .	6
3.4.5. Evaluadores de los resultados de clasificación . . . . .	6
<b>4. Materiales y métodos</b>	<b>6</b>
4.0.1. Extracción de datos . . . . .	6
4.0.2. Preprocesamiento de datos . . . . .	6
4.0.3. Descubrimiento y análisis de características . . . . .	6
4.0.4. Entrenamiento y evaluación del modelo 1 . . . . .	6
4.0.5. Entrenamiento y evaluación del modelo 2 . . . . .	6
4.0.6. Entrenamiento y evaluación del modelo 3 . . . . .	6
4.0.7. Comparación de modelos . . . . .	6
4.1. Análisis estadístico . . . . .	6
4.1.1. Pruebas de normalidad . . . . .	6
<b>5. Referencias bibliográficas</b>	<b>7</b>
<b>6. Cronograma de trabajo</b>	<b>8</b>
<b>7. Presupuesto y financiamiento</b>	<b>8</b>

## Resumen

Las celdas solares son dispositivos que convierten energía lumínica en energía eléctrica útil a través de fenómenos electroquímicos. El desarrollo de materiales sostenibles y con mayor eficiencia es un reto de investigación y también ambiental, ya que resolverlo conlleva explorar miles de millones de compuestos y hallar dicho tipos de materiales favorecerían el consumo de energías limpias. Esta exploración requiere recursos humanos y materiales que muchos investigadores no pueden proporcionarse, es entonces que las simulaciones computacionales y el enfoque dirigido por datos cobran importancia. El uso de experimentos *in silico* han ...

# 1. Introducción

## 1.1. Problemática

La dependencia de los combustibles fósiles trae consigo problemas como la inseguridad energética y el incremento de gases de efecto invernadero. Las sociedades que concentran la producción de los combustibles fósiles podrían controlar geopolítica y económicamente a otras [1], esta inseguridad energética se evidenció en la crisis del petróleo de 1973 [2] y en la disputa entre Rusia y la Unión Europea por el suministro de gas en este año [3]. En contraste, el problema del incremento de la temperatura global es un problema mediáticamente más discreto pero no menos alarmante. Tanto así que, en el 2018, la *Intergovernmental Panel on Climate Change* (IPCC) emitió un informe sobre los impactos que causaría dicho incremento en 1.5°C con respecto a los niveles preindustriales para el 2040, que en resumidas cuentas, se prevé un detrimento crítico y sin retorno de la civilización y la biósfera [4]. A fin de hacer frente a estos problemas, la transición hacia una matriz energética mundial donde predominen fuentes energéticas menos contaminantes y descentralizadas es la solución.

Las fuentes renovables reúnen dichas características, por lo que muchos gobiernos y organizaciones han tomado acciones para aprovecharlas. En ese marco, la *International Renewable Energy Agency* (IRENA) realizó un análisis multisectorial en el 2020 donde propuso una hoja de ruta para que las energías renovables generen el 86 % de la electricidad global [5]. En ese documento también se menciona que esta cifra no se alcanzaría sin la investigación ni el desarrollo de tecnologías que aprovechen dichas fuentes para hacerlas sostenibles y comercialmente viables.

Hace un año, la energía hidroeléctrica generó más de 4 mil tera-vatios por hora (TWh) en el mundo, lo que representó más del 50 % de la generación de todas las renovables [6]. No obstante, construir centrales hidroeléctricas favorecería solo a los países con ríos en zonas montañosas. Sin contar esta fuente, las energías eólica y solar son las que encabezan la producción energética. Con el propósito de contextualizar

en el 2021, y que su vez es mayor que el total de potencia consumida por las otras energías renovables, cerca de 3.5 mil TWh, no es rentable la construcción de centrales hidroeléctricas en diversos territorios a diferencia de las termoeléctricas. En contraste a esta situación, las otras fuentes son menos costosas de implementar y sus unidades de generación energética son modulables pero el reto actual es incrementar su potencia de generación y su adaptación en el mercado. Entre dichas fuentes, resaltan la energía eólica y solar estas son las que más porcentaje de potencia de consumo han tenido en el 2021 a nivel mundial, pero la energía solar es la que más crecimiento mostró sobre todo en sudamérica y el Perú (ver Figura 1). Este hecho puede deberse a causas políticas, económicas y tecnológicas; pero hace evidente que serán una fuente renovable no hídricas preponderante en los próximos años.

La energía solar es accesible para las personas en todos los terrenos habitables y gracias a los 140 mil TWh/h que el sol genera se considera una fuente casi ilimitada [8]. Sin embargo, el aprovechamiento de esta energía requiere de dispositivos que deben ser capaces de capturar, transformar y almacenar la mayor cantidad de energía lumínica en trabajo útil en rangos de tiempo variantes a causa del tiempo atmosférico y los estadíos solares.

Existen dos tipos de tecnologías que aprovechan la energía solar: La termosolar y la fotovoltaica. En la tecnología termosolar se condensa el haz solar en un contenedor de un fluido de tal manera que este se caliente y mueva un alternador eléctrico, mientras que la energía fotovoltaica convierte la radiación solar en energía electroquímica directamente por medio de fenómenos fotovoltaicos que se ocurren dentro de los materiales. Las centrales termosolares tienen el potencial de generar miles de kilos-vatios por hora (KWh) no obstante existen desventajas que limitan la difusión de su uso tales como su alto costo en la instalación, mantenimiento y distribución por cableado eléctrico,

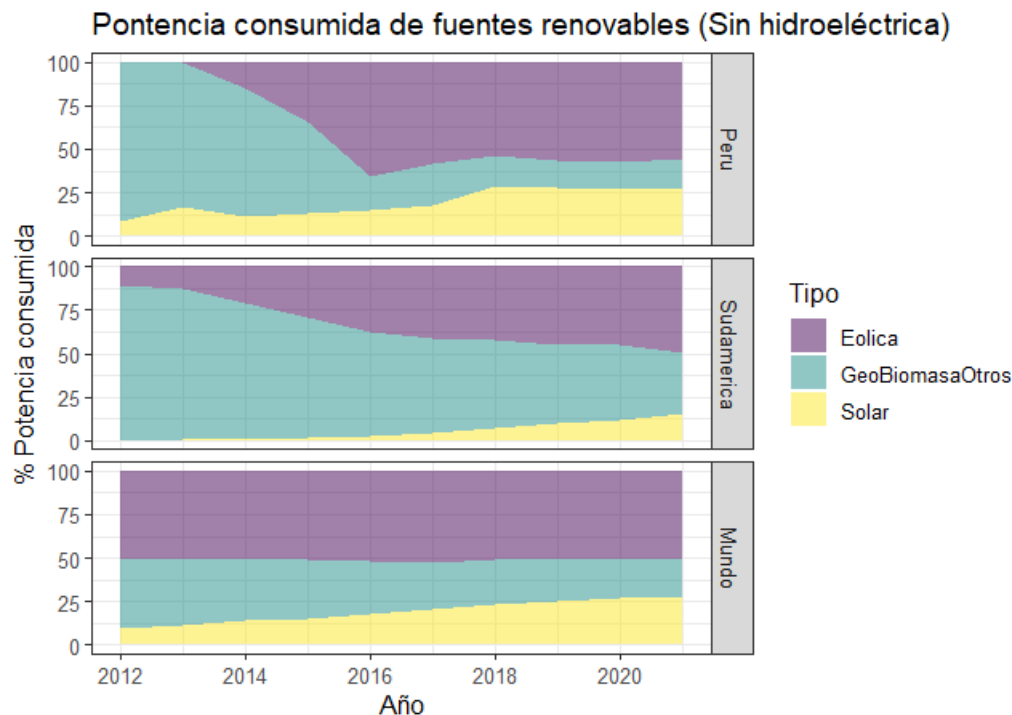


Figura 1: Comparación del porcentaje de la cantidad de potencia consumida por tipo de energía renovable sin considerar la hidroeléctrica. Fuente: Our World in Data [7]

así como la gran cantidad de energía desperdiciada en los procesos de transferencia de calor. Estos problemas están casi ausentes en los sistemas fotovoltaicos ya que se usan paneles y módulos que facilitan su uso, sin embargo aún es un reto equiparar la eficiencia de estos dispositivos con los termosolares. Si bien existen tecnologías que unen la termosolar y fotovoltaica, el presente trabajo se enfoca en mejorar los dispositivos fotovoltaicos.

Las celdas solares, centros funcionales de los sistemas fotovoltaicos, llevan en el mercado más de 60 años, desde su descubrimiento en los Laboratorios Bell en 1954 [9] y su primer uso comercial en el satélite Vanguard [10]. En términos de impacto ambiental, los sistemas fotovoltaicos generan entre 14 a 73 gr- $CO_2$ /kWh, lo cual es casi 90 % menos del impacto que generan los sistemas basados en gas [11]. Gracias

Blakers cita a la *National Renewable Energy Laboratories* (NREL) con el propósito de describir la clasificación que esta institución hace a varios tipos de celdas según los mejores rendimientos que mostraron en sus fases experimentales [12], por otro lado Rathore y su equipo se basaron en la antigüedad comercial y madurez técnica de estas [13]. Por fines prácticos se hace uso de esta última para describir las tres generaciones que engloban varios tipos de celdas así como sus características principales.

En la primera generación están las celdas construídas a base de obleas de silicio monocristalinas y policristalinas. Las monocristalinas han alcanzado rendimientos mayores al 20 % [14], no obstante su fabricación es costosa [15] debido al "recristalizado" de silicio por el método de Czochralski [16] para formar lingotes de monocristal que luego se cortan. Frente a esto, el uso de diferentes policristales en las celdas abarató el costo gracias a que se fabrican con procesos de

enfriamiento simples, además sus eficiencias de conversión energética oscilan de entre 12-14 %, por lo que estas características le permiten ser el tipo de celda con más prevalencia en el mercado [17].

En la segunda generación se encuentran las celdas en base películas delgadas de silicio amorfo (a-Si) [18], cadmio-teluro (CdTe) [19] y cobre-indio-galio-selenio (CIGS) [20], la característica principal de estas celdas es que son económicas en comparación con las de otras generaciones [13], sin embargo sus limitaciones son la inestabilidad de la eficiencia [21] así como los impactos negativos en el ambiente y a la salud que causan sus materiales, como el cadmio por ejemplo [20].

Finalmente, la tercera generación agrupa a las celdas más recientes tales como aquellas basadas en concentradores, nano cristales, polímeros y sensibilizadas por tintes. Las celdas por concentradores son dispositivos rígidos que aprovechan la radiación y el calor que se generan en al incidir luz en lentes convexos [19], si bien tiene viabilidad comercial se demandan mejoras en los materiales fotovoltaicos. Las celdas basadas en nano cristales o *quantum dots* (QD) pueden capturar mejor la luz mientras suprimen fenómenos de recombinación además de tener un rendimiento estable por encima del 20 % [22], no obstante no son comercialmente viables porque son costosos de fabricar [23]. Por el contrario, las celdas de polímero se caracterizan por su flexibilidad pero sus rendimientos están en promedio por debajo del 10 % y a su vez son inestables porque son sensibles a cambios de temperatura y presión [24]. Por último, las celdas sensibilizadas por tintes son dispositivos que tienen semiconductores fotosensibilizados y ensamblados a un sistema electrolítico que permite un flujo continuo de electrones [25], se caracterizan por su construcción versátil y de bajo costo gracias a la diversidad materiales de sus componentes, sin embargo sus rendimientos aún no se equiparan a las comerciales de silicio [26].

Las celdas sensibilizadas por tintes (CSPT) pueden ser

... debido a todas estas razones, la investigación de nuevos materiales y procesos de fabricación para celdas solares sensibilizadas por tintes más costo eficientes cobran relevancia.

Las estructura de las celdas solares

Es claro que deben de validarse experimentalmente los resultados obtenidos por los métodos computacionales, pero estos pueden discriminar cientos de candidatas

pero dado un grado de precisión razonable en la predicción de la propiedad del material, estos modelos computacionales aceleran el proceso de elegir mejores estructuras o moléculas candidatas con la correcta combinación de propiedades necesarias para cumplir satisfactoriamente con el propósito para las que fueron hechas *Leon R. Devereux*.

## 1.2. Justificación

La referencia es citada en [27]

Los procesos en el descubrimiento de celdas solares con mayor eficiencia de producción energética y sostenibilidad tales como el diseño, la prospección, la construcción son lentos, costosos y poco eficientes. Y el tiempo que toma que estas nuevas celdas solares tengan rentabilidad comercial para las partes que financiaron la investigación es reducido [28].

La ciencia de materiales dirigida por datos

## 1.3. Antecedentes

En los antecedentes se

## **2. Pregunta de investigación y objetivos**

### **2.1. Pregunta de investigación**

### **2.2. Objetivos**

## **3. Marco Teórico**

### **3.1. Celdas solares sensibilizadas por tintes (CSPT)**

#### **3.1.1. Componentes de las CSPT**

#### **3.1.2. Funcionamiento de las CSPT**

#### **3.1.3. Tipos de sensibilizantes**

### **3.2. Colorantes naturales**

#### **3.2.1. Características físicas y químicas**

#### **3.2.2. Familias de colorantes naturales usados como sensibilizantes**

### **3.3. Enfoque de investigación direccionado por datos**

#### **3.3.1. Tipos de enfoques de investigación en química**

#### **3.3.2. Tipos de estructuras de datos en química**

#### **3.3.3. Extracción y preprocesamiento de datos**

#### **3.3.4. Tratamiento de datos**

#### **3.3.5. Descubrimiento y análisis de características**

#### **3.3.6. Entrenamiento y evaluación de modelos**

#### **3.3.7. Validadores estadísticos de los modelos**

### **3.4. Modelos de aprendizaje automático**

#### **3.4.1. Naturaleza de los modelos de aprendizaje automático**

#### **3.4.2. Interpretatividad de los modelos de aprendizaje automático**

#### **3.4.3. Clasificadores clásicos**

#### **3.4.4. Clasificadores ensamblados**

#### **3.4.5. Evaluadores de los resultados de clasificación**

## **4. Materiales y métodos**

### **4.0.1. Extracción de datos**

### **4.0.2. Preprocesamiento de datos**

### **4.0.3. Descubrimiento y análisis de características**

### **4.0.4. Entrenamiento y evaluación del modelo 1**

### **4.0.5. Entrenamiento y evaluación del modelo 2**

### **4.0.6. Entrenamiento y evaluación del modelo 3**

### **4.0.7. Comparación de modelos**

## **4.1. Análisis estadístico**

### **4.1.1. Pruebas de normalidad**

La verificación de la normalidad de la distribución de un conjunto de datos se puede obtener a través de técnicas gráficas y por pruebas.

En primer lugar, aquellas basadas en gráficas tienen como

#### **Data**

La cien Actualmente, cerca del 50 % de las nuevas publicaciones de artículos son de acceso abierto y estimaciones hechas por (Ref 33 Y 34) indican que estás casi en su totalidad lo serán también para el 2040.

#### **Herramientas de libre acceso**

El software de acceso abierto es el conjunto de herramientas computacionales que permiten las operaciones de la ciencia de datos basada en datos. Este tipo de software adquiere importancia desde la aparición pública del internet y el sistema operativo LINUX en 1990, cuyos desarrollos fueron netamente colaborativos. Adicionalmente, el nacimiento de herramientas de control de versiones y lenguajes de programación de alto nivel permitieron el nacimiento de repositorios proyectos científicos (46), notebooks electrónicos (47), paquetes de simulación y experimentos(49) y librerías de aprendizaje automático y ciencia de datos (50).

## **5. Referencias bibliográficas**

### **Referencias**

1. Mayer A. Fossil fuel dependence and energy insecurity. *Energy, Sustainability and Society* 2022; 12:1-13
2. Vernon R. Oil Crisis. 1976
3. Rodriguez-Fernandez L, Carvajal ABF y Tejada VF de. Improving the concept of energy security in an energy transition environment: Application to the gas sector in the European Union. *The Extractive Industries and Society* 2022; 9:101045
4. Guilyardi E, Lescarmontier L, Matthews R, Point SP, Rumjaun AB, Schlüpmann J y Wilgenbus D. IPCC Special Report “Global Warming of 1.5° C”: Summary for Teachers. 2018
5. Asmelash E, Prakash G, Gorini R y Gielen D. Role of IRENA for global transition to 100 % renewable energy. *Accelerating the transition to a 100 % renewable energy era*. Springer, 2020 :51-71
6. IRENA. Renewable Energy Statistics 2022. Renewable Energy Target Setting, Abu Dhabi, UAE 2022
7. Ritchie H, Roser M y Rosado P. Energy. Our World in Data 2020. <https://ourworldindata.org/energy>
8. Hammarström L. Overview: capturing the sun for energy production. *Ambio* 2012; 41:103-7
9. Green MA. The path to 25 % silicon solar cell efficiency: History of silicon cell evolution. *Progress in photovoltaics: research and applications* 2009; 17:183-9
10. Singh GK. Solar power generation by PV (photovoltaic) technology: A review. *Energy* 2013; 53:1-13
11. Tawalbeh M, Al-Othman A, Kafiah F, Abdelsalam E, Almomani F y Alkasrawi M. Environmental impacts of solar photovoltaic systems: A critical review of recent progress and future outlook. *Science of The Total Environment* 2021; 759:143528
12. Blakers A, Zin N, McIntosh KR y Fong K. High efficiency silicon solar cells. *Energy Procedia* 2013; 33:1-10
13. Rathore N, Panwar NL, Yettou F y Gama A. A comprehensive review of different types of solar photovoltaic cells and their applications. *International Journal of Ambient Energy* 2021; 42:1200-17



14. Gul M, Kotak Y y Muneer T. Review on recent trend of solar photovoltaic technology. *Energy Exploration & Exploitation* 2016; 34:485-526
15. Srinivas B, Balaji S, Nagendra Babu M y Reddy Y. Review on present and advance materials for solar cells. *International Journal of Engineering Research-Online* 2015; 3:178-82
16. Yu X y Yang D. Growth of crystalline silicon for solar cells: Czochralski Si. *Handbook of Photovoltaic Silicon* 2019 :129-74
17. Sharma S, Jain KK, Sharma A y col. Solar cells: in research and applications—a review. *Materials Sciences and Applications* 2015; 6:1145
18. Kaur M y Singh H. A review: comparison of silicon solar cells and thin film solar cells. *Int. J. Core Eng. Manag.(IJCEM)* 2016; 3:15-23
19. Bertolli M. Solar Cell Materials. Course: Solid State II. Department of Physics, University of Tennessee 2008
20. Bagher AM, Vahid MMA y Mohsen M. Types of solar cells and application. *American Journal of optics and Photonics* 2015; 3:94-113
21. Gul M, Kotak Y y Muneer T. Review on recent trend of solar photovoltaic technology. *Energy Exploration & Exploitation* 2016; 34:485-526
22. Kim M, Jeong J, Lu H, Lee TK, Eickemeyer FT, Liu Y, Choi IW, Choi SJ, Jo Y, Kim HB y col. Conformal quantum dot-SnO<sub>2</sub> layers as electron transporters for efficient perovskite solar cells. *Science* 2022; 375:302-6
23. Jean J, Xiao J, Nick R, Moody N, Nasilowski M, Bawendi M y Bulović V. Synthesis cost dictates the commercial viability of lead sulfide and perovskite quantum dot photovoltaics. *Energy & Environmental Science* 2018; 11:2295-305
24. Gusain A, Faria RM y Miranda PB. Polymer solar cells—Interfacial processes related to performance issues. *Frontiers in chemistry* 2019; 7:61
25. Suhaimi S, Shahimin MM, Alahmed Z, Chyskỳ J y Reshak A. Materials for enhanced dye-sensitized solar cell performance: Electrochemical application. *Int. J. Electrochem. Sci* 2015; 10:2859-71
26. Sharma K, Sharma V y Sharma S. Dye-sensitized solar cells: fundamentals and current status. *Nanoscale research letters* 2018; 13:1-46
27. Wen Y, Fu L, Li G, Ma J y Ma H. Accelerated Discovery of Potential Organic Dyes for Dye-Sensitized Solar Cells by Interpretable Machine Learning Models and Virtual Screening. *Solar RRL* 2020; 4:2000110
28. Eagar TW. Bringing new materials to Market. en. *ABI/INFORM Global* 1995; 98:42

## **6. Cronograma de trabajo**

## **7. Presupuesto y financiamiento**