Predicting Superconducting Critical Temperatures with Supervised Machine LearningWork supported by the National Science Foundation

K. Kleinasser, Cornell University, Ithaca, NY*

CONTENTS

* Lycoming College, Williamsport, PA; klekirk@lycoming.edu

I. INTRODUCTION

I.1. Background on Superconductors

Superconductors are materials that lose all electrical resistance at low temperatures. These materials have a critical temperature (T_C) at which they lose their resistance. Most have very low cricial temperatures, but "unconventional superconductors" can have critical temperatures as high as room temperature under non-atmospheric conditions.

Electrons in superconductors form Cooper Pairs below their critical temperature. These pairs of electrons are held together with phonouns, which are atomic-level collective excitations. Phonouns are similar to photons in that they also have particle-like properties [citation here]. Unconventional superconductors are still not well understood and remain an open question in Physics. Understanding them could lead to the discovery of superconducting materials stable at room temperature under atmospheric conditions. Such a material would have large implications, such as super efficient electricity transfer and vast efficency improvements for applications like particle accelerators and powerlines.

I.2. Background on Matminer

Most superconducter databases do not include enough information to train an effective machine learning model, but such data can be extracted from the data they do provide. We use matminer to produce our features from the provided material data. Matminer is a python library that generates data from various measured properties of a material. Matminer collects existing calculations into a machine learning friendly python package. Our matminer workflow is shown in Figure 1.

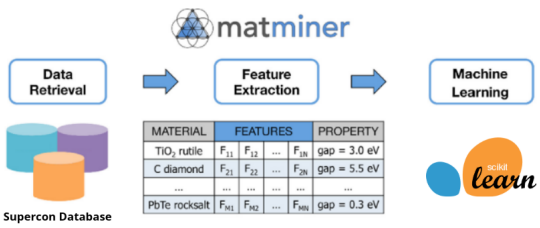

FIG. 1. Flowchart illustrating our matminer usage, modified from official matminer graphic [1].

Our database only provides the superconductor composition data. Matminer's featurizers can generate 53

features from the composition of a material. If we had band structure or other data, we could produce more information that we could use in our model.

[FLOWCHART FIGURE]

I.3. Background on Machine Learning

Previous papers have used random forest models to predict critical temperature [citation needed], but this paper will examine eight models before settling on two for further investigation. These models are descibed below. Each model's hyperparameters[2] was optimized with Scikit-Learn's GridSearchCV, which tests combinations from a grid of hyperparameters and returns the best performing model based on a specifed metric.

We started our model search with some linear models. Besides the base Linear Regression model, we used linear (and polynomial) Support Vector Regression (SVR) models. SVR uses decision boundaries, which are lines parallel to the regression line. The model aims to maximize the amount of data within the decision boundaries and has hyperparamters to modify sensitivity to prevent overfitting.[3] We also trialed Elastic Net and Bayesian Ridge models. Elastic Net uses L1 and L2 penalties to stabilize the model, and Bayesian Ridge uses probablity distributors instead of point estimates.

Additionally, we trialed Decision Tree and KNeighbors (KNN) models. Decision trees are very interpretable - they break predictions into nodes of the tree, eventually leading to a prediction value. These trees can be represented graphically and show how they produce results, unlike most machine learning models. KNN models are a little different, they store all the data and predict values based on a similarity measure. The model looks at a specified number of similar neighbors to produce a prediction.

Finally, we tried multiple ensemble models - Random Forest Regression (RFR), Extra Trees, and a superlearner. RFR models use numberous decision trees and subsamples the data with replacement. This means that the model replace data after using it in a subset. Extra Trees is like RFR, but it does not replace the data after use in a subset. The final ensemble model we tested is a superlearner, a model that can combine mutliple high-scoring Scikit-Learn model predictions and sometimes improve the performance from the indvidual models.

II. METHODOLOGY

We evaluated our models using several metrics - R2 scores for regression evaluation, Mean Squared Error (MSE) and Mean Absolute Error (MAE) for error evaluation, and prediction intervals for uncertainty evaluation.

III. FORMATTING

III.1. Preprint, reprint, and twocolumn options

REVTEX 4.2 offers a `reprint` class option to typeset a manuscript in a format that is a close approximation to the actual journal's appearance. It should be emphasized that this is only an *approximation*; a manuscript may be substantially different in length or appearance after it goes through our production process. This is mostly due to the choice of fonts and the scaling of figures.

REVTEX 4.2 is designed to make it straightforward to switch between two-column and single-column formatting just by changing the class option. Authors may submit with either the `reprint` or the `twocolumn` class options. The `preprint` primarily does three things: It increases the font size to 12pt, increases the line spacing, and changes the formatting to single column.

III.2. Paper size

Manuscripts should be submitted to APS formatted for letter size paper. Papers are sent electronically to referees who may want to print them out. Letter size formatting ensures that this will be trouble free for all referees.

IV. MARKING UP FRONT MATTER

Perhaps the most important macros are those pertaining to the markup of the front matter (title, authors, affiliations, abstract, etc.). Note that proper use of the REVTEX 4.2 macros means that explicit centering environments in the front matter are not needed and should not be used.

IV.1. Title

The title of the manuscript should be specified using the `\title` macro. A double backslash `\\` may be used to force a line break in a long title.

IV.2. Authors, affiliations, and collaborations

REVTEX 4.2 makes it straightforward to input author names and link them up properly with affiliations. Authors should let REVTEX 4.2 do the work of grouping authors and affiliations and, if using the superscript style, numbering affiliations. Please follow these guidelines:

- Use a single `\author` macro for each author's name. REVTEX 4.2 automatically puts in all commas and the word 'and.'

- Use the \surname macro to explicitly indicate if an author's family name consists of more than one name or if the family name is not the author's last name.
- The \email macro may be used to specify an author's e-mail address. The \thanks macro must not be used for this. Only the e-mail address itself may appear in the macro's required argument.
- The \homepage macro may be used to specify a URL associated with an author. The \thanks macro must not be used for this. Only the URL may appear in the macro's required argument.
- The \altaffiliation macro may be used to specify an alternate affiliation or temporary address for an author. The \thanks macro must not be used for this. Only the affiliation may appear in the macro's required argument.
- The \thanks macro may be used only if one of the more specific macros list above does not apply.
- Use a single \affiliation for each affiliation.
- Superscripts linking authors to affiliations must be accomplished using the superscriptaddress class option rather than putting in explicit superscripts by hand.
- A collaboration may be specified by using the \collaboration macro. The \author macro must not be used for collaborations.

IV.3. Abstract

The abstract must be specified using the abstract environment. Note that in REVTEX 4.2, the abstract must come before the \maketitle command. REVTEX 4.2 now allows the the use of the description environment within the abstract to provide *structured abstracts*. For instance, *Physical Review C* would like authors to provide abstracts with sections summarizing the paper's **Background**, **Purpose**, **Method**, **Results**, and **Conclusions**. This can be accomplished in the following manner:

```
\begin{abstract}
\begin{description}
\item[Background] This part would describe the
context needed to understand what the paper
is about.
\item[Purpose] This part would state the purpose
of the present paper.
\item[Method] This part describe the methods
used in the paper.
\item[Results] This part would summarize the
results.
\item[Conclusions] This part would state the
conclusions of the paper.
\end{description}
\end{abstract}
```

V. REFERENCES AND FOOTNOTES

Authors are strongly encouraged to use BibTEX when preparing their bibliographies. If BibTEX is used, current production processes require that the .bbl file be included directly into the manuscript's main .tex file. REVTEX 4.2 comes with two BibTEX style files for formatting references, one for the *Physical Review* journals and one for *Review of Modern Physics*. In 4.2, the BibTEX styles have been modified to display journal article titles in the bibliography.

The following apply whether BibTEX is used or not.

- Authors should use the \cite and \bibitem commands to create bibliographies and to refer to items in the bibliography. "By hand" numbering of references should be avoided.
- REVTEX 4.2 provides new syntax for combining multiple citations into a single entry in the bibliography and for putting extra text before and after a reference. Please refer to *REVTEX 4.2 Author's Guide* included with the REVTEX 4.2 distribution for full details.
- Footnotes must be specified using the \footnote macro. REVTEX 4.2 will place the footnotes in the bibliography for the *Physical Review* journals. Please note that even if you don't use BibTEX, you may have to run BibTEX to get the footnotes to appear. Footnotes giving additional information about authors (such as e-mail addresses) must not be specified using the \footnote macro (see Section IV.2).
- Avoid custom footnotes using \footnotemark and \footnotetext [except in the context of tables (see Section IX.5)].
- References should be formatted and specified according to the *Physical Review Style Guide*. Note that using BibTEX automatically ensures this.
- URLs should be specified using the \url macro. BibTEX will automatically take care of this if the url field is used.
- E-print identifiers should be included using the \eprint macro. BibTEX will automatically take care of this if the eprint field is used.

Please see the REVTEX 4.2 Author's Guide for new features in REVTEX 4.2's APS BibTEX styles, including support for citing data sets, journals that use DOIs in place of page numbers, and journals that use year and issue instead of volume to uniquely identify articles.

VI. BODY OF THE PAPER

VI.1. Sectioning and cross-referencing

For sectioning a manuscript, the basic rule is to use the appropriate sectioning commands (\section, \subsection, \subsubsection, *etc.*). Cross-referencing a section must be done by using the proper \label and \ref commands. Cross-referencing by hand is not allowed. \part, \chapter, and \subparagraph should not be used.

VI.2. Appendices

Appendices should be specified using the \appendix command which specifies that all following sections create with the \section commands are appendices. If there is only one appendix, then the \appendix* command should be used instead.

VI.3. Acknowledgments

Any acknowledgments should be included by using the acknowledgments environment. Note that in REVTEX 4.2, this is an environment and not a command.

VI.4. Counters

No counters may be created and the standard ones may not be altered. If an exceptional label is needed for an equation, the \tag command (requires the amsmath class option) should be used. Please note that the use of the \tag command may conflict with the use of the hyperref package due an incompatibility between amsmath and hyperref.

VI.5. Fonts

It is preferable to avoid the older TEX and LATEX 2.09 macros for controlling fonts such as \rm, \it, *etc.* Rather, it is better to use the macros introduced in LATEX 2_ε. If the older font commands are used (they really should be avoided!), be sure to use curly braces to properly limit the extent of the font change. {\bf ...} is the correct method. Commands for controlling text and math font changes are summarized in Table I.

Bold Greek letters and other bold math symbols should be accomplished with the use of bm.sty which is distributed as a required tool with the latest versions of LATEX 2_ε and should be loaded via \usepackage{bm}. This package introduces the \bm macro. Some bold characters may require using the amsfonts class option.

New fonts may not be declared with \newfont. Font attribute commands for selecting a font family, shape, and series are all disallowed; the standard LATEX 2_ε font selection macros list above should be used instead.

Finally, the \symbol macro is also not allowed.

TABLE I. LATEX 2_ε and AMS-LATEX font summary.

\textit	Italics. Replaces \it
\textbf	Bold face. Replaces \bf
\textrm	Roman. Replaces \rm
\textsl	Slanted. Replaces \sl
\textsc	Small caps. Replaces \sc
\textsf	Sans serif. Replaces \sf
\texttt	Typewriter. Replaces \tt
\textmd	Medium series
\textnormal	Normal
\textup	Upright
\mathbf	Bold face
\mathcal	Replaces \cal
\mathit	Italics
\mathnormal	Replaces \mit
\mathsf	Sans serif
\mathtt	Typewriter
\mathfrak	Fraktur: Requires amsfonts or amssymb class option
\mathbb	Bold blackboard: Requires amsfonts or amssymb class option
\bm	Bold Greek and other math symbols: Requires \usepackage{bm} and may require the amsfonts class option

VI.6. Environments

VI.6.1. Lists

The standard list environments itemize, enumerate, and description are allowed. The \item macro with or without the optional argument is also allowed. Customization of the list environments (with macros such as \labelstyle, \labelitemi, \labelenumi, \itemsep, etc.) is allowed but may be ignored in production. Generalized lists (\begin{list}) and trivial lists (\begin{trivlist}) are not allowed.

VI.6.2. Other Environments

Creating generalized new environments with \newenvironment is not allowed. Creating a new theorem environment with \newtheorem is allowed though.

The tabbing environment and the macros \=, \>, \‘, and \’ are allowed but may be ignored in production. Conversion programs used in production should recognize

the escapes \a=, \a', and \a` for using the corresponding accents within a tabbing environment though.

The verbatim environment is allowed.

VI.7. Boxes

Most boxes and macros to manipulate them are not allowed. These include \raisebox, \parbox, \minipage, \rulebox, \framebox, \mbox, \fbox, \savebox, \newsavebox, \sbox, \usebox, and the environment \begin{lrbox}. Rules produced with \rule are not allowed.

VI.7.1. Margin Notes

Margin notes created with \marginpar are not allowed, as are the associated style parameters \marginparwidth, \marginparsep, and \marginparpush.

VII. MATH MARKUP

In general, all math markup and the standard math environments from LaTeX 2$_\varepsilon$ are allowed. These include \begin{math}, \begin{displaymath}, \begin{equation}, \begin{eqnarray}, and \begin{eqnarray*}. The shortcuts $, $$, \[, and \] are allowed. In addition, authors may use almost all of the additional markup introduced by AMS-LaTeX by using the amsmath class option. The explicit exceptions are \genfrac, \boxed, and \smash. The markup contained in amsextra and amsthm may not be used though. Commutative diagrams created with the amscd package are acceptable.

VIII. FIGURES

VIII.1. Figure inclusions

Figures should be included into a REVTeX 4.2 manuscript by using the standard LaTeX 2$_\varepsilon$ macros. LaTeX 2$_\varepsilon$ includes several powerful packages for including the files in various formats. The two main packages are graphics and graphicx. Both offer a macro called \includegraphics; they mainly differ in how arguments for controlling figure placement (*e.g.*, scaling and rotation) are passed to the \includegraphics.

The figure environment should be used to add a caption to the figure and to allow LaTeX to number and place the figures where they fit best. If a figure needs to be referred to in the text, rather than manually numbering the figures a \label should be added to the figure environment (best practice is to put the label within the argument of the \caption command) and the \ref macro should be used to reference this label. Figures that span the page should use the \figure* environment. The picture environment must not be used directly (one can include an Encapsulated PostScript figure that was produced using the picture environment of course).

VIII.2. Figure placement

Figures should be placed as close as possible to the point where they are first referenced. There is no need to place all figures separately at the end of the manuscript and it is preferred that authors leave the figures in their natural locations. Authors may also find useful the REVTeX 4.2 floatfix class option which adds emergency float placement processing to avoid "stuck" floats which would otherwise be deferred to the end of the job (and can lead to the fatal ``Too many unprocessed floats'' message).

IX. TABLES

The standard LaTeX 2$_\varepsilon$ table formatting environments are supported as is the use of the longtable package. Tables may be reformatted during production to meet APS style guidelines. Here are some helpful hints for trying to get tables formatted correctly:

- Use the longtable package to get tables to break across pages.
- The macro \squeezetable will reduce the font size of the table. This macro must occur within a group outside the table environment. The proper markup is:

```
\begingroup
\squeezetable
\begin{table}
...
\end{table}
\endgroup
```

- Try using the float placement option H which will enable LaTeX to break a float across pages. Long tables are more attractively set with longtable however.

```
\begin{table}[H]
\begin{ruledtabular}
\begin{tabular}
...
\end{tabular}
\end{ruledtabular}
\end{table}
```

IX.1. Doubled rules and table formatting

REVTeX 4.2 provides the ruledtabular environment which automatically puts the scotch rules (double lines)

around tables and formats all enclosed `tabular` environments to the full width of the tables and improves intercolumn spacing. This environment should be used whenever possible.

IX.2. Wide tables

When typesetting using `twocolumn`, tables can either span a single column or both columns. Using the '`*`'-ed version of the `table` or `longtable` environments produces wide tables that span the columns.

Tables that are very wide and that may be better typeset in a landscape orientation (rotated 90 degrees) should be enclosed in a `turnpage` environment. This will place the rotated table on its own page. Note that some dvi previewers may not be able to show the table properly, but `dvips` and `pdflatex` work correctly.

IX.3. Table placement

Tables should be placed as close as possible to the point where they are first referenced. There is no need to place all tables separately at the end of the manuscript and this is not desirable for APS purposes. The class option `floatfix` may be helpful for table placement as well as figure placement (see Section VIII.2).

IX.4. Aligning columns on a decimal point

The standard LaTeX 2_ε macro package `dcolumn` should be used to accomplish this.

IX.5. Tablenotes

Footnotes in tables (tablenotes) should use the `\footnote` macro. However, if more than one reference to the same footnote is needed, authors may use `\footnotetext` and `\footnotemark`. This will produce notes (labeled by lower-case roman letters) inserted below the table rather than in the reference section or at the bottom of the page.

X. AUTHOR-DEFINED MACROS

Authors may define convenience macros to save keystrokes. This means that the macros may not invoke TeX macros such as `\if` or other context dependent commands. Also, LaTeX 2_ε provides three macros for declaring new commands: `\providecommand`, `\newcommand`, and `\renewcommand` (as well as their '`*`'-ed versions). These should be used. Authors may not use TeX's low-level commands `\def`, `\edef`, and `\gdef`.

XI. SUMMARY

To ensure the best use of TeX manuscripts, authors need to follow the guidelines specified here. Use of low-level formatting commands to finely control horizontal and vertical spacing may be ignored during production, or even worse, make it impossible to convert the manuscript to XML. Authors should keep things as simple as possible and correctly use the proper REVTeX 4.2 or LaTeX 2_ε macros. Any questions about usage may be directed to `revtex@aps.org`.

[1] L. Ward, A. Dunn, A. Faghaninia, N. E. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K. A. Persson, G. J. Snyder, I. Foster, and A. Jain, Computational Materials Science **152**, 60 (2018).

[2] Hyperparameters are machine learning parameters that change how a model is trained.

[3] Overfitting occurs when a model is trained to be too specific to a particular dataset and is not generalizable.