

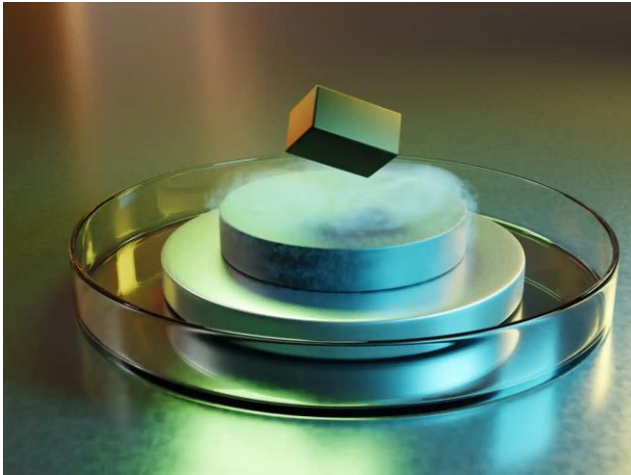
Introducción a la Ciencia de Datos

Explorando la Relación entre la Composición de los Elementos y la Superconductividad a Altas Temperaturas mediante Redes Neuronales Densas

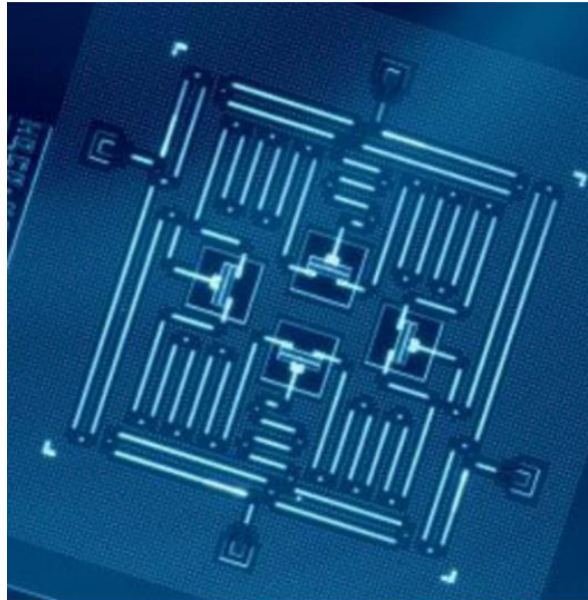


Eber David Gaytán Medina

INTRODUCCIÓN



El desarrollo de superconductores a temperatura ambiente es uno de los desafíos más importantes en la física de materiales



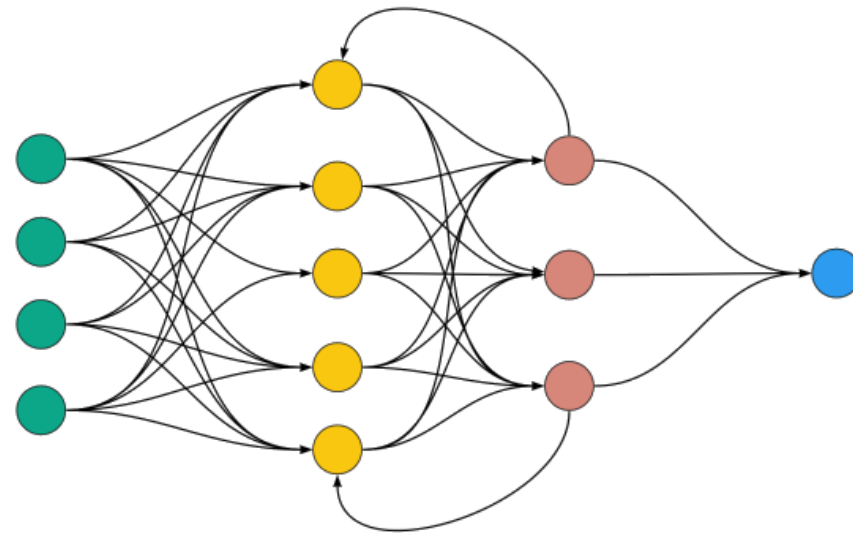
Un área de gran impacto es la computación cuántica, donde los superconductores son esenciales para la operación de los qubits



Los superconductores ofrecen soluciones potenciales para problemas críticos en la producción y transporte de energía eléctrica

● INTRODUCCIÓN

Para acelerar el descubrimiento de nuevos materiales con estas propiedades, el uso de herramientas avanzadas como el **aprendizaje automático** se vuelve crucial, permitiendo predecir y validar compuestos prometedores antes de su fabricación y prueba experimental.

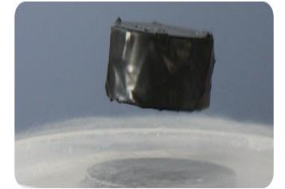


METODOLOGÍA

- Se utilizaron dos dataset.
- El **primer dataset** se hizo a partir un conjunto de datos con descriptores químicos y físicos de materiales, con el objetivo de establecer una relación entre la composición de los elementos y la superconductividad a altas temperaturas.
- El **segundo dataset** es la base y tiene el formato de material y temperatura crítica.
- Para asegurarse de condensar la mayor cantidad de información en un solo dataset se hizo un “merge” a partir de la temperatura crítica, uniendo 82 columnas del primer dataset y 87 del segundo.

Superconductivity Data Data Set

Superconductivity Data Data Set



Data Card Code (7) Discussion (0) Suggestions (0)

About Dataset

Source:

Kam Ham idieh, khamidieh '@' gmail.com, University of Pennsylvania, Statistics

Data Set Information:

There are two files: (1) train.csv contains 81 features extracted from 21263 superconductors along with the critical temperature in the 82nd column, (2) unique_m.csv contains the chemical formula broken up for all the 21263 superconductors from the train.csv file. The last two columns have the critical temperature and chemical formula. The original data comes from [Web Link] which is public. The goal here is to predict the critical temperature based on the features extracted.

Usability ⓘ

10.00

License

[CC0: Public Domain](#)

Expected update frequency

Never

Tags

Science and Technology

● CONJUNTO DE DATOS

- El conjunto de datos empleado consta de **168 atributos** que describen la composición y las propiedades de los materiales superconductores, con temperaturas críticas que oscilan entre 0 y 140 K.
- Los atributos se generaron a partir de análisis químicos, físicos y estadísticas descriptivas de los elementos que componen cada material.

Número de elementos	Masa Atómica	Entropía	Rango de masa atómica	Valencia	Material	Densidad	Radio atómico	Conductividad térmica	Temperatura crítica
---------------------	--------------	----------	-----------------------	----------	----------	----------	---------------	-----------------------	---------------------

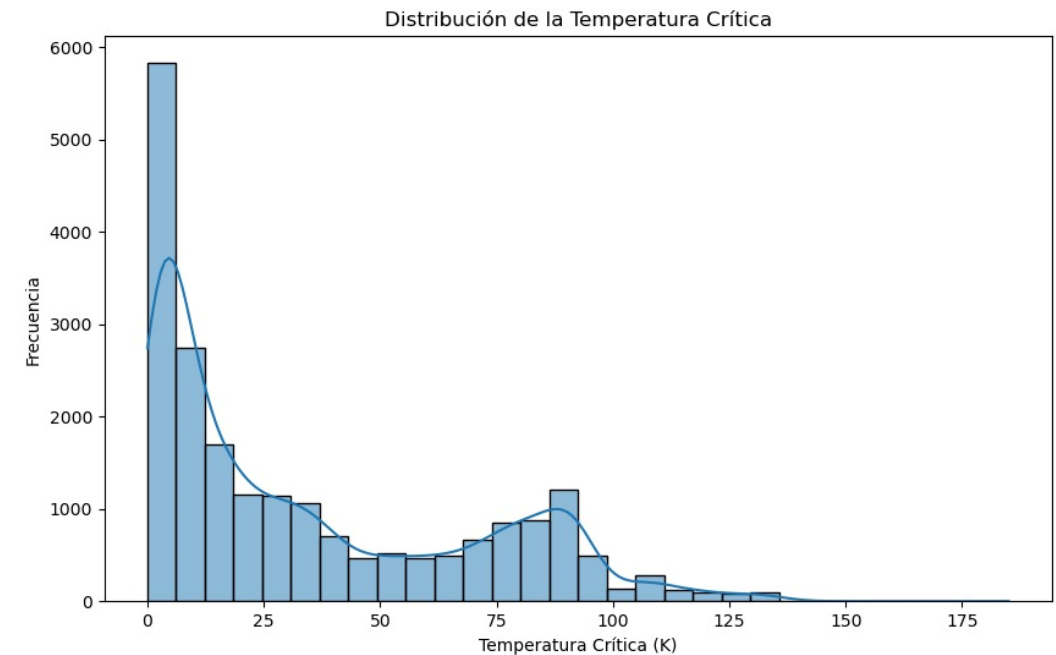
PREPROCESAMIENTO

GESTIÓN DE DATOS BALANCEADOS

- Se exploró la distribución de las temperaturas críticas para identificar posibles desbalances.
- Se emplearon histogramas y gráficos de densidad para visualizar la distribución.

ESTANDARIZACIÓN

- Se obtuvo que los datos estaban desbalanceados hacia temperaturas cercanas al 0 absolute.
- Se hizo una transformación z-score para que todos los atributos contribuyan de manera equitativa al entrenamiento del modelo.



● ARQUITECTURA DEL MODELO

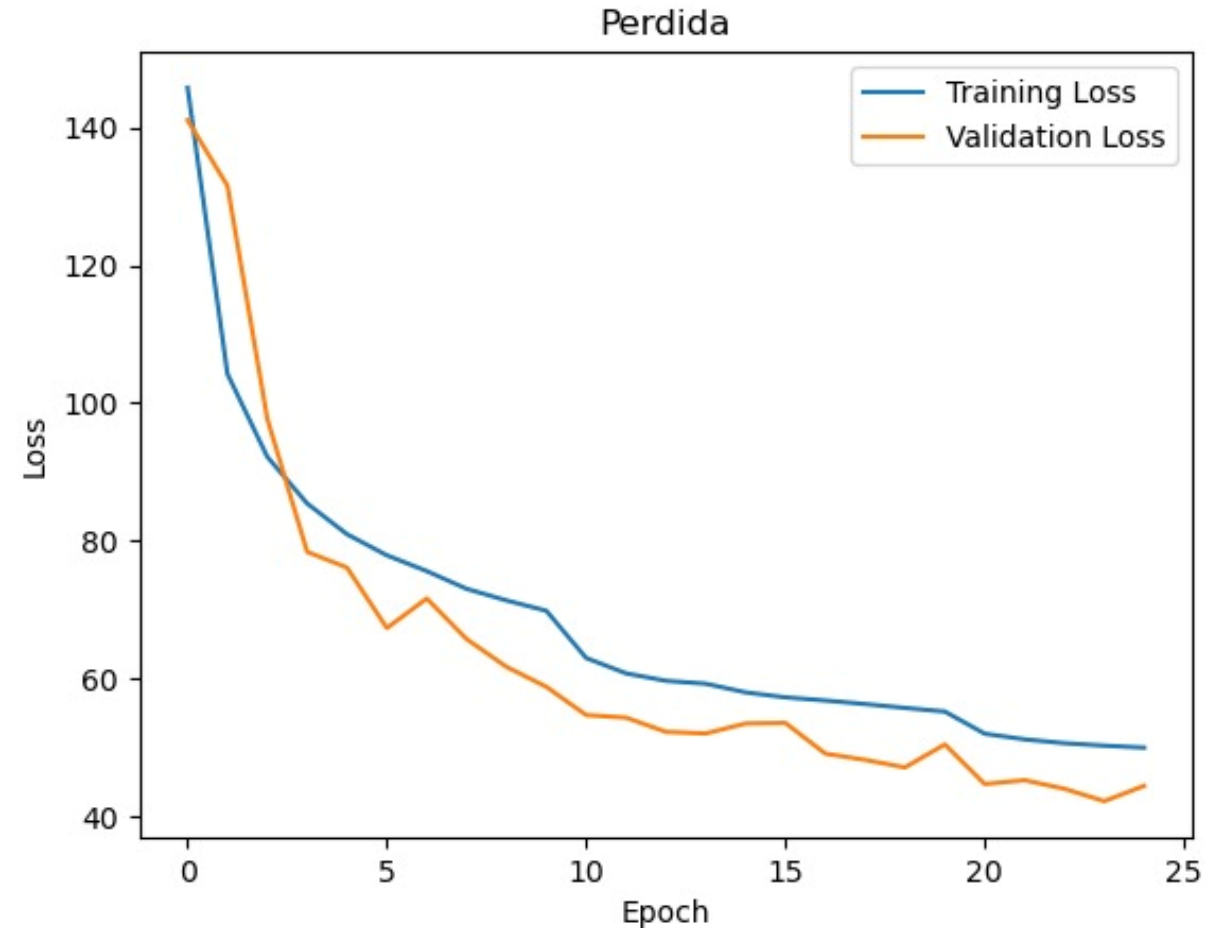
Se diseñó un modelo de red neuronal profunda utilizando Keras y se usaron capas densas con la justificación de atacar una problemática con muchos atributos, con la siguiente arquitectura:

- Una capa densa de 256 neuronas con función de activación ReLU, seguida de una capa de Dropout (0.3) para prevenir el sobreajuste.
- Una segunda capa densa de 128 neuronas con ReLU, seguida de una capa de Dropout (0.3).
- Capas adicionales de 64, 32 y 16 neuronas con ReLU para capturar características complejas.
- Una capa de salida de 1 neurona sin función de activación, adecuada para un problema de regresión.

El modelo fue compilado con el optimizador Adam y una tasa de aprendizaje inicial de 0.001, utilizando el error cuadrático medio (MSE) como función de pérdida y el error absoluto medio (MAE) como métrica de evaluación.

ENTRENAMIENTO

- Se entrenó el modelo durante 25 épocas con un tamaño de lote de 32 muestras.
- Se implementó un programador de tasa de aprendizaje que redujo la tasa a la mitad cada 10 épocas para optimizar la convergencia.
- Se usaron conjuntos de datos de validación para monitorear el rendimiento y ajustar los hiperparámetros.



● EVALUACIÓN

- El rendimiento del modelo se evaluó utilizando el MAE, el RMSE y el coeficiente de determinación (R^2) en el conjunto de prueba.
- Se compararon los resultados con modelos de referencia, como Random Forests y métodos de regression tradicionales, reportados en estudios previos.
- Se generaron gráficos de dispersión para analizar visualmente las predicciones frente a los valores reales de temperatura crítica.

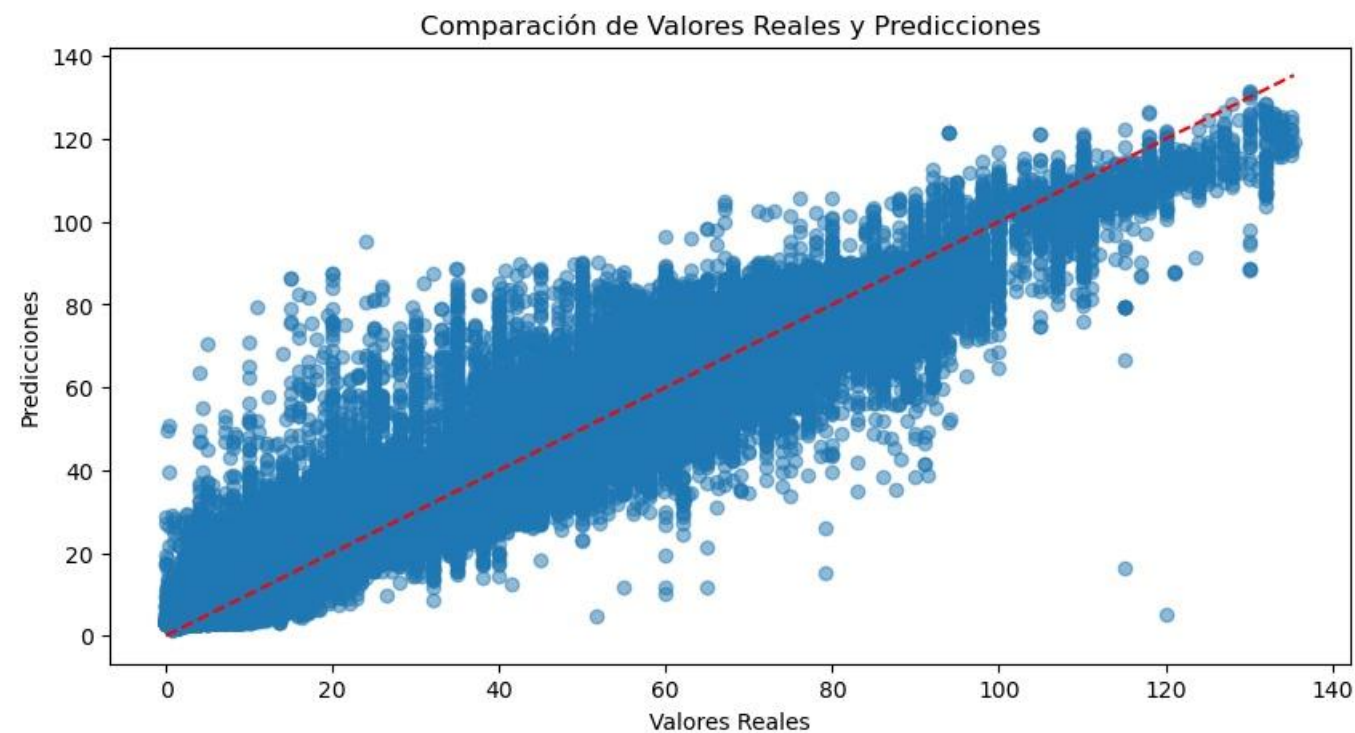
RESULTADOS

$RMSE : 6,4965$

$R^2_{Train} : 0,9621$

$R^2_{Test} : 0,9592$

Métrica / Modelo	SVD	Bagged Tree	DNN (Propuesto)
R^2	0.90	0.93	0.96
RMSE	9.67	8.91	6.49



CONCLUSIONES

- Se obtuvo un modelo de Redes Neuronales Densas capaz de predecir con una R^2 de 95 %, el cual, podría ayudar a validar aquellos superconductores teóricos sin necesidad de crearlos físicamente.
- Para una mayor exactitud al predecir críticas, es necesario combinar la tabla de composición con la extensión de características del material y no solo entrenar a partir de la tabla del material.
- Se demostró la eficacia de las redes neuronales densas para predecir temperaturas críticas en materiales superconductores.

● TRABAJO FUTURO Y LIMITACIONES

- Como trabajo a futuro se propone automatizar el proceso de la generación del dataset extendido de tal manera que baste con ingresar la fórmula química para obtener la mayor cantidad de datos útiles.
- Se identificaron limitaciones como la sensibilidad a atributos específicos que requieren validación experimental.
- Se sugiere explorar arquitecturas más avanzadas como Transformers para tareas similares.

● REFERENCIAS

- [1] L. Bassman Oftelie, A. De Pasquale, and M. Campisi, "Dynamic cooling on contemporary quantum computers," PRX Quantum, vol. 5, July 2024.
- [2] I. Sarajcev, M. Majstrovic, and I. Medic, "Calculation of losses in electric power cables as the base for cable temperature analysis," WIT Transactions on Engineering Sciences, vol. 27, 2000.
- [3] G. Gonz'alez, "Almacenamiento de energ'ia magn'etica por superconducci'3n," 2013.
- [4] M. Coll and A. Palau, "Superconductividad," 2017.
- [5] Tunguz, "Superconductivity data dataset." <https://www.kaggle.com/datasets/tunguz/superconductivity-data-data-set/code>, 2020. Último acceso: 19 de noviembre de 2024.
- [6] B. Roter and S. V. Dordevic, "Predicting new superconductors and their critical temperatures using machine learning," Physica C: Superconductivity and its applications, vol. 575, p. 1353689, 2020.

Introducción a la Ciencia de Datos

Explorando la Relación entre la Composición de los Elementos y la Superconductividad a Altas Temperaturas mediante Redes Neuronales Densas



Eber David Gaytán Medina