

Copper Alloy Discovery System using Machine Learning Techniques (COMP4560)

A thesis submitted in partial fulfilment of the degree of
Bachelor of Advanced Computing (Honours)

Submitted by Abhinav Pandey

Supervised by Professor Nick Birbillis

Examiner – Dr. Charles Martin



College of Engineering and Computer Science
The Australian National University

Submitted in June 2020

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of the author's knowledge, it contains no material previously published or written by another person, except where due reference is made in the text.

Abhinav Pandey
15 June 2020

Acknowledgements

This section is to express my sincerest gratitude to all those who supported the completion of this paper. Firstly, I would like to thank Professor Nick Birbillis, Deputy Dean of the College of Engineering and Computer Science at the Australian National University for supervising this project. His words of encouragement motivated me to push myself and teach myself all that was necessary to complete this project. I also extend my heartfelt appreciation to my team member, Venkata Prakash, without whom this project would not have achieved what it does now. I would also like to thank Dr. Weifa Liang for the support that he provided in the successful completion of this project. Finally, I would like to thank the Australian National University for providing me with this wonderful opportunity to grow academically and helping me develop an ability to face any challenges come my way.

ABSTRACT

Copper and copper alloys are some of the most versatile engineering materials available^[4] and their ability to freely form alloys opens a potential to get excellent physical properties such as high tensile strength, conductivity (thermal and electrical), ductility, corrosion resistance and more. There exist more than 400 copper alloys already, each with a unique combination of properties^[4]. The development of new copper alloys, however, relies largely on a combination of expert judgment, trial and error and intuition, which makes the process slow and expensive^[2]. This indicates a requirement of system that can quickly and reliably recommend compositions and processing conditions for targeted property values. This study aims to develop a system for generating copper alloys with user-defined physical properties viz. tensile strength and thermal conductivity using Random Forest models. The results of the study indicate that the system is able to generate copper alloys compositions that are not present in the training data.

CONTENTS

Table of Contents

ABSTRACT.....	1
Chapter 1: Introduction	7
1.1 Introduction	7
1.1.1 Predictive Modelling	7
1.1.2 Composition Generation.....	7
Chapter 2: Background	8
2.1 Copper Alloys and Some Properties	8
2.1.1 Tensile Strength :	8
2.1.2 Thermal Conductivity :	8
2.1.3 Electrical Conductivity :	9
2.2 Machine Learning.....	9
2.2.1 Supervised Learning	10
2.2.2 Regression	10
2.2.3 Decision Tree Regression	10
2.2.4 Random Forest Regression	11
2.3 Overview of the Copper Alloy Dataset.....	11
2.4 Data Pre-processing	12
2.5 Model Selection/Evaluation.....	12
2.5.1 Train Test Split.....	12
2.5.2 Overfitting Example	13
2.5.3 10-Fold Cross Validation	13
2.6 Hyperparameter Tuning and Some Approaches.....	15
2.6.1 Grid Search.....	15
2.6.2 Random Search	15
Chapter 3: Methodology.....	16
3.1 Data Collection	16
3.1.1 Data Overview :	16
3.2 Data Pre-processing	17
3.3 Train Test Split.....	17
3.4 Model Selection	17
3.5 Hyperparameter Tuning.....	19
3.6 Model Evaluation	21
3.7 Inverse Modelling	21

Chapter 4: Results and Analysis	22
4.1 Prediction Performance	22
4.2 Analysis of Features' Importance:	23
4.3 Results of SAGS system :	24
Chapter 5: Conclusions and Future Work	25
5.1 Conclusions	25
5.2 Limitations.....	25
5.3 Future Work	26

Table of Figures

Figure 1. Regression using Decision Trees	11
Figure 2. Regression using Random Forests	11
Figure 3. Visualizing Train-Test Split of a dataset	13
Figure 4. Demonstration of overfitting by regression models (blue and red lines) on data (dots)	13
Figure 5. Visualising 10-fold cross validation	14
Figure 6. Performance of different machine learning models using 10-fold cross validation (left) and prediction performance on test data (tensile strength)	18
Figure 7. Performance of different machine learning models using 10-fold cross validation (left) and prediction performance on test data (for thermal conductivity)	19
Figure 8. Comparison of the cross-validation performance of the model with default hyperparameters with the best models obtained after hyper parameter tuning using Grid Search and Random Search (for tensile strength)	19
Figure 9. Demonstration of lack of performance improvement after hyperparameter tuning (for tensile strength).....	20
Figure 10. Comparison of the cross-validation performance of the model with default hyperparameters with the best models obtained after hyper parameter tuning using Grid Search and Random Search (for thermal conductivity).....	20
Figure 11. . Demonstration of lack of performance improvement after hyperparameter tuning (for thermal conductivity).....	21
Figure 12. Predicted vs Actual Values of our models	22
Figure 13. Visualising feature importance for tensile strength	23
Figure 14. Visualising feature importance for thermal conductivity	24

Chapter 1: Introduction

1.1 Introduction

Copper and copper alloys are some of the most versatile engineering materials available.^[4] Pure copper has the best electrical and thermal conductivity of any commercial metal and forms alloys more freely than other metals which makes it suitable for a wide range of applications^[4]. The ability to freely form alloys opens a potential to get excellent physical properties such as high tensile strength, conductivity (thermal and electrical), ductility, corrosion resistance and more. These properties can be further enhanced with variations in composition and manufacturing processes.^[5]

There exist more than 400 copper alloys already, each with a unique combination of properties^[4]. The development of new copper alloys, however, relies largely on a combination of expert judgment, trial and error and intuition, which makes the process slow and expensive^[2]. This indicates a requirement of system that can quickly and reliably recommend compositions and processing conditions for targeted property values.

The main objective of this study is to propose an alloy design system (referred as Smart Alloy Generation System (SAGS) from here) that utilizes machine learning to obtain multi-element copper alloy compositions and processing conditions for a desired value of either ***Tensile Strength*** or ***Thermal Conductivity*** (referred as relevant properties from here).

The development of the SAGS system can be broadly divided into broad processes –

1.1.1 Predictive Modelling

Two distinct Random Forest models, one for tensile strength and one for thermal conductivity predictions, are built to learn the relationship between tensile strength, thermal conductivity the compositions of copper alloys and their processing conditions using samples obtained from multiple sources including public databases and educational packages.

1.1.2 Composition Generation

The models obtained in the previous step are used to generate candidate alloy compositions (and processing conditions in the case of tensile strength) for a user-defined value of either of the relevant properties. This is done by making property predictions on synthetically generated composition sample space using combinations of the compositions of existing alloys.

The proposed system demonstrates the ability to generate suitable compositions for a given value of the relevant properties and shows potential to alter the conventional practices of alloy discovery.^[2]

Chapter 2: Background

2.1 Copper Alloys and Some Properties

Copper alloys are metal alloys that have copper as their principal component^[9] and are useful in a wide range of applications in the electrical industry because of their desirable physical properties such as high conductivity, strength, ductility, and resistance to corrosion.

Some important properties of materials that are relevant to this study are -

2.1.1 Tensile Strength :

The tensile strength of a material is a measure of the maximum stress that it can withstand while being stretched or pulled before breaking. In technical terms, the tensile strength of a material is the force per unit area at which it breaks in two.

Tensile strength is measured in units of force per unit area. The unit is newton per square meter (N/m^2), kilogram (force) per square centimeter (kg/cm^2) or kilopounds per square inch (ksi).^[10]

The tensile strength of an alloy is dependent not only the composition of the alloy but also the processing methods used during manufacturing of the alloy, among other things.

2.1.2 Thermal Conductivity :

The thermal conductivity refers to the intrinsic ability of a material to transfer or conduct heat. In technical terms, it is defined as the amount of heat per unit time per unit area that can be conducted through a plate of unit thickness of a given material, the faces of the plate differing by one unit of temperature.^[11] In the International System of Units (SI), thermal conductivity is measured in watts per meter-kelvin ($\text{W/(m}\cdot\text{K)}$). In imperial units, thermal conductivity is measured in $\text{BTU}/(\text{h}\cdot\text{ft}\cdot^\circ\text{F})$.

The thermal conductivity of an alloy, at a given temperature, is fixed for a given composition. However, it is independent of processing methods, unlike tensile strength.

2.1.3 Electrical Conductivity :

The electrical conductivity of a material refers to its ability to conduct electricity. In technical terms, it is the ratio of the current density in the material to the electric field which causes the flow of current.^[13]

According to the Wiedemann–Franz law , at a given temperature the electrical conductivity of a material is directly proportional to its thermal conductivity.^[12] Since, there is a direct relationship between thermal conductivity and electrical conductivity, we will be focusing on modelling only a) the tensile strength and (b) the thermal conductivity of the copper alloys in this study.

2.2 Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.^[15]

Machine Learning tasks can be broadly divided into the following few categories :

- a. Supervised Learning
- b. Unsupervised Learning
- c. Semi-Supervised Learning
- d. Reinforcement Learning

2.2.1 Supervised Learning

The objective of supervised learning algorithms is to infer a function from labeled training data consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). These training data and their labels are used to produce an inferred function, which can be used for mapping new examples.^[16] In an optimal scenario, a supervised learning algorithm will then be able to correctly determine the class labels for unseen instances.

Supervised Learning can be further divided into two broad class of problems –

- a. Regression
- b. Classification

2.2.2 Regression

If the labels provided to training examples in a problem are discrete values of a variable, then the problem belongs to category regression or prediction. In other words, regression finds a function that takes an input vector and maps it to a discrete value.

There are several regression models available today, however we will limit our discussion only to those models which are used in this study viz. Random Forest Regressors. But before we discuss Random Forests it is important to understand their building blocks which are Decision Tree Regressors.

2.2.3 Decision Tree Regression

Decision Tree Regression is a predictive modelling approach in machine learning which uses decision trees to go from observations about a training item (represented as branches) to model and predict a real-valued target label. They are among the most popular techniques because of their simplicity and interpretability.^[17]



Figure 1. Regression using Decision Trees

2.2.4 Random Forest Regression

Random Forest Regression is another powerful predictive modelling approach that builds up on the concept of decision trees and tries to address the problem of overfitting associated with them.^[18] A Random Forest Regressor is an ensemble bagging approach, which means that it creates several decision tree regressors at training time and its prediction is the mean value of the all the decision tree predictions.

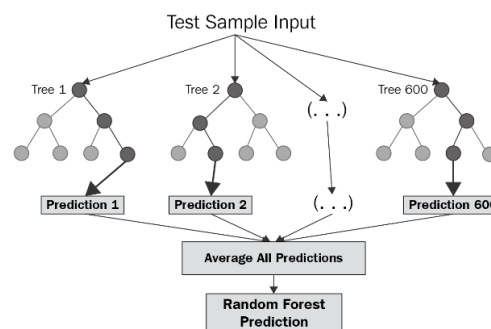


Figure 2. Regression using Random Forests

2.3 Overview of the Copper Alloy Dataset

The data used in this study has been collected from multiple sources including the public databases^[14], research papers and educational software.

The dataset contains a total of 34 columns that contain the following information for each alloy instance –

- 2 Identifier Columns viz. *Alloy Name* and *Temper Code*
- 2 columns for Processing Conditions viz. *Form* and *Temper*
- 28 columns for constituents of the alloy (each of these columns contains the contribution of the element to the alloys' percentage by mass composition)
- Tensile Strength - measured in kilopounds per square inch (ksi) at room temperature (68 °F)
- Thermal Conductivity – measured in BTU/(h·ft·°F) at room temperature (68 °F)

2.4 Data Pre-processing

Data pre-processing is an integral step in Machine Learning as the quality of data and the useful information that can be derived from it directly affects the ability of our model to learn; therefore, it is extremely important that we pre-process data before feeding it into our model.^[19] Some common data pre-processing are –

- Handling missing values
- Transformations such as data normalization
- Encoding categorical data
- Handling duplicate data

2.5 Model Selection/Evaluation

2.5.1 Train Test Split

Train-test split is an essential part of the data science process where we train models on certain part of the data and test the model on unseen data to verify its generalization capabilities. Some models can have amazing scores on training data but poor performance on test data. This is an indication of overfitting by the model on the

training data. In other words, it indicates that the model has trained on the noise in the data and memorized it

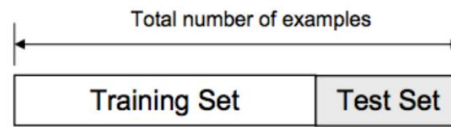


Figure 3. Visualizing Train-Test Split of a dataset

2.5.2 Overfitting Example

An example of overfitting can be seen in Fig. where we can compare the predictions of two different models represented by the red and blue lines on the data (represented by the dots).

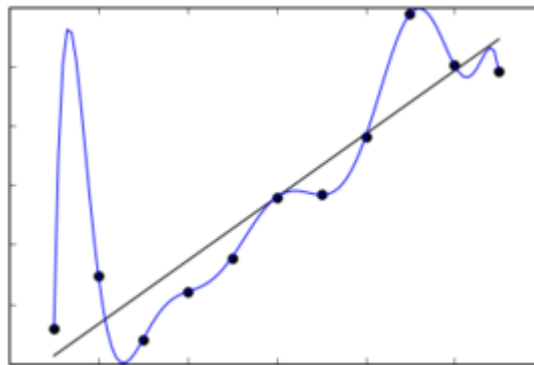


Figure 4. Demonstration of overfitting by regression models (blue and red lines) on data (dots)

The function identified by model represented by the red line indicates a good fit to the training data and indicates good generalization capabilities. On the other hand, the model represented by the blue line has determined a complex function which, although, predicts all the training data correctly, is likely to fail in predicting unseen data. The blue line model is an overfitted model.

2.5.3 10-Fold Cross Validation

K-Fold Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. It is one of the most reliable methods for evaluating and comparing model performances. In this study we use $k = 10$ since it has been

found through experimentation to generally result in a model evaluation metric with low bias a modest variance.^[20]

The general procedure is as follows and can be visualised as in Figure 5.:

- Shuffle the dataset randomly.
- Split the dataset into k groups
- For each unique group:
 - Take the group as a hold out or test data set
 - Take the remaining groups as a training data set
 - Fit a model on the training set and evaluate it on the test set
 - Retain the evaluation score and discard the model
- Summarize the skill of the model using the sample of model evaluation scores

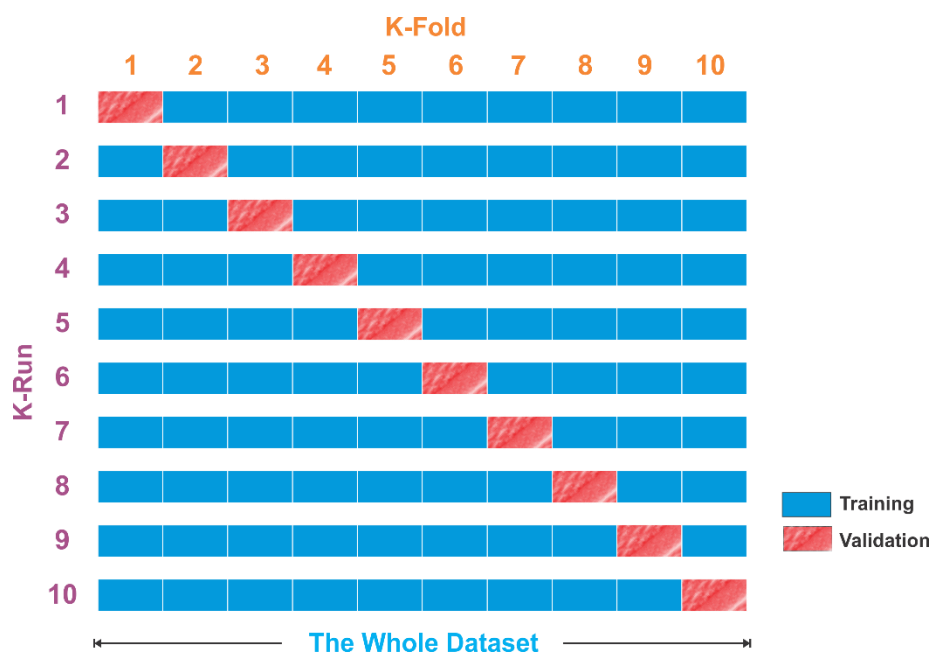


Figure 5. Visualising 10-fold cross validation

2.6 Hyperparameter Tuning and Some Approaches

Hyperparameter tuning is defined as choosing a set of hyperparameters^[21] for a given machine learning model to optimise model performance.

2.6.1 Grid Search

Grid search is a traditional way to perform hyperparameter optimization. It works by searching exhaustively through a specified subset of hyperparameters.^[21]

Strength - *Guaranteed to find the optimal combination of parameters supplied.*

Weakness - *Can be very time consuming and computationally expensive.*

2.6.2 Random Search

Random search differs from grid search mainly in that it searches the space of hyperparameters randomly instead of exhaustively.^[21]

Strength - *Decreased processing time*

Weakness – *Does not guarantee to find optimal solution from the specified search space*

Chapter 3: Methodology

3.1 Data Collection

Since, our objective for the project is to build supervised learning models to predict the tensile strength and thermal conductivity of alloys, the first and foremost step is to create a reliable dataset.

As discussed earlier, the tensile strength of an alloy varies with multiple factors such as chemical composition, form factors and temper factors. The thermal conductivity of an alloy is fixed for a given chemical composition, however a difference in composition will almost always bring a change in thermal conductivity. Therefore, our dataset should contain features that incorporate all the above-mentioned information for each alloy.

The data used in this project was collaboratively compiled from multiple sources including the web, educational software and past research papers. These sources are Copper Org website, CES Edu pack and past research papers.

3.1.1 Data Overview :

Each alloy in the dataset is represented by a single row of the csv file represented by the following features :

1. Name (Identifier)
2. Temper Conditions
3. Form Conditions
4. Mass Percentage Composition of alloys –
Named as the corresponding chemical symbol of each element in the periodic table. eg. 'Cu' column contains the percentage (by mass) of copper (Cu) in each alloy, 'Al' column contains the percentage of aluminium (Al) and so on.
5. Tensile Strength (ksi units) - Measured at room temperature, 68°F (20°C)
6. Thermal Conductivity ($Btu/sq\ ft/ft\ hr/^{\circ}F$) - Measured at room temperature, 68°F (20°C)

The data that has been collected has been made such that there are no missing values for any of the alloys. If an alloy does not contain an element, the feature column corresponding to that element is set as 0.

3.2 Data Pre-processing

The data was pre-processed for our supervised modelling step using the following steps:

- Remove the identifier column (Name):
Since, identifiers like name do not have any effect on the tensile strength or the thermal conductivity of an alloy, we will remove them from the dataset.
- Encode categorical variables (Temper and Form):
This was done using the concept of dummy variables, which is a suitable method of encoding categorical variables for machine learning models .
- Removing duplicates of alloys from the dataset:
For those alloys (if any) which occurred in the dataset more than once we retained only a single instance of the alloy.
- Removing Temper and Form columns (only for Thermal Conductivity prediction):
Since, the Thermal Conductivity of an alloy is only dependent on its chemical composition, columns containing other information (Temper and Form) were also removed for the dataset compiled for thermal conductivity.

3.3 Train Test Split

The data was divided into train and test sets. For the following sections (3.4 and 3.5) only the training dataset has been used, while the test dataset is saved for the final evaluation.

3.4 Model Selection

For Tensile Strength

The following regression models from scikitlearn's machine learning library were used to predict the tensile strength of copper alloys using their composition, form, and temper:

- Random Forests
- Linear Regression
- Decision Trees
- K-Nearest Neighbours
- Multi-Layered Perceptron
- XGBoost

Figure 6 shows a comparison of their cross-validation performances on the training data.

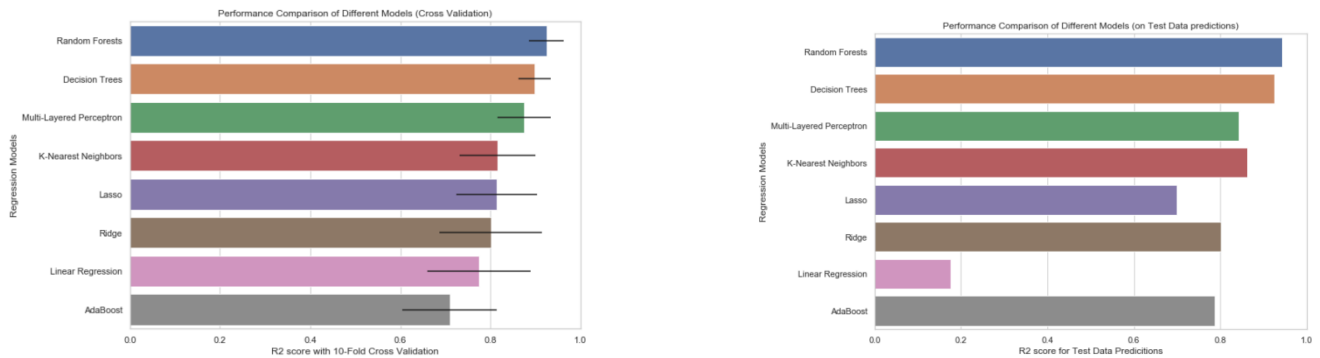


Figure 6. Performance of different machine learning models using 10-fold cross validation (left) and prediction performance on test data (tensile strength)

As can be seen from Figure 6, Random Forests had the best performance score. Therefore, the Random Forest model was selected and taken to the step of hyperparameter tuning (3.5).

For Thermal Conductivity

The following regression models from scikitlearn's machine learning library were used to predict the thermal conductivity of copper alloys using only their composition features:

- Random Forests
- Linear Regression
- Decision Trees
- K-Nearest Neighbours
- Multi-Layered Perceptron
- XGBoost

Figure 7 shows a comparison of their cross-validation performances on the training data.

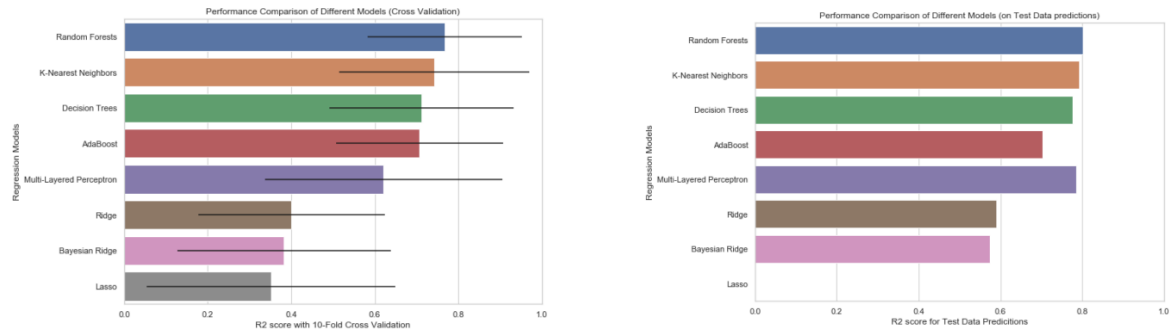


Figure 7. Performance of different machine learning models using 10-fold cross validation (left) and prediction performance on test data (for thermal conductivity)

As can be seen from Figure 7, Random Forests had the best performance score. Therefore, the Random Forest model was selected and taken to the step of hyperparameter tuning(3.5).

3.5 Hyperparameter Tuning

The hyperparameter tuning step is necessary one since it is shown to optimise model performance. The tuning process was carried out separately for tensile strength and for thermal conductivity.

The figures in this section are used to visualise and compare performance improvement (or the lack of it) after the hyperparameter tuning steps.

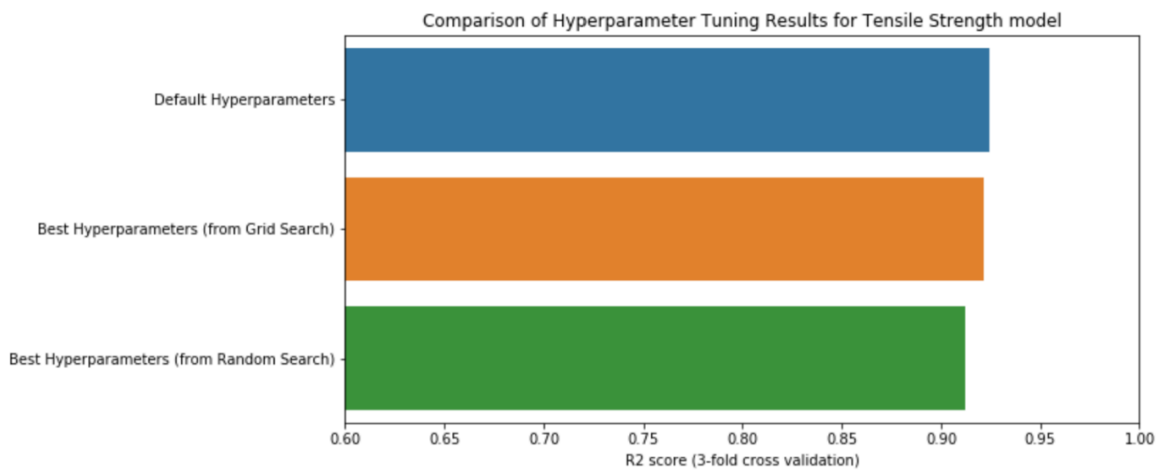


Figure 8. Comparison of the cross-validation performance of the model with default hyperparameters with the best models obtained after hyper parameter tuning using Grid Search and Random Search (for tensile strength)

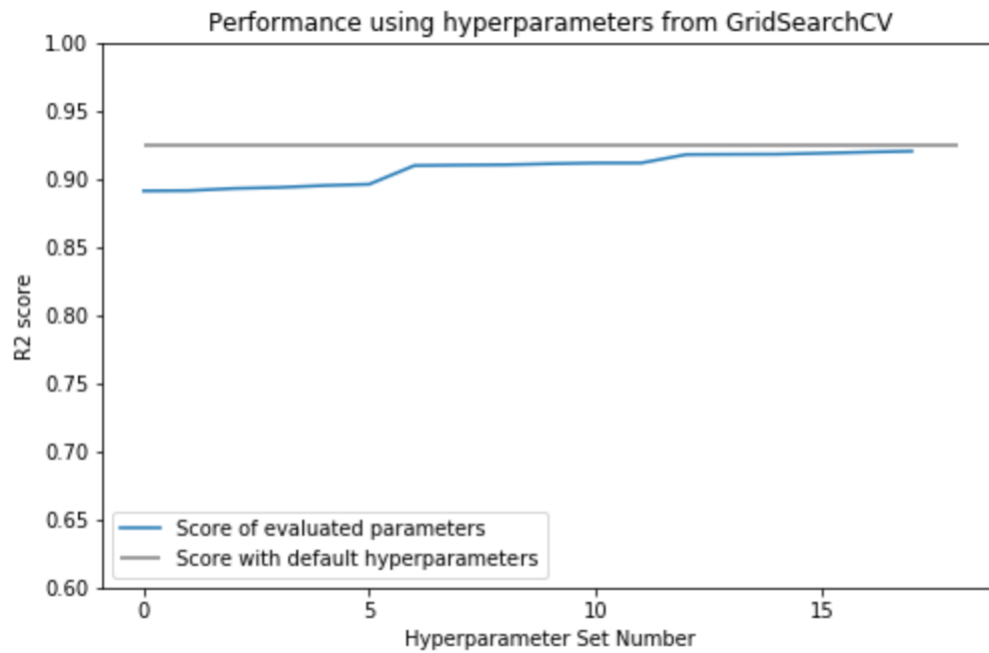


Figure 9. Demonstration of lack of performance improvement after hyperparameter tuning (for tensile strength)

A set of hyperparameters were defined to create separate models and were evaluated on the basis of their cross-validation performances using scikitlearn's inbuilt module GridSearchCV. (The workings of GridSearchCV are as discussed earlier).

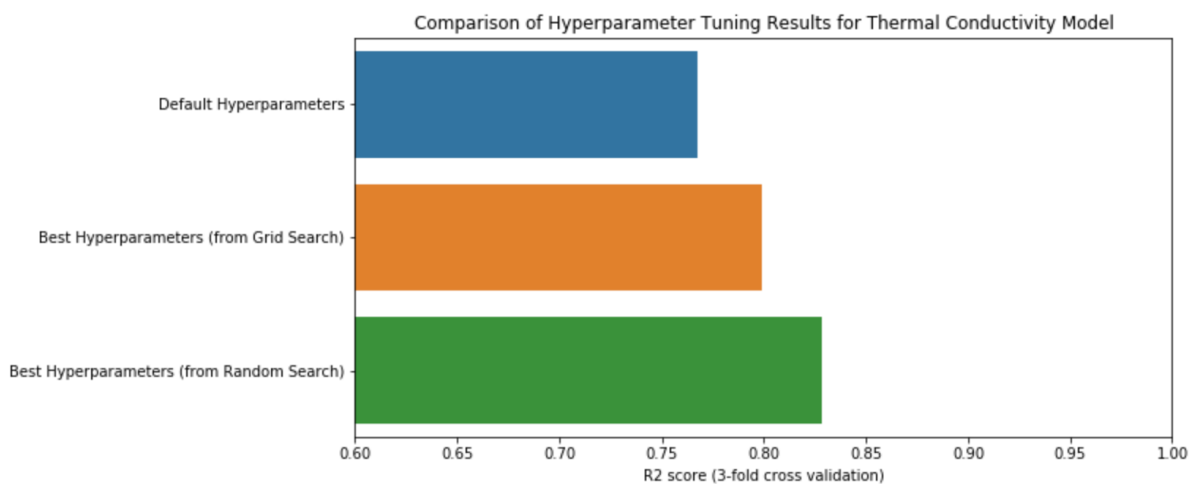


Figure 10. Comparison of the cross-validation performance of the model with default hyperparameters with the best models obtained after hyper parameter tuning using Grid Search and Random Search (for thermal conductivity)

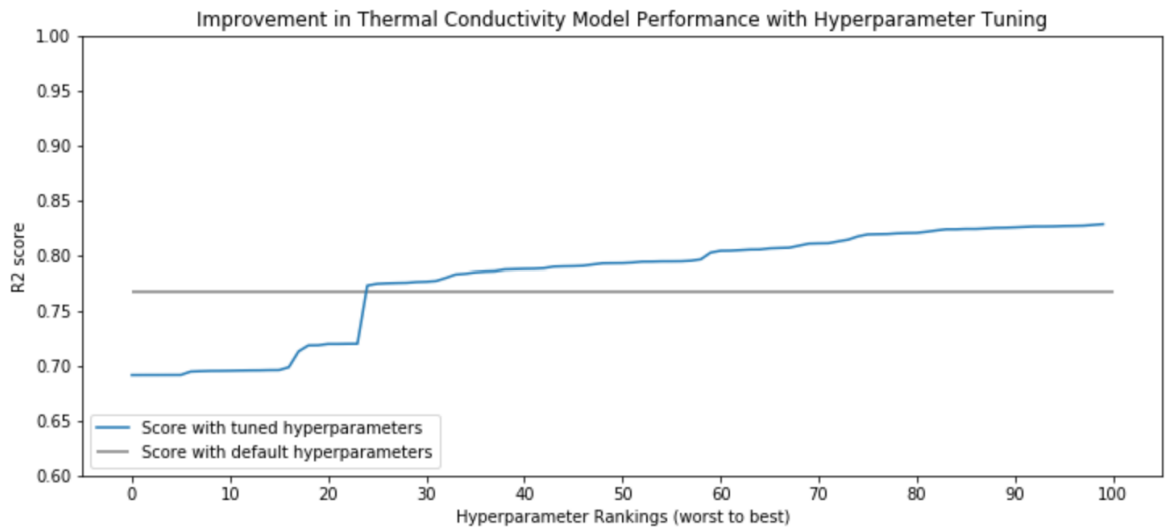


Figure 11. . Demonstration of lack of performance improvement after hyperparameter tuning (for thermal conductivity)

The parameters which created the best performing model was ultimately selected.

3.6 Model Evaluation

Following hyperparameter tuning, the models were used to make predictions on both training and test data for evaluation. The evaluation was done using the R-squared metric which has a scale of 0 to 1, with 1 being the best.

3.7 Inverse Modelling

The objective of this step is to create a system with the function which takes an input of the desired value a physical property and uses the model above to output alloy compositions which are likely to have that desired property value.

The following procedure was followed for the inverse modelling:

1. Take user inputs (desired value, property, number of predictions)
2. Choose 10 samples from the dataset with values closest to the desired value
3. Generate new data by shuffling values within these samples
4. Use models to make predictions on newly generated samples.
5. Return alloys from dataset (if any) and from the predictions, which have the closest values of the desired property (along with a confidence percentage)

The predicted compositions were then evaluated against the publicly available copper alloy data^[14].

Chapter 4: Results and Analysis

4.1 Prediction Performance

Both the models showed good predictive capabilities, performing equally well on both training and test data with no indication of overfitting, especially the model developed for tensile strength predictions.

The prediction quality on unseen data can be observed (Figure 12) where we observe that the predicted values are awfully close to the actual values with only slight error for majority of the samples. The model for thermal conductivity performs well given that it was trained using only 200 samples. However, there is a large error for some of the cases, suggesting scope of improvement.

Figure 12. visualises the closeness of the predicted values to the actual values of relevant properties for unseen data.

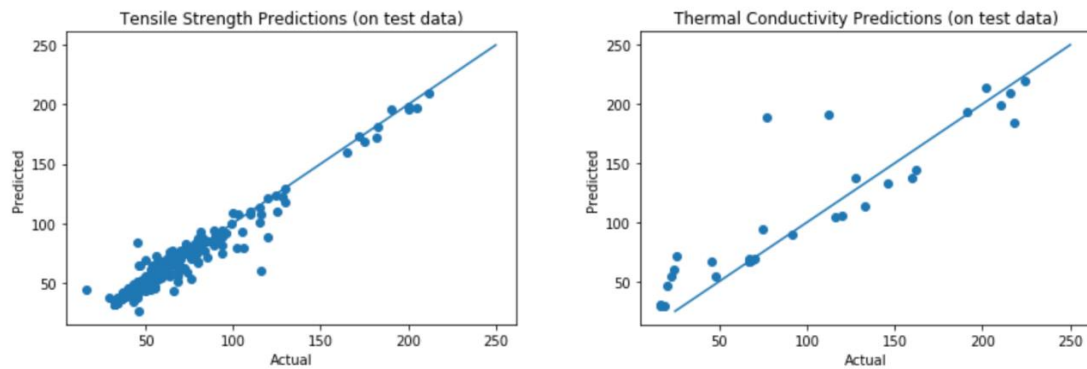


Figure 12. Predicted vs Actual Values of our models

The following table contains the summary of the model performances for both models.

Target Property	10-Fold Cross Validation Score summary			R-squared score (Test Data)
	Mean R-squared score	Median R-squared score	Standard Deviation	
Tensile Strength	0.925510	0.937542	0.035971	0.938538
Thermal Conductivity	0.822672	0.838199	0.168483	0.757116

4.2 Analysis of Features' Importance:

Random forests model allows to access a hierarchy of the features which prove to be most useful for predicting each the relevant properties. These importance metrics are represented in descending order of their contribution percentages for both the models (Figure 13). These values indicate that the features for beryllium and processing factors have a large contribution to determining the tensile strength of a copper alloy. The reliability of these results are affirmed by the excellent performance by the tensile strength model.

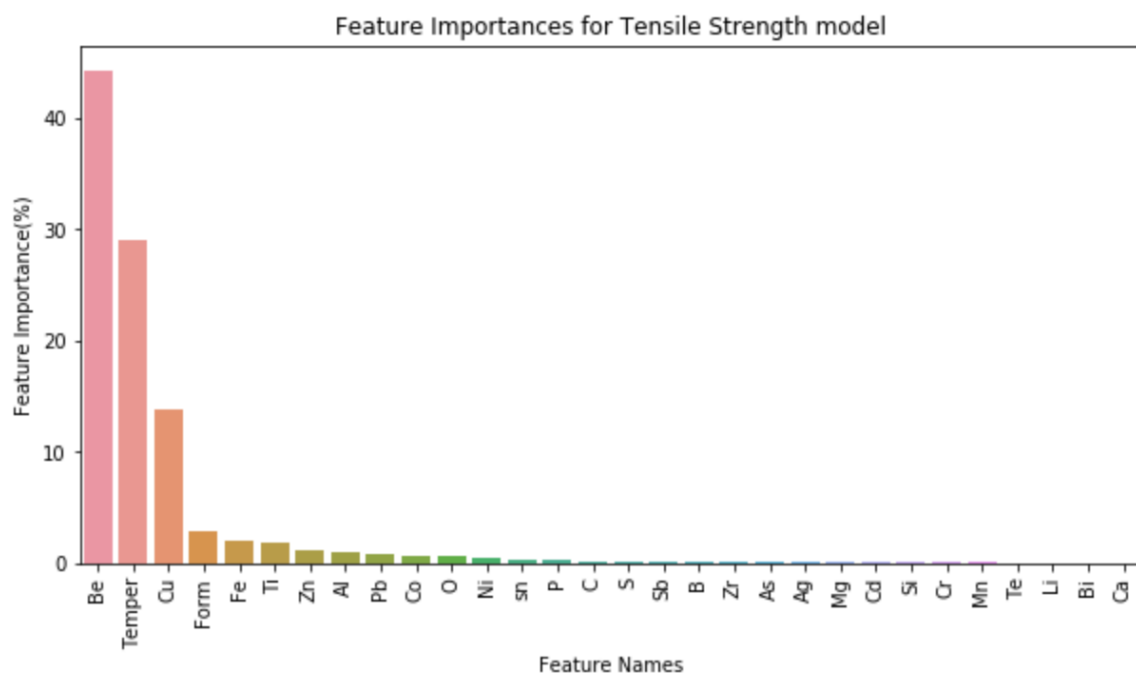


Figure 13. Visualising feature importance for tensile strength

The features' importance for thermal conductivity prediction are not given much regard in this analysis since the model still has scope for improvement. Nonetheless, the hierarchy of features can be seen in Figure 14.

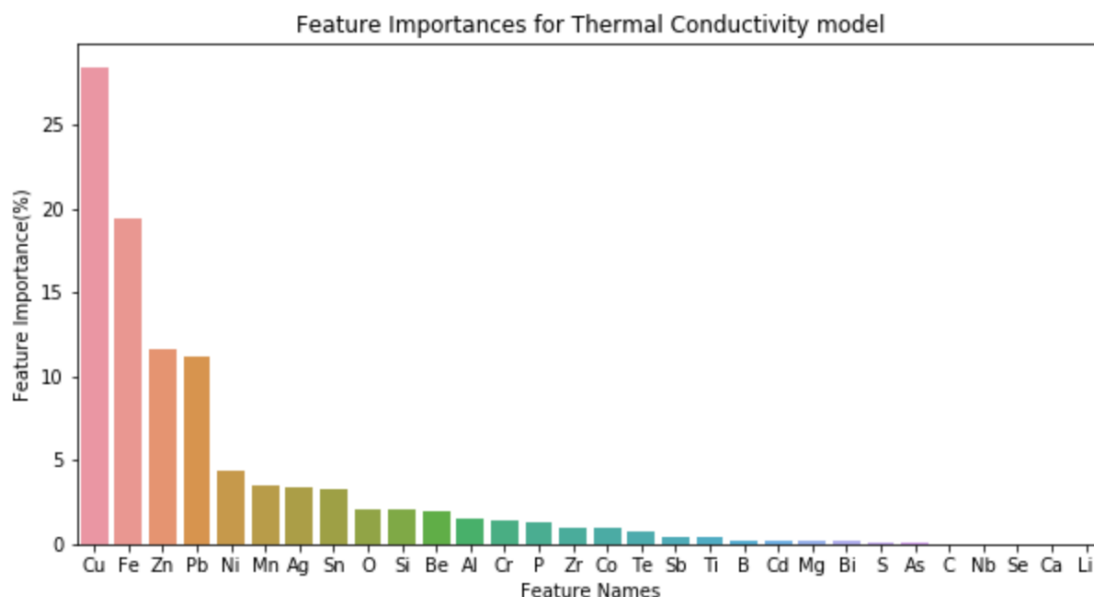


Figure 14. Visualising feature importance for thermal conductivity

4.3 Results of SAGS system :

The inversed model generates candidate copper alloys as their a percentage by mass composition of their constituent elements (also, form and temper features in the case of alloy generation using tensile strength) based on a specified target value of tensile strength or thermal conductivity.

The candidate solutions for a few values of both the properties were inspected. This step was challenging because of the lack of an accessible copper alloy data or similar resources. There were multiple instances where the system returned compositions with the selected value of the relevant properties. Moreover, there were also a few cases where the alloys described by the system were not in the training dataset and had the same value of the desired properties, which we were able to verify using the public database.

Chapter 5: Conclusions and Future Work

5.1 Conclusions

This paper delivers my study of application of machine learning to improve the process of alloy discovery. It focusses not only on building and evaluating machine learning models for predicting properties of copper alloys using their composition but also provides a solution to creating a system which can be used to efficiently generate new copper alloys. The structure that has been followed in this paper is designed with the intention that it can be applied in the discovery of alloys of metals other than copper, starting from data pre-processing followed by model selection and ultimately creating an inverse model that can generate candidate alloy compositions with desired physical properties.

This use of machine learning as a tool for building predictive models as opposed to deep learning models (as done by (Wang et al., 2020)), has proven to be an effective way of identifying which features of alloys have a large contribution to their tensile strength and thermal conductivity as well as ones which don't. This is a significant strength of this paper which can guide the feature engineering process in future studies on related topics. This unambiguous nature of machine learning models, as opposed to the black box like nature of deep learning models, bring significance the findings of this paper. Given the demonstration of powerful predictive abilities of neural network like models, we speculate the possibility of a combination of the two for this task. We strongly believe that the predictive abilities of deep learning architectures can be combined with the insights of the relevant features identified in this paper. We encourage that further research be devoted to this domain.

We believe that the fact that we were able to generate significantly accurate predictions for most of the alloys with no indications of overfitting demonstrates the existence of a relationship between alloy compositions and their physical properties. This knowledge can help the development of more sophisticated systems for alloy discovery using the power of artificial intelligence to drive the process.

5.2 Limitations

Although, there is evidence that our alloy generation system provides good potential candidates, there is still a lot of scope for improvement.

The performance of the current models, although impressive, are not sufficient to be deployed in a real-world system, especially for thermal conductivity.

The dataset used to train the models in this paper contains alloys containing only 26 different elements (excluding copper). And the proposed system only generates combinations of existing alloys. So, the generated alloys only come from this small subset of the potential search space. For example, Alloys containing carbon have shown to have many of the desirable properties of good alloys but are not included in our dataset, therefore, copper alloys containing carbon cannot be generated by our system.

5.3 Future Work

The insights of important features acquired from the Random Forest models can be used to engineer new and powerful indicators of the tensile strength and thermal conductivity and be applied to more sophisticated models for improving the performance of the current model.

We believe that the thermal conductivity model's performance could be significantly improved by the addition of more data and is essential before the system can truly provide utility in the space of alloy discovery.

Currently, only a few of the alloy compositions generated by the inverse model have been inspected. Further analysis of the quality of the generated alloy by the inverse modelling technique is required before the system can be deployed as a solution for copper alloy generations, either by investigating existing copper alloy databases or by a process of developing the candidate alloys in a controlled environment (such as a lab) and evaluating their tensile strength and thermal conductivity properties.

References

1. Wang, C., Fu, H., Jiang, L. et al. A property-oriented design strategy for high performance copper alloys via machine learning. *npj Comput Mater* 5, 87 (2019).
2. Ling, Julia, Antono, Erin, Bajaj, Saurabh, Paradiso, Sean, Hutchinson, Maxwell, Meredig, Bryce, and Gibbons, Brenna M. "Machine Learning for Alloy Composition and Process Optimization." *Proceedings of the ASME Turbo Expo 2018: Turbomachinery Technical Conference and Exposition. Volume 6: Ceramics; Controls, Diagnostics, and Instrumentation; Education; Manufacturing Materials and Metallurgy.* Oslo, Norway. June 11–15, 2018. V006T24A005. ASME.
3. Ren, F., Ward, L., Williams, T., Laws, K. J., Wolverton, C., Hattrick-Simpers, J., & Mehta, A. (2018). Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. *Science advances*, 4(4), eaaq1566.
4. European Copper Institute. 2020. Copper And Copper Alloys, Suited To A Wide Range Of Applications. [online] Available at: <<https://copperalliance.org.uk/about-copper/copper-alloys/>> [Accessed 15 June 2020].
5. Aalco.co.uk. 2020. Copper And Copper Alloys - General Information - Introduction To Copper And Its Alloys. [online] Available at: <http://www.aalco.co.uk/datasheets/Copper-and-Copper-Alloys-Introduction-to-Copper-and-its-Alloys_68.ashx> [Accessed 15 June 2020].
6. Hu, X., Wang, J., Wang, Y., Li, J., Wang, Z., Dang, Y., & Gu, Y. (2018). Two-way design of alloys for advanced ultra supercritical plants based on machine learning. *Computational Materials Science*, 155, 331-339.
7. Kim, K., Kang, S., Yoo, J., Kwon, Y., Nam, Y., Lee, D., ... & Son, W. J. (2018). Deep-learning-based inverse design model for intelligent discovery of organic molecules. *npj Computational Materials*, 4(1), 1-7.
8. Aalco.co.uk. 2020. *Copper And Copper Alloys - General Information - Introduction To Copper And Its Alloys.* [online] Available at: <http://www.aalco.co.uk/datasheets/Copper-and-Copper-Alloys-Introduction-to-Copper-and-its-Alloys_68.ashx> [Accessed 15 June 2020].
9. En.wikipedia.org. 2020. *List Of Copper Alloys.* [online] Available at: <https://en.wikipedia.org/wiki/List_of_copper_alloys#:~:text=Copper%20alloys%20are%20metal%20alloys,and%20brass%2C%20using%20zinc%20instead.&text=Today%20the%20term%20copper%20alloy,be%20substituted%2C%20especially%20by%20museums.> [Accessed 15 June 2020].
10. Rueylung.com.tw. 2020. *Tensile Strength.* [online] Available at: <http://www.rueylung.com.tw/home/site_index/Tensile_Strength#:~:text=Tensile%20strength%20is%20measured%20in,which%20it%20breaks%20in%20two.> [Accessed 15 June 2020].
11. Thermtest Inc. 2020. *What Is Thermal Conductivity? An Overview - Thermtest.* [online] Available at: <<https://thermtest.com/what-is-thermal-conductivity>> [Accessed 15 June 2020].
12. En.wikipedia.org. 2020. *Wiedemann–Franz Law.* [online] Available at: <https://en.wikipedia.org/wiki/Wiedemann%E2%80%93Franz_law> [Accessed 15 June 2020].
13. Wikipedia contributors. (2020, June 6). Electrical resistivity and conductivity. In *Wikipedia, The Free Encyclopedia.* June 15, 2020
14. Copper.org. 2020. *Official Site Of Copper Development Association, Inc. (USA).* [online] Available at: <<https://www.copper.org/>> [Accessed 15 June 2020].
15. Medium. 2020. *Hyperparameter Tuning.* [online] Available at: <<https://towardsdatascience.com/hyperparameter-tuning-c5619e7e6624>> [Accessed 15 June 2020].
16. Wikipedia contributors. (2020, June 12). Supervised learning. In *Wikipedia, The Free Encyclopedia.*, June 15, 2020
17. Wikipedia contributors. (2020, June 14). Decision tree learning. In *Wikipedia, The Free Encyclopedia.* June 15, 2020

18. Wikipedia contributors. (2020, May 2). Random forest. In *Wikipedia, The Free Encyclopedia*. June 15, 2020
19. Medium. 2020. *Introduction To Data Preprocessing In Machine Learning*. [online]
20. Beaulieu, K. and Dalisay, D., 2020. *Machine Learning Mastery*. [online] Machine Learning Mastery. Available at: <<https://machinelearningmastery.com/>> [Accessed 15 June 2020].
21. Medium. 2020. *Hyperparameter Tuning The Random Forest In Python*. [online] Available at: <<https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74>> [Accessed 15 June 2020].