

seminar report 2

by Rajrishi Sarkar

Submission date: 23-Apr-2021 11:34AM (UTC+0530)

Submission ID: 1567403308

File name: for_turnitin_1.docx (289.26K)

Word count: 1465

Character count: 7912

Seminar Report (MM 694)

On

**DATA DRIVEN APPROACH TO PREDICT THE
CRITICAL TEMPERATURE OF A
SUPERCONDUCTOR**

Submitted by

RAJRISHI SARKAR

(203110061)

In partial fulfillment for the award of the degree

Prep. (ETS)

Of

MASTER OF TECHNOLOGY

In

METALLURGICAL ENGINEERING AND MATERIALS SCIENCE

Guided by

Prof. HINA AMOL GOKHALE

Article Error (ETS)



INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY

Powai, Mumbai- 400076

APRIL 2021



ABSTRACT:

We know that superconductivity has been the focus of research for quite some time now. But the features responsible for this particular phenomenon remains quite mysterious. In this study we review some past papers which tries to estimate the superconducting critical temperature on the basis of a statistical model. Features were extracted based on atomic radius, atomic mass, thermal conductivity etc. To improve the model accuracy even new features were incorporated using the materials data from the AFLOW online repositories. Based on RMSE the models showed reasonable predictions with a deviation of 9.5K to 10K. The models don't say much about superconductivity. It only predicts the critical temperature of the superconductors.

LIST OF FIGURES:

Fig no.	Fig details	Page no.
Fig 1:	Histogram of materials categorized by critical temperatures T_c	7
Fig 2:	The distribution of critical temperatures T_c	8
Fig 3:	The relationship plot between the standard deviation per element and the mean critical temperature	8
Fig 4:	Plot of observed critical temperature vs the predicted critical temperature values based on the regression model.	9
Fig 5:	Plot between the predicted T_c versus observed value of T_c for XG-boost model	10

INTRODUCTION:

Materials that conduct current without offering any resistance are called superconductors, they have several applications. One of the applications would be Magnetic Resonance Imaging (MRI) used vastly in the field of medical science. Among other applications, would be Large Hadron Collider (LHC) and SQUIDS. Although superconductivity is a very interesting area to explore but there are two major drawbacks:

- a) It only behaves like a superconductor below a certain critical temperature (T_c) which is impractical and difficult to achieve.
- b) The model that is responsible for predicting T_c is an open problem which has become quite unfathomable in the scientific community.

¹ Application of statistical methods in order to predict the superconductivity started in the early 1980s with simple clustering methods. But due to absence of any theoretical solid models just simple empirical rules which are based experimental results have helped the scientists for several years. Say, the physicist Matthias concluded that T_c is related to the number of valence electrons per atom, is one of the empirical rules. Many of these empirical rules are violated.

² In this report we look into the papers which takes entirely Data-driven approach to create a statistical model that predicts the critical temperature T_c . The required data is extracted from the superconductivity material Database maintained by NIMS in Japan. Features are derived based on the elemental properties of the superconductor. Several Features are derived from the existing features creating a total of 81 features for each superconductor. Of course all the properties are not required for the model. Various statistical models were tried out in different papers but the most efficient are: A multiple regression model and a gradient boosting model. These are the models that generated maximum accuracy. Unlike regression based models the Gradient boosting models creates an ensemble of trees to predict T_c . After fitting a gradient boosting model the weighted average of all the trees are taken into consideration that finally gives the required prediction. It is found that in maximum cases the gradient boosting model outperformed the regression model as they were able to account for the complex correlations among the features. The multiple regression model shows an root mean squared error (R-squared) of 0.75 where as the XG boost model delivered an R-squared value of about 0.95.

LITERATURE SURVEY:

From the above discussion it is evident that statistical modelling has been a keen interest of many researchers to predict the critical temperature of the superconductors. Although we discussed about the best two models for the same but the main issue with these approaches is feature selection, after extracting them. Because out of all those features many will show correlation that will affect the accuracy of the models. This means various data pre-processing steps are required before fitting the required model into the data, followed by model tuning to improve the accuracy. We look into various literatures in order to figure out the best features and several data engineering techniques adopted to generate the best output i.e a model of maximum accuracy. All of those approaches are summarized below:

Data Pre-processing:

Stage 1: Element Data Preparation

The data consists of total 86 features i.e that many rows. Conder (2016) is a big help for choosing the important properties out of all those features but different paper uses different properties as per they see fit. Variables like "boiling point" are dropped and the corresponding feature which is correlated to it is used, like "Fusion heat" variable. After several trial and errors and gaining various insights literature survey suggests that there are 8 important variables that are the most important ones and hence used in the statistical model.

Stage 2: Pre-processing of the Superconducting Material data:

As mentioned before, the Superconducting Material data is maintained by NIMS, a Japan based institution. This database is the ultimate source of the superconductors along with their critical temperatures and elemental formulas. The elemental formulas are used to generate the required feature. This is considered as the most authentic database. The features that were extracted from the database are mentioned in the table:

Data Analysis:

This is the most important step towards **model** building as it helps to gain insight on how the features of the dataset are distributed, the correlation among the variables and the dependency of the target variable on the extracted features.

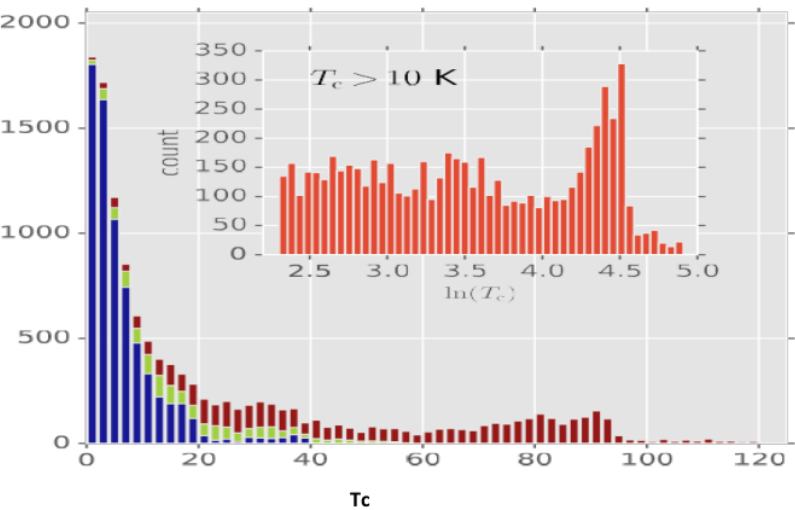


Fig 1: The Histogram of materials categorized by critical temperatures T_c . Blue, green, red denotes “low- T_c ”, iron based and cuprate semiconductors.

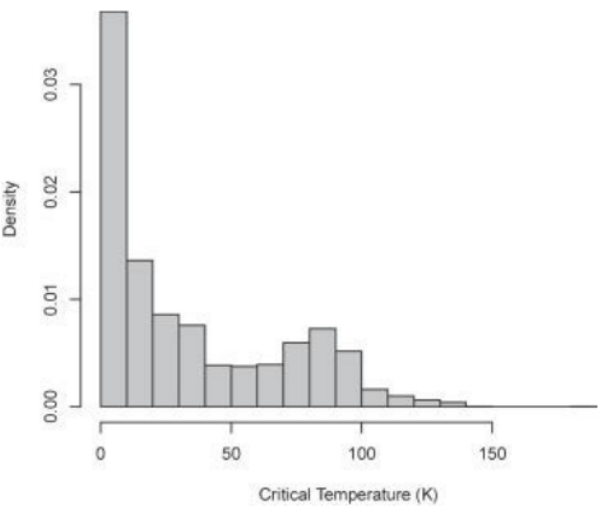


Fig 2: The above figure shows the distribution of critical temperatures T_c

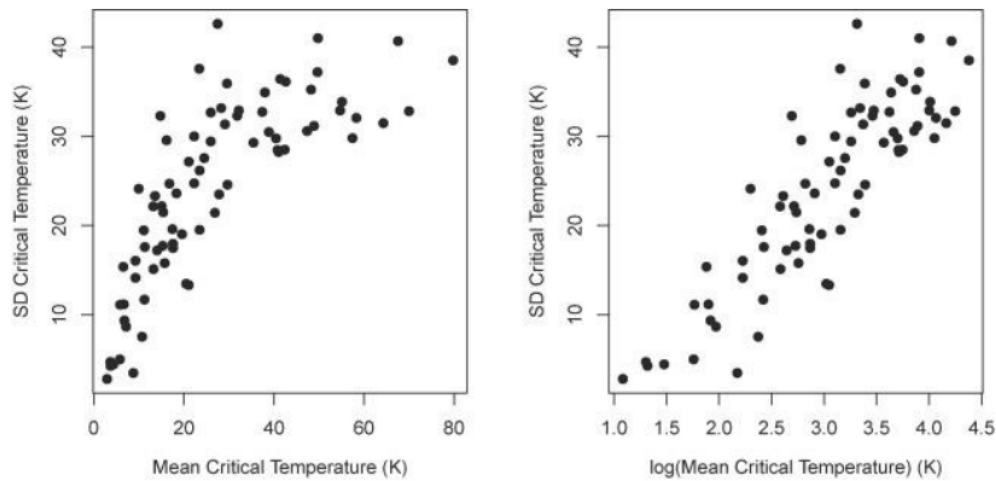


Fig 3: The first figure shows the relationship between the standard deviation per element and the mean critical temperature. The right panel shows the relation of $\log T_c$ with SD of the elements. It is evident that with the increase in mean critical temperature, value the standard deviation of critical temperature per element increases.

Model analysis:

Multiple Regression Model:

The R-squared value for multiple regression models lies between 0.70-0.76 approximately. The figure shown below shows the observed T_c versus the Predicted values of T_c for the superconductors. Actually the multiple regression model under-predicts critical temperature of high temperature superconductors, not only that, it also overpredicts the critical temperature values for the low temperature superconductors. Although it serves as the basis of other models, it should not be used for predicting the critical temperature values.

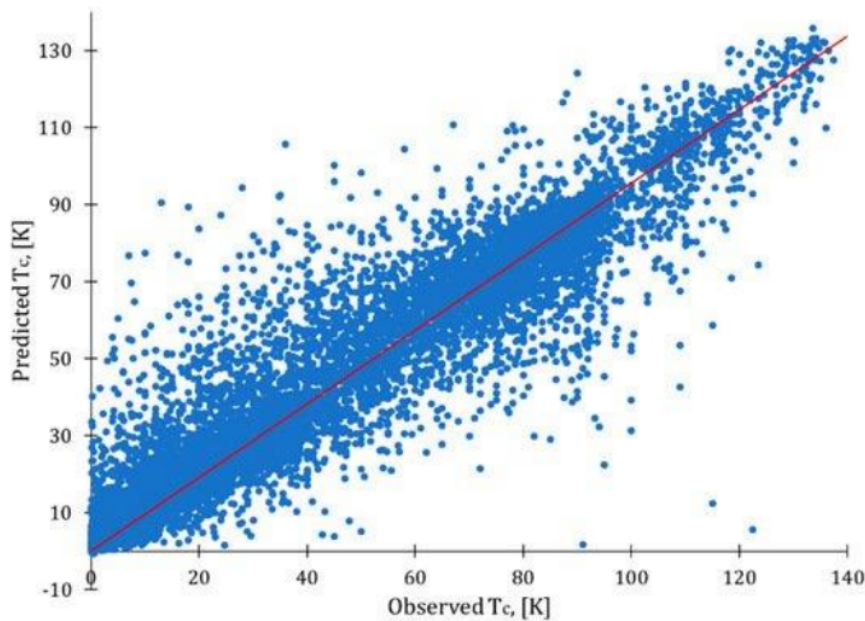


Fig 4: This is the plot of observed critical temperature vs the predicted critical temperature values based on the regression model.

XG-boost model:

The XG-boost model performs on:

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (2)$$

where $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$

Now x_i is the input feature, y_i is the feature that we need to predict and f_1, \dots, f_T is basically sequence of trees. Here L is the expected loss function for the i th predicted response and Ω is the penalty function. T is the no. of leaves per tree. The penalty function in the XG-boost model is beneficial for increasing the accuracy of the model. We can also use parameter tuning to achieve higher level of accuracy and increase the efficiency of the model. This includes tuning the learning parameter η and also feature sub sampling, which means to choose a fraction of features randomly while adding a new tree at each stage. The figure below shows the observed critical temperature vs the predicted critical temperatures for XG-boost model with an R-squared value of 0.90 to 0.97 approximately.

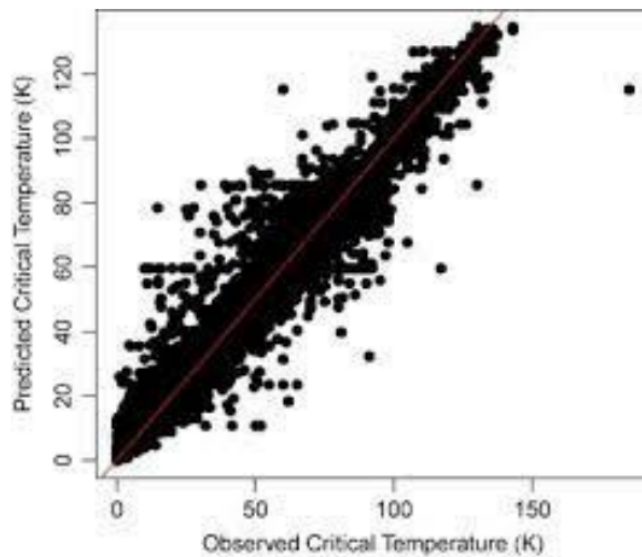


Fig 6: The plot between the predicted Tc versus observed value of Tc for XG-boost model

Well certainly this model performs far better judging by the plots. The R squared value of regression model do not measure up-to the standards of the XG-boost model. Now we look into the most important features for the model. The features with maximum importance with respect to the given model can be calculated by the following formulae: Prep. (ETS)

$$\text{Feature Gain} = (\text{Sum of gains for the feature}) * (\text{sum of gains for all the features})^{-1}$$

Features having higher gain are more important. The features with most importance are selected for next model evaluation, dropping the ones with least gains. This in turn increases the model accuracy and the model XG boost model improvises.

CONCLUSION:

In this report many studies on statistical modelling in order to predict the critical temperature of superconductors are reviewed. This work basically demonstrates the significant role statistical models play in superconductivity research. Several another databases may be used. Out of all the different models we narrowed down our focus to Multiple Regression and XG-boost models. Data preparation and Data analysis also plays a significant role to visualize the data and increase the efficiency of the models. Application of these models has the potential to enhance the search for high temperature superconductors.



seminar report 2

ORIGINALITY REPORT

6%

SIMILARITY INDEX

4%

INTERNET SOURCES

3%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1

www.nature.com

Internet Source

4%

2

Submitted to Columbia University

Student Paper

2%

Exclude quotes Off

Exclude bibliography Off

Exclude matches Off

seminar report 2

PAGE 1



Prep. You may be using the wrong preposition.



Article Error You may need to use an article before this word. Consider using the article **the**.

PAGE 2

PAGE 3



Missing "," Review the rules for using punctuation marks.



Possessive Review the rules for possessive nouns.

PAGE 4



Proofread This part of the sentence contains an error or misspelling that makes your meaning unclear.



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word. Consider using the article **the**.

PAGE 5



Run-on This sentence may be a run-on sentence.



Missing "," Review the rules for using punctuation marks.



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word. Consider using the article **a**.



P/V You have used the passive voice in this sentence. You may want to revise it using the active voice.



Confused You have used either an imprecise word or an incorrect word.



Prep. You may be using the wrong preposition.



Missing ", " Review the rules for using punctuation marks.



Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.



S/V This subject and verb may not agree. Proofread the sentence to make sure the subject agrees with the verb.



Confused You have used either an imprecise word or an incorrect word.



S/V This subject and verb may not agree. Proofread the sentence to make sure the subject agrees with the verb.



Missing ", " Review the rules for using punctuation marks.



Article Error You may need to use an article before this word.



Missing ", " Review the rules for using punctuation marks.



Article Error You may need to use an article before this word.



P/V You have used the passive voice in this sentence. You may want to revise it using the active voice.



Wrong Form You may have used the wrong form of this word.



Missing ", " Review the rules for using punctuation marks.



Prep. You may be using the wrong preposition.



Sentence Cap. Review the rules for capitalization.



Missing "?" Review the rules for using punctuation marks.



Wrong Article You may have used the wrong article or pronoun. Proofread the sentence to make sure that the article or pronoun agrees with the word it describes.



Proofread This part of the sentence contains an error or misspelling that makes your meaning unclear.



P/V You have used the passive voice in this sentence. You may want to revise it using the active voice.

PAGE 7



Article Error You may need to use an article before this word.



Proofread This part of the sentence contains an error or misspelling that makes your meaning unclear.

PAGE 8



Missing "," Review the rules for using punctuation marks.



Article Error You may need to use an article before this word.



Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.



Verb This verb may be incorrect. Proofread the sentence to make sure you have used the correct form of the verb.



Missing "," Review the rules for using punctuation marks.

PAGE 9



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word.



Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.



Sentence Cap. Review the rules for capitalization.



Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.



Frag. This sentence may be a fragment or may have incorrect punctuation. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.



Sentence Cap. Review the rules for capitalization.



Article Error You may need to use an article before this word. Consider using the article **a**.



Article Error You may need to use an article before this word. Consider using the article **the**.



Article Error You may need to use an article before this word. Consider using the article **the**.



Article Error You may need to use an article before this word.



Prep. You may be using the wrong preposition.



Article Error You may need to use an article before this word.



Prep. You may be using the wrong preposition.



Article Error You may need to use an article before this word.



Wrong Article You may have used the wrong article or pronoun. Proofread the sentence to make sure that the article or pronoun agrees with the word it describes.



Article Error You may need to use an article before this word.