**Predicting Critical Temperature of Superconductors using Machine Learning Methods**

Superconductors are a material that under a certain critical temperature, the material is able to conduct current with zero resistance. This makes superconductors very favorable in certain applications. The most common application is the use of them in fMRI machines used by the healthcare industry, as well as various research facilities, such as the Large Hadron Collider.

Since the discovery of superconductors in 1911 by Heike Kamerlingh Onnes, there have been two main limitations that prevent more abundant use of superconductors. One being that superconductors favorable properties emerge only at very low temperatures making it challenging to keep the material under its critical temperature. The other issue is that there is no current model to predict the critical temperature of superconductors.

The goal of this project is to attempt to apply machine learning methods in order to predict the critical temperature of a superconductor given its chemical formula.

The data will be accessed from the Superconducting Material Database maintained by Japan’s National Institute for Materials Science (NIMS). There are 21,263 superconductor entries with 81variables per entry. These variables will be features of chemical properties of the material such as atomic mass, thermal conductivity, or number of valance electrons.

Initially, dimensionality reduction techniques will be used to determine if all 81 inputs are required to accurately predict the critical temperature, and if not how many will be sufficient. Neural Networks and Regression will be utilized to attempt to predict the critical temperatures of superconductors, and the two model will be compared in accuracy, efficiency, and reliability.

**Dataset**

The dataset used for this project was taken from 2018 paper titled, “A Data-Driven Statistical Model for Predicting the Critical Temperature of a Superconductor” [1]. It contains 21,263 rows and 81 input columns. The 81 input columns are made up of various feature extractions of each semiconductor material. Eight element properties were used to create features. The eight properties are atomic mass, first ionization energy, atomic radius, density, electron affinity, fusion heat, thermal conductivity, and number of valence. For each of the eight properties, ten methods were applied to extract features.

The following are the ten ways that the element data is manipulated.

*Mean:*

*Weighted Mean:*

*Geometric Mean:*

*Weighted Geometric Mean:*

*Entropy:*

*Weighted Entropy:*

*Range:*

*Weighted Range:*

*Standard Deviation:*

*Weighted Standard Deviation:*

Where,

t­i is an elements given property,

pi is the ratio of an element in a material,

is the fraction of a given property of a material,

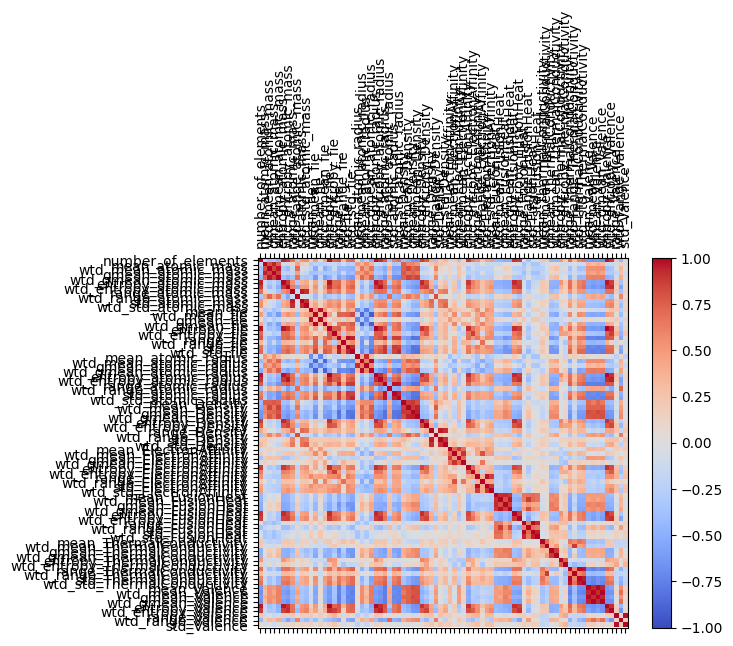
A

B =

This is how 80 input columns were generated with the last one being the number of elements in the material. The element data was gathered using ElementData from Mathematica, and the superconductors material information was obtained from the Superconducting Material Database supported by NIMS.

**Dimensionality**

Knowing that the 80 input dimensions were extracted from only 8 original inputs, some analysis was done to better understand how the data is related to each other. First, a Pearson Correlation was done to see if any two variables were linearly related with each other. The full correlation matrix is shown below with a subsection of the atomic mass property. It is seen that there are groups of variables that show strong relationship as well as groups that show no relationship at all. For example, the standard deviation and range of atomic mass have a Pearson Correlation of 0.971. This shows a very strong linear relationship and can lead to dimensionality reduction.



*Figure # - Complete Pearson Correlation Matrix for 80 dimensions*

A picture containing treemap chart

Description automatically generated

*Figure # - Sample Pearson Correlation matrix for single element property (atomic mass)*

Next, Principal Component Analysis was done to see if the provided dataset can be represented in less dimensions. A Scree Plot is shown below. The first two components make up 50% of the variance of the model. Expanding to 20 principal components captures 97% of the variance.

Chart, line chart

Description automatically generated

*Figure # - Scree Plot*

**Future Work**

The next steps include implementing the two machine learning models, Neural networks and Regression will be used to evaluate these models ability to predict critical temperature accurately. Also, using these models, the effects of dimensionality reduction will be analyzed.

**References**

Hamidieh, Kim. (2018). *A Data-Driven Statistical Model for Predicting the Critical Temperature of a Superconductor.* Available*:* <https://arxiv.org/abs/1803.10260>

MDR SuperCon Datasheet. National Institute for Materials Science Available: <https://mdr.nims.go.jp/collections/5712mb227>