**Abstract**

**The objective of the case study is to identify what elements, characteristics, and exploratory variables correlate to superconductors by identifying relationships with high critical temperatures. This is done by regressing the critical temperature of many superconductors of varying elemental composition based on a range of quantitative material characteristics.**

**Introduction**

**A superconductor is a material that will exhibit extremely low to no electrical resistance at a sufficiently low enough temperature, resulting in infinite conductivity. This still little-understood phenomenon at which a material reaches the physical state of superconductivity occurs suddenly at a temperature known as the critical temperature. In 1911 the Dutch physicist Heike Kamerlingh Onnes discovered Mercury (Hg) as the first superconductor at a critical temperature of 4 degrees Kelvin (or -424 degrees Fahrenheit). Since then, researchers have discovered and developed many superconducting compounds, including metal alloys and ceramics, intending to find materials with higher critical temperatures (or ideally, those as close to room temperature as possible), as the requirement of extreme cooling is both expensive and challenging to achieve.**

**The dataset contains 21,263 observations of materials. The information for these observations includes the ratios of 77 different elements and 81 quantitative material characteristics of the compounds. The goal is to predict the Critical Temperature and find which combination of these variables can positively or negatively correlate most with increased critical temperature to identify valuable new superconductors' traits.**

**Methods**

**Dataset provided two files containing measurements of material characteristics and another with the elemental compositions. The two files were combined for use in this analysis along with the target response variable, critical temperature. There were no missing values in the data provided for this analysis. The team also identified a few outliers' (82 rows) in the clean-up process. The team worked on a model that will help determine which individual features contribute most to increases in critical temperature. Therefore, linear regression is best for this case as the response variable is continuous. The team split the data as a training set to build the model and a test set to validate the model's outcome.**

**After combining the two data files, the data set contained 158 features to start. This dataset was very complex to interpret and drive any kind of correlation and may result in over-fitting the model. After initial exploration of the dataset, the team built a linear model with all 158 features as a baseline. Followed by the linear model using a LASSO algorithm to select the most pertinent variables, and then Ridge regression to fit the model using the smaller set of variables chosen to reduce the chances of overfitting without further reducing the number of variables . The primary objective was to reduce the number of explanatory variables, to build a meaningful model and ensure the accuracy and explain ability of a dependable variable by the independent variable is not compromised.**

**The prime criteria to evaluate the model was to minimize the error (mean square error) and increase the explanation of the relationship between independent and dependent variables (r – square). The team also made a few manual adjustments to optimize (trade-off) the model's accuracy and enhance the model interpretation. The team investigated the correlations, checked the multi-collinearity, and kept the most relevant features.**

**Results**

**The coefficients representing the feature importance in the final linear model and the summary table with vital statistics are shown in the figures below.**

**The team started building the model with the notion of predicting the critical temperature based on material composition and properties given. Entropy, Thermal Conductivity and Atomic Radius has the high influence on critical temperature (show positive correlation) of superconductors to achieve superconductivity. Also, the presence of Barium (Ba), Calcium (Ca) and Bismuth (Bi) helps increase the critical temperature.**

**We observed an interesting result that only around nine (9) features and thirteen (13) materials, out of 158, dominate the outcome. This model could potentially provide the scientific community with some insight to narrow their research to a few key factors and elements.**