Predicting the Critical Temperature of a Superconductor

Evan Walker

***Abstract*—Understanding what chemical properties most significantly determine a superconductors critical temperature has been of utmost importance since its discovery by Dutch physicist Heike Kamerlingh Onnes in 1911. Several theories have been made to describe the behavior of superconductors, most importantly the BCS theory, or Bardeen-Cooper-Schrieffer theory, which has been expanded upon since 1957 [2]. The theory accurately describes what we expect the atomic behavior of superconductors to be, in turn shedding light on the macroscopic properties. The issue at hand between BCS and other theories regarding superconductors is identifying new superconductors that are also simple, effective, and feasible to make. Hence, this paper aims to support theory with a purely empirical approach, attempting to make accurate predictions of the superconductor's effectiveness, namely by the critical temperature. Given a data set of 21,263 superconductors and 82 relevant chemical properties, including critical temperature, we find significant features that accurately explain the critical temperature and use them for model training. Methods used for feature extraction and predictions are PCA, linear and general linear regression, deep learning architecture, generalized boosted regression model, and random forests. Our linear model serves as a baseline model, with no preprocessing, we get an out-of-sample root-mean-squared-error of ± 17.93K and a multiple R^2 of .73.**

KeyWords: Superconductor; Critical Temperature; Machine Learning; Statistical Models; Data Mining

I. Background

SINCE the discovery of superconductors in 1911, many physicists have attempted to create a theory that accurately describes their properties. Most prominent among them are the BCS theory and the Ginzburg-Landau theory. Much like the bitter disconnect of Einstein's theory of gravitation and Quantum mechanics, the two predict on different scales but don't meet in the middle. BCS theory was published in 1957 and can accurately describe the microscopic effect of a superconductor at temperatures near absolute zero [2]. However, as the superconductor compound becomes more complex with more elements or as the temperature increases, BCS fails to adequately describe such phenomenon on a microscopic scale. Similarly, the Ginzburg-Landau theory(1950) can be derived from the BCS theory through restrictions of parameters [3]. After the restrictions have been made, the resulting equations are nonlinear partial differential equations of a complex-valued function [3]. It can describe the macroscopic phenomenons of Type I superconductors without taking microscopic properties into account. Note that GL theory was developed seven years before BCS theory, and was later shown that they are related. The limitations of the theory would take a full paper in itself dealing purely with intense equations and physics jargon, hence I leave the reader to acknowledge that the theory does not perfectly describe all superconductors and can read about the theory in the paper cited in the references section [3]. Not to say that either of these theories are failures, theoretical shortcomings provide important discourse in science such as the discourse between Einstein and Niels Bohr where the famous quote "god doesn't play dice" comes from (and the famous response from Bohr, "Don't tell god what to do"). What we have accomplished thus far is already a great feat considering that superconductivity has been one of the hardest problems facing physics of the 20^{th} century.

As to address what a Type I superconductor is, we will briefly discuss superconductor classification as it is based on important properties of the superconductor. Superconductors have been categorized into a few main groups based on chemical makeup, critical temperature, and magnetic properties [4]. Chemically, there are two prominent categories, those containing Oxygen and Copper called cuprates, and those containing Iron. A conventional superconductor is a superconductor that can be explained using the BCS theory, an unconventional superconductor is one that cannot be described by BCS theory. The most common classification is Type I and Type II, both based on the magnetic property of the superconductor. Taking a weak magnetic field, such as the fields created by magnets, and bringing it into close proximity to the superconductor can cause one of two things. It's the case that either the superconductors magnetic field excludes the weak magnetic field from the interior, or that the weak magnetic field permeates the superconductors magnetic field and causes the superconductor to no longer behave as such. In other words, when a material makes the transition to a superconducting state, it will exclude magnetic fields from its interior through perfect cancellation of the two opposing fields. This effect is known as, and described by, the Meissner effect [4]. We classify a superconductor as Type I or Type II through this critical field strength threshold, which is commonly written as H_c. The classification is as follows: Type I can only exist below the critical field strength, while Type II can exist above it and still function as a superconductor. We find that most Type I superconductors are also conventional superconductors. As a side note, the Meissner effect of zero magnetic field inside a superconductor is distinct from perfect diamagnetism [4].

Why is critical temperature so important to superconductors, or rather, why is it so important to us? All current technology that uses super conductors need to be cooled in order to keep it in its superconductive state. Some important machines that use superconductors are Magnetic Resonance Imaging Machines, Nuclear Magnetic Resonance, and of course the Large Hadron Collider [4]. NMR's are typically used for identifying organic molecules through matching the oscillation frequency of the magnetic field and the intrinsic frequency of the nuclei. The much more infamous MRI is used for making anatomical images of the body, and unlike any other scan, does not use X-rays but magnetic fields and radio waves to identify structures. In summary, there is a plethora of extra-expensive machines that use superconductors. They are expensive due to the superconductor, since they need coolant, tons of power to operate, are difficult to make. Hence, the next step would be to identify superconductors that have a higher critical temperature and are simple to make, resulting in a wider range of machines than can be built using superconductors. Consider the pitfalls of modern electronics for a moment. Attempt to generate 1,000 binomial distributions with 100,000 samples each, surely your computer will get extremely hot, regardless of your execution environment being CPU or GPU. Now consider the following trade off: resistance vs. heat. As you pump more Hz into your CPU, the energy lost due to resistance become heat, and in turn the heat increases resistance, and so on. This is because as the heat increases, the atoms jiggle more vigorously, resulting in collisions between electrons passing through, hence increasing resistance. This is why we see desktop computers with powerful GPU's being water cooled, and being fan cooled. Hypothetically speaking, we can build computers using superconductors to replace semiconductors, given that the technology to create superconductor circuits exist, and using liquid nitrogen as coolant. We would see a massive improvement in the speed of processing as well as energy requirements, and temperatures of our processors (which would extend their lifespan) [4]. Not only that, but imagine the implications that zero resistance would have on signal processing and error-correcting codes. Hence, the importance of superconductor investigation cannot be understated.

II. Introduction

Attempting to predict the critical temperature of a superconductor is no simple matter. In the data presented in the next section, there are 81 variables that reflect chemical properties of a superconductor, of which there is 21,263. The true question is which of the 82 variables are significant predictors of the critical temperature. All of the computational work, as well as visualizations, was done in R and deep-learning in python using keras and Tensorflow as the backend. I compare the models using the AIC metric, which penalizes model complexity, the root-mean-square-error of the models, and R^2. In the upcoming sections, I will elaborate on the form of the data, and continue on to the methodology employed in the analysis, followed by preprocessing of the data. Lastly, I will cover the models used to make predictions as well as the results of the models used, and finally the conclusion.

III. Superconductivity Data

The data collected is available on the UCI data set repository. It has been previously analyzed by Kam Hamidieh in the statistics department of the University of Pennsylvania [1]. The data supplied, according to K.Hamidieh, was merged from multiple research papers, but predominantly from Japan's National Institute for Materials Science (NIMS) [1]. Kam Hamidieh's paper dealt with a similar objective but with a different methodology, hence this papers purpose is not only to execute a different methodology, but to confirm any findings he made about the predictability of critical temperatures (and maybe get better results using machine learning). Kam Hamidieh is a statistician, and his paper shows this as he is predominantly concerned with multiple regression models. Before going into the form of the data, I would like to mention that K.Hamidieh took great care in preprocessing the data. As well as removing erroneous data, Hamidieh created the transformations in Table II. To see what else he did to the data, see his paper in the references. It's enough for our purposes to accept that the data is in the right form and free of NA's.

The data found in the UCI repository contains two data sets. One consisting of 81 features extracted from the 21,263 superconductors which relate to chemical properties in Table I, and transformations of each of those features using Table II. Both tables are derivative of tables in K.Hemidieh's paper, as it was the only documentation of the seemingly encrypted variable names. The second data set corresponds to the chemical formula split up into 88 features, each corresponding to the observed element in the superconductor. If the element

TABLE I
Variable transformations (for the case of a two element superconductor)

Transformation	Formula
Mean	$\frac{(t_1+t_2)}{2}$
Weighted Mean	$(p_1 t_1) + (p_2 t_2)$
Geometric Mean	$(t_1 t_2)^{1/2}$
Weighted Geometric Mean	$(t_1)^{p_1}(t_2)^{p_2}$
Entropy	$-w_1 ln(w_1) - w_2 ln(w_2)$
Weighted Entropy	$-Aln(A) - Bln(B)$
Range	$t_1 - t_2(t_1 > t_2)$
Weighted Range	$p_1 t_1 - p_2 t_2$
Standard Deviation	$\frac{\sqrt{((t_1-\mu)^2+(t_2-\mu)^2}}{2}$
Weighted Standard Deviation	$\sqrt{(p_1(t_1-\mu)^2 + p_2(t_2-\mu)^2}$

Where p_1 and p_2 are weights calculated from the ratio of the number elements from a superconductor, t_1 and t_2 are the thermal conductivity coefficient of each element in a superconductor. And,

$$w_1 = \frac{t_1}{t_1 + t_2}, \quad w_2 = \frac{t_2}{t_1 + t_2}$$

Lastly,

$$A = \frac{p_1 w_1}{p_1 w_1 + p_2 w_2}, \quad B = \frac{p_2 w_2}{p_1 w_1 + p_2 w_2}$$

(All above equations can and are extended into multiple dimensions, not just 2. We use 2 for legibility and ease of description, otherwise there would be summations and products in the table. The dimension is defined by the number of elements in a specific superconductors chemical formula.)

TABLE II
VARIABLES IN DATA (DERIVATIVE A TABLE IN K.HAMIDIEHS PAPER)

Variable	Unit	Description
Atomic Mass	Atomic mass units(AMU)	Total proton and neutron rest masses
First Ionization Energy	Kilo-Joules per mole (kJ/mol)	Energy required to remove a valence electron
Atomic Radius	Picometer (pm)	Calculated atomic radius
Density	Kilograms per meters cubed (kg/m^3)	Density at standard temperature and pressure
Electron Affinity	Kilo-Joules per mole (kJ/mol)	Energy required to add an electron to neutral atom
Fusion Heat	Kilo-Joules per mole (kJ/mol)	Energy to change from solid to liquid without temperature change
Thermal Conductivity	Watts per meter-Kelvin ($W/(mK)$)	Thermal conductivity coefficient k
Valence	No units	Typical number of chemical bonds formed by the element

was not observed, it was set to zero. Preliminary analysis was run on the second data set, however results were significantly worse than baseline models ran on the the chemical properties, hence the second data set is primarily used for visual analysis and linking critical temperatures to specific formulae.

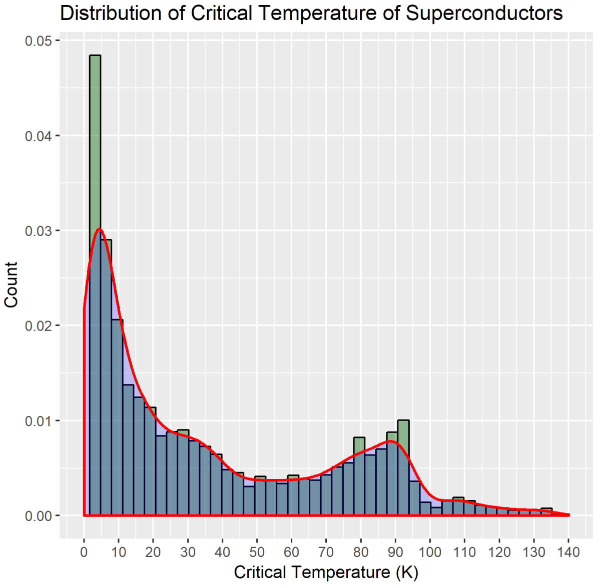


Fig. 1. Histogram and fitted density function for Critical Temperature

The above Fig. 1 shows the bimodal distribution of the critical temperature T_c. This shows the rarity of high T_c in superconductors. Further discussion of this distribution in upcoming sections.

Refer to the next page for Fig. 3. It should be noted that a similar diagram to Fig. 3 is presented in K.Hamidieh's paper, however Fig. 3 objectively shows more information than the one in his paper. In Fig. 3, we see that cuprates are extremely common by noticing the size of the points for Oxygen and Copper. The average T_c is extremely high for superconductors containing Mercury, Barium, Calcium, Titanium, and Copper. Notice the disconnect between the average T_c and the tail of the T_c distribution. This tells us that certain elements highly vary in critical temperature. For some superconductors that have $T_c > 100$, the individual elements in those superconductors never guarantee a tight distribution for T_c. For example, take Hg in Fig. 3, we have an average of approximately 80K, which is our highest average T_c. The tail of the distribution has T_c as high as 140K. Then, we may expect that the distribution of T_c for superconductors with a specific element, such as Hg, to vary highly as to attribute to the fact that the average is much lower than the tail of the histogram in Fig. 1.

To elaborate on our specific example, calculating the standard deviation of Hg, we get $38.52K$. The distribution of the T_c of superconductors containing mercury is shown in Fig. 2, which confirms the intuition that the distribution varies highly. However this intuition need not apply to elements with lower T_c, since it falls closer to the mean of the histogram of

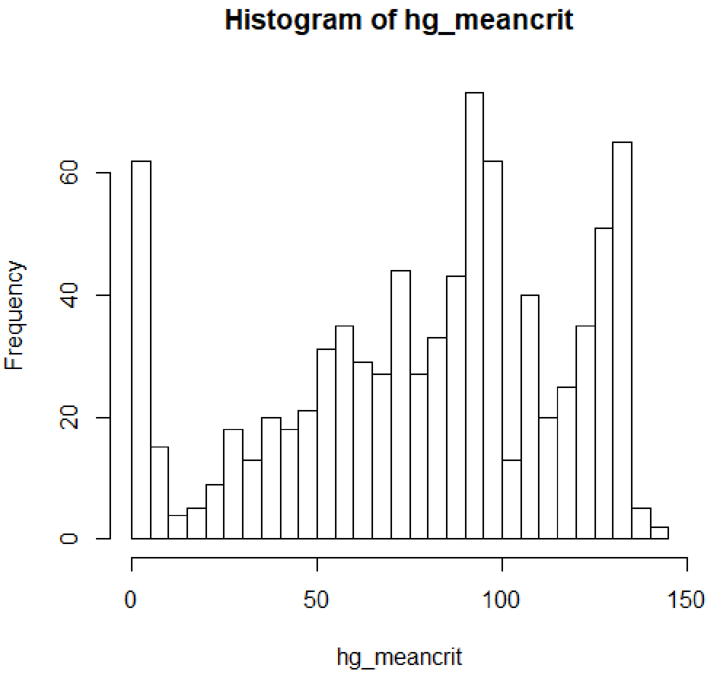
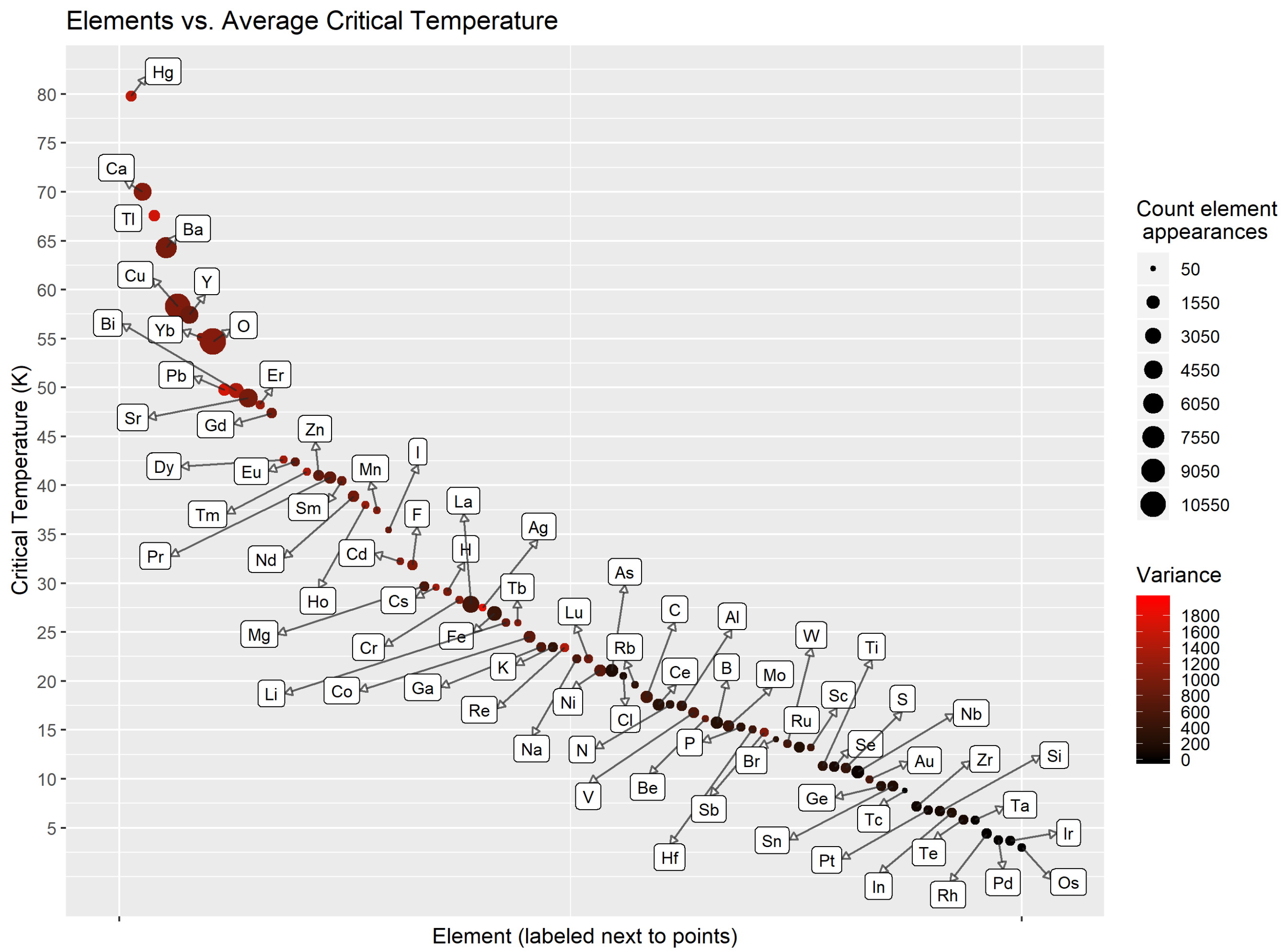


Fig. 2. Histogram of T_c for superconductors containing mercury



The critical Temperature is an average of the critical temperature of all superconductor that element appears in. The variance scale is the variance of the critical temeprature for the superconducdcotrs that contain that element

Fig. 3. Average T_c of all superconductors that contain the specified element. color shows tha variance of T_c for superconductors with that element. the size of each point is in accordance with the number of times that element shows up in the data

T_c, it's harder to make inferences. Hence, take note to the color of the points in Fig. 3, as they represent the variance of T_c for superconductors containing that element. Interestingly, elements with higher average T_c show more variance in T_c. This shows that elements that have higher average T_c cannot be trusted to always give us higher T_c in a superconductor, an important observation. This explains why regression would have trouble predicting T_c based on the elemental makeup alone, since the critical temperature has the ability to vary widely with each element.

IV. Methodology

In general, there is an important trade off when it comes to modeling data sets like this one. As described in the previous section, we are over saturated with features, and any model we create using all features will be overly complicated and get terrible AIC scores relative to other models with less features. So feature reduction/selection is a must, however if we were to transform our data, we lose interpretation at the gain of model accuracy and information criterion metrics. We have a trade off between accuracy and interpretability.

Hence, we may have three schools of thought for how we may go about this problem. One, where we invoke PCA on the training data, then model on the resulting components. This will reduce the number of features greatly, and guarantee no correlations between our new transformed features. This decision is mainly inspired by the transformations in Table II. Since each chemical property has many variations based on these transformations, some of which are linear, we expect many features to be explaining the same variance as other features. PCA would fix this problem at a loss of interpretability of relevant features. As I mentioned in the abstract as well as the background, we don't want to simply predict accurately, we would like to understand the relationships between critical temperature and the chemical properties.

The second would be through hand selecting features, perhaps through significance according to linear models or letting stepAIC() do some heavy lifting. stepAIC is a function in the MASS package of R that essentially, given a baseline model that is typically regressed on all features, will add/subtract features with every iteration and attempt to move down the AIC gradient to what we hope is a global minimum, guaranteeing a model with simplicity and significance. stepAIC is not limited to main effects, it can work on any level of interactions between features. However, including 2-way or even 3-way interactions may leave you with a more complicated model than you started with. This does not only have to serve as a form of modeling but also feature extraction. Other interpretable models that select relevant features are Lasso and Ridge regression.

Furthermore, I had mentioned Type I and Type II superconductors in the background section, however those labels do not exist in the data. To determine such a label, we must calculate the London Penetration Depth (λ_L) which specifies the depth of penetration of an incoming magnetic field. However, to calculate λ_L we must know certain properties of the charge and magnitude of the magnetic field, which cannot be derived from the data's variables alone. The distribution of T_c shows a bimodal distribution, indicating that there may be two clusters in the data. Understanding that Type II superconductors would have higher critical temperature, we may safely assume that the two clusters portray Type I and Type II superconductors, and can be cross validated with already known Type I and Type II superconductors. Further discussion regarding clustering in upcoming sections.

For the third, we could have a combination of the first and second. that is, use PCA as a preprocessing method and fit models on the principal components that explain most of the variance. PCA may be difficult to interpret the results with respect to the initial variables, but it doesn't leave us completely clueless. We can use the loadings of the PCA to indicate its correspondence with the initial variables, leaving us with something to infer about the initial un-transformed chemical properties.

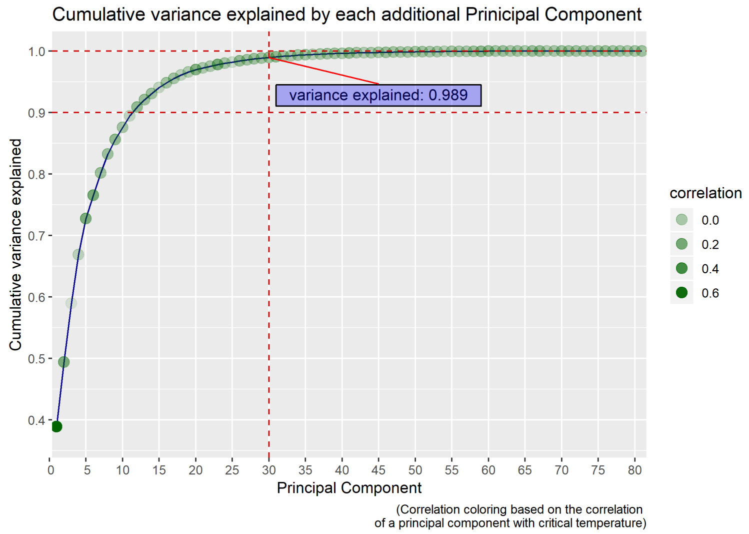
To choose the proper methodology, we may first run a regression on the un-transformed data and compare it to a regression on the transformed data on a basis of analysis of variance or simply the R^2 and AIC metrics. Based on the results, PCA may or may not be completely worth while. After that initial preprocessing decision is made, we can continue on to more complicated models such as deep neural networks and random forests.

Given the distribution of the target variable critical temperature, it's highly unlikely that any generalized linear model will fit well. The distribution is clearly not normal, leading me to consider transforming it to make the most of the regressions. We need to consider the posterior distribution of our data to even start considering a generalized linear model being a good fit. If we do not transform it, we would have to choose something like a gamma or inverse Gaussian distribution since we are dealing with a non-negative continuous target variable that is skewed right with a second smaller mode down the tail.

Lastly, some of the models presented work better (or worse) when the target varaible is continuous or categorical. We note that for a scientist, knowing the precise value of T_c may be just as beneficial as knowing a simple label for the T_c being very low, low, medium, high, very high, or super high. Since these models should be used as preliminary tests to see if creating a hypothetical superconductor is worth the time and cost. Hence, categorizing a superconductors T_c into the previously stated categories may be as useful as knowing the specific T_c $\pm\sigma^2$. Therefore, it may be fruitful to see the results of both.

V. Preprocessing and Feature Selection

As mentioned before, the data was preprocessed by K.Hamidieh. Hence, feature selection will be the main topic of this section. First and foremost, I run PCA to identify how much correlation between features there is in the data. To show the results of this, I plot the cumulative variance explained by adding each principal component in Fig. 4. As labeled in Fig. 4, the first 30 principal components explain 98.9% of the variance observed in the data. If this is used as preprocessing, then the features can be reduced from 81 to 30. Furthermore, the transparency of each point in Fig. 4 represents the Pearson correlation. We see that the first principal component is highly correlated with T_c and the third has almost no correlation with T_c.



(Correlation coloring based on the correlation of a principal component with critical temperature)

Fig. 4. Principal Component Analysis. The transparency of each point indicates the correlation of that principal component with T_c based on the scale in the legend on the right

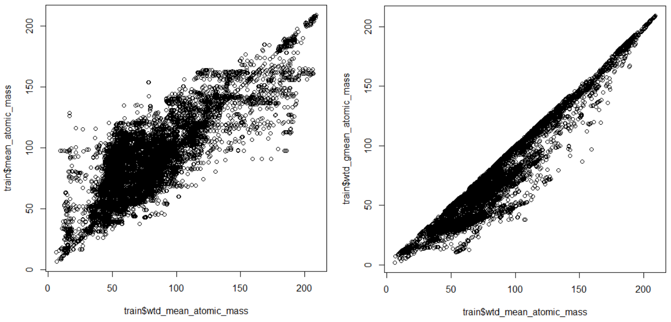


Fig. 5.
(left: $wtd\_mean\_atomic\_mass$ vs. $mean\_atomic\_mass$ cor: .82, right: $wtd\_mean\_atomic\_mass$ vs. $wtd\_gmean\_atomic\_mass$ cor: .96)

As to explain why such a high percentage of the variance is explained so quickly, refer to Fig. 5. The plots show $wtd\_mean\_atomic\_mass$ vs. $mean\_atomic\_mass$ and on the right $wtd\_mean\_atomic\_mass$ vs. $wtd\_gmean\_atomic\_mass$, with .82 and .96 correlation respectively. It is clear that some transformations made to create new variables are linearly correlated, hence PCA picks up on this pretty quick. Specific models require specific preprocessing, hence will be addressed in the next section.

To end this section, we will discuss clustering the data, The results of which were not so fruitful. Silhouette score was used for the metric to measure the ”goodness of fit” of the number of clusters. The number of clusters tested was in the range 2-10, using kmeans. The results are shown in Fig. 6.

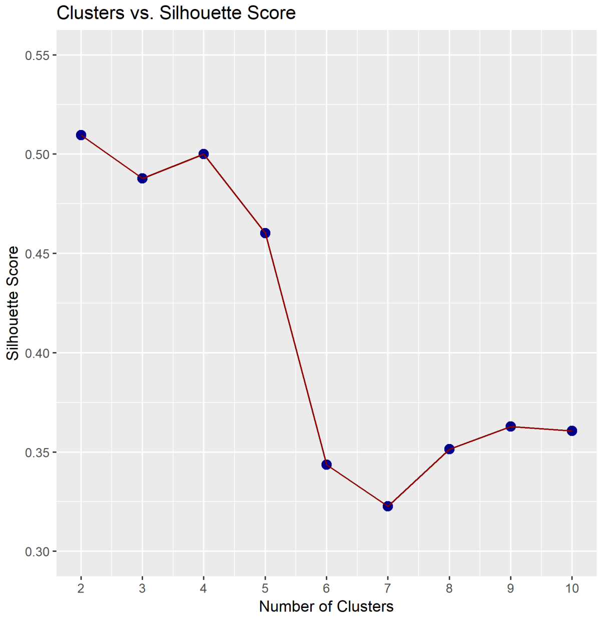


Fig. 6.

We can see that two clusters and four clusters are most likely to describe the structure of the data. However, without one being definite over the other, its difficult to choose one. I suspect that two and four work the best since, as stated in the background section, there are multiple different classifications of superconductors, all of which are binary. Hence, how can we discriminate between two clusters representing Type I and Type II, and Type I and II being sub-clusters of the four clusters. The most logical way of attempting to tell is to see the distributions of T_c for the two (or four) clusters. The label for Type I and Type II could have been very fruitful for analysis, but it’s not worth taking the risk of an estimated guess. Hence, clustering is not a part of the analysis.

As Mentioned in the methodology section, analysis would also be considering the categorical form of critical temperature. Fig. 7 shows the distribution of the factorized T_c. Factorization was based on the quantiles of the continuous variable T_c. This enables us to turn a complicated regression problem into a classification problem.

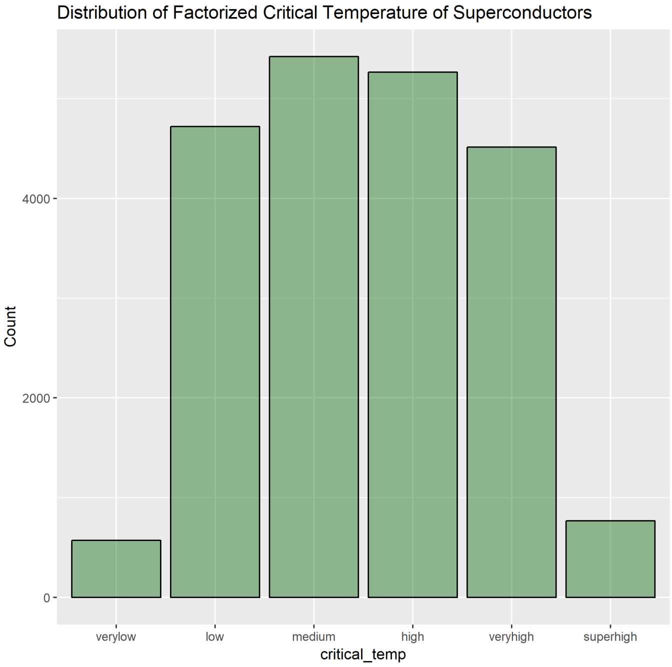


Fig. 7.

VI. Model Building

In this section, I lay out the structure and parameters of each model. Some models may have as many as three different layouts corresponding to the methodologies mentioned earlier. For example, I may present a random forest full of regression trees on the raw data, and on PCA data, as well as a random forest full of classification trees. Furthermore, depending on the models sensitivity to sparse data, all features were standardized to have $\mu = 0$ and $\sigma^2 = 1$. All models have and $80/20$ split for training and test set respectively.

A. Multiple Regression (Baseline)

There isn’t so much to be said about the baseline model, since it is simply a linear model regressed on all features. Recall that the distribution of T_c is by no means normal, however linear regression assumes normality. The general formula for regression is below.

$$T_c = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_{81} X_{81} + \epsilon$$

where X_i represents the i^{th} feature in the data which is a chemical property.

B. Step AIC

Much like the last model, there is not much more that one can say about stepAIC than what was already said in the methodology section. This model bases ”goodness” off of the AIC, throwing out variables until an optimal model is found that minimizes complexity and maximizes R^2. The baseline model is passed into the stepAIC function, so that it has all relevant features to work with so to speak.

TABLE III
RESULTS OF MODELS

Model	Raw R^2	PCA R^2	Raw RMSE	PCA RMSE
Baseline	.73	.68	17.55	19.13
Baseline (log)	.71	.67	18.70	20.01
Baseline (sqrt)	.77	.73	16.89	18.42
StepAIC (sqrt)	.77	.73	16.89	18.43
GBM	-	-	11.61	12.90
GBM (sqrt)	-	-	11.81	-
Random Forest	-	-	8.8	9.36
Random Forest (classifier)	acc: .98	-	-	-
Neural Network	-	-	11.53	12.06

C. General Boosted Regression

General Boosted models are a infamous type of ensemble method. Not only does it return a model slightly better than non-ensemble methods, but it also gives you a rank for which features were most influential to the target variable. The function in R is typically used for decision or regression trees.

D. Random Forest

The random forest implemented uses 5,000 separate regression or decision trees. However, soon after implementation I found that 100 trees produced the same result. This is the only model used for classification of T_c. As we will see in the next section, random forests have great performance. Random forests were computed using the ranger package in R.

E. Deep Neural Network

I implemented two neural networks of similar architecture on the raw data and on the PCA data. As this is a regression problem, most layers have rectified linear activation functions and have weights initialized according to the Gaussian distribution. Whereas in classification, one commonly uses softmax for their final output layer, a linear activation function is applied. The hope with using deep architecture is that it will learn the subtleties of how chemical properties relate to T_c.

More specifically, for the raw data the input layer size is 81, and for PCA data the layer size is 30. For each neural network, a dropout layer was added to prevent overfitting. For the PCA net, 4 layers were used, not including the output layer. The input layer, a dense layer of size 25, another dense layer of size 20, another dense layer of size 15, then the output layer of size 1. All layers had relu activation except the last layer which had linear activation. Every layer had weights initialized to the normal distribution.

Similarly the raw data network had 6 layers, not including the output layer. The layer sizes decreased somewhat linearly much like the PCA net. Again, all layers were relu except the last one, and all were initialized to the normal distribution.

This is the only model where scaling was done feature by feature before passing into the model. This is important preprocessing for neural networks as it puts the data in terms of activation for the neurons.

VII. RESULTS

As you will notice in Table III, for some of the multiple regression models, a transformation of the target variable is used. In this case it is log and square root. This is done in an attempt to somewhat normalize the distribution of T_c since this is an assumption of linear regression. Error is calculated with this transformation taken into account.

The first thing to notice is that stepAIC could perform no better than a linear regression with the square root transformation applied. It seems that stepAIC could not "decide" on what variables to throw out as they were somewhat significant to the outcome variable. Furthermore, throwing out a variable would reduce both AIC and R^2, however the algorithm found it more important to conserve the R^2. Second, we notice that of all the linear regression models, the one with the square root function applied to the outcome variable produces the best results. This is of no surprise since the histograms of the square root appeared more normal than the log or the original target T_c.

When comparing the GBM to the other regression models, we find that we lose the metric of R^2. On terms of root-mean-square-error, it out performs all other linear regression models. However, moving down the table still, we see that the random forest out performs all other regression models and the Neural Network. The Neural Network performs on par with our GBM. In terms of classification, random forests were the only algorithm used and it performs extremely well. Accuracy for classification was calculated through how many classifications made correct on the test set divided by the length of the test set.

One transformation that was not mentioned is scaling. As to combat any sparseness in the data, I ran models on scaled data, where each feature was z-scored. The results to which was identical to the raw data, indicating the data was not incredibly sparse.

A. *Brief Discussion of Results*

The hope of this paper was to out perform the models proposed in K.Hamidiehs's paper with the deep neural network, however it did not perform as expected. various combinations of layer sizes and number of layers was used and the best results were reported. However, in light of this, this paper serves to confirm and elaborate on K.Hamidieh's paper.

The models ran on data that was transformed through Principal Component Analysis did not perform as well as those that were ran on the raw data. This was a surprise since I expected that it was the right fit considering the large number of features and multicollinearity of the data.

The general boosted regression tree model reported finding that the range of thermal conductivity was the most significant factor for predicting T_c. Likewise, for the PCA analysis, The boosted model reported PC1 being the most significant predictor. This is understandable since, in Fig. 4, PC1 has the highest correlation with the target variable, hence most likely is highly influence by the range of thermal conductivity.

VIII. Related Works

The superconductivity data used was released to the public this year. Hence, the most prominent related work was K.Hamidieh's paper on this dataset. Many algorithms implemented served to cross reference his paper as well as attempted to optimize models that he presented. What Hamidieh did not do was use PCA as preprocessing, use a deep learning architecture, use stepAIC, or classify using a random forest. I was hoping for better performance using these approaches, however that was not the case since the best model implemented in this paper was no better than his proposed model. Other papers were used for fact checking the background section. Two of which were dissertations on theories of superconductivity. The last one is a website kept up since 1999 by Joe Eck, this was used for general history and uses of superconductivity.

IX. Conclusion

In conclusion, The results of this paper were no better than that of K.Hamidieh's. The new approaches that I presented in the last section were hoped to improve the accuracy of the model, but that was not what was found. The classification using random forests performed much better than expected considering it was a six class classification. The random forest regression model predicts the T_c with a root-mean-square-error of $8.8K$ on the raw data, and $9.36K$ for the PCA data. The high hopes for the neural network was misplaced as it performed on par with the general boosted model, both being second for best model implemented.

References

[1] K. Hamidieh, A data-driven statistical model for predicting the critical temperature of a superconductor. *IEEE Computational Materials Science*, vol. 154, pp. 346-354, Nov. 2018.

[2] J. Schmalian, Failed theories of superconductivity. *Department of Physics and Astronomy Iowa State University*, 1999

[3] S. Gustafson, Some Mathematical Problems in Ginzburg-Landau Theory of Superconductivity. *Graduate Department of Mathematics University of Toronto*

[4] J. Eck, Superconductors. *Superconductor History*, www.superconductors.org/ 1999