A regression-based model evaluation of the Curie temperature of transition-metal rare-earth compounds

Duong-Nguyen Nguyen1, Tien-Lam Pham1,2,3, Viet-Cuong Nguyen4, Anh-Tuan Nguyen5, Hiori Kino2,3, Takashi Miyake2,3,6, and Hieu-Chi Dam1,2,7

1Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
2CMI2, NIMS, Tsukuba 305-0047, Japan
3ESICMM, NIMS, Tsukuba 305-0047, Japan
4HPC Systems Inc., Japan
5Hanoi Metropolitan University, 98 Duong Quang Ham, Cau Giay, Hanoi, Vietnam
6CD-FMat, AIST, Tsukuba 305-8568, Japan
7JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan

E-mail: dam@jaist.ac.jp

Abstract. The Curie temperature \( T_C \) of \( RT \) binary compounds consisting of 3\( d \) transition-metal (\( T \)) and 4\( f \) rare-earth elements (\( R \)) is analyzed systematically by a developed machine learning technique called kernel regression-based model evaluation. Twenty-one descriptive variables were designed assuming completely obtained information of the \( T_C \). Multiple kernel regression analyses with different kernel types: cosine, linear, Gaussian, polynomial, and Laplacian kernels were implemented and examined. All possible descriptive variable combinations were generated to construct the corresponding prediction models. As a result, by appropriate combinations between descriptive variable sets and kernel formulations, we demonstrate that a number of kernel regression models can accurately reproduce the \( T_C \) of the \( RT \) compounds. The relevance of descriptive variables for predicting \( T_C \) are systematically investigated. The results indicate that the rare-earth concentration is the most relevant variable in the \( T_C \) phenomenon. We demonstrate that the regression-based model selection technique can be applied to learn the relationship between the descriptive variables and the actuation mechanism of the corresponding physical phenomenon, i.e., \( T_C \) in the present case.

1. Introduction

The development of strong permanent magnets is an urgent technological issue as well as a fundamental challenge in materials science. Most strong permanent magnets are rare-earth magnets that mainly consist of transition-metal and rare-earth elements. To date, various rare-earth transition-metal compounds have been synthesized. The strongest magnetic compound, \( \text{Nd}_2\text{Fe}_{14}\text{B} \), which is the main phase of neodymium magnets, was developed by Sagawa et al. [1]. Because of its relatively low Curie temperature \( (T_C) \) of 586 K, its coercivity at high temperatures is not sufficient for technological usage; therefore, dysprosium (Dy) is added to ensure sufficient coercivity. In fact, Sagawa et al. intended to increase the \( T_C \) of \( \text{R}_2\text{Fe}_{17} \) (where \( R \) is a rare-earth...
element). However, the $T_C$ was much lower than that of Sm$_2$Co$_{17}$, which was the strongest magnet before the invention of neodymium magnets. By adding boron to Nd$_2$Fe$_{17}$ ($T_C = 335$ K), they obtained a new compound with a higher $T_C$, Nd$_2$Fe$_{14}$B. However, the $T_C$ values are still much lower than those of Sm$_2$Co$_{17}$ (1,193 K) and $\alpha$-Fe (1,043 K). Owing to this complexity, it is highly desirable to establish a technique to predict accurate $T_C$ values and to clarify the controlling parameters (variables) that influence this property for the development of a new strong magnet.

From a theoretical point of view, an accurate description of $T_C$ is a demanding task. The spin magnetic moment at each atomic site and the magnetic exchange coupling between them can be evaluated within the framework of density functional theory [2]. $T_C$ can be obtained by solving a derived classical spin model [3, 4]. This approach works reasonably well for $\alpha$-Fe, whereas it gives a value that deviates significantly from the experimental observation for Ni, because the longitudinal spin fluctuation is not captured in the classical spin model [5]. In the case of rare-earth compounds, the $4f$ electrons add further complications. Both the electron correlation and spin–orbit interaction are strong in these compounds; hence, an advanced theoretical treatment is needed for a reliable description of the magnetic properties.

The recent rapid development of data-driven approaches in materials research offers another possibility [6, 7, 8, 9, 10]. At present, $T_C$ data can be obtained for many compounds from databases or literature. Machine learning may be utilized to predict the $T_C$ of a new compound from existing data. In the present work, we focus on the $T_C$ of transition-metal rare-earth bimetals. We explain the Kernel ridge regression (KRR) technique with various forms of the kernel functions and how to select variables to represent the collected data. We demonstrate how the $T_C$ value of a compound can be reproduced from the existing data by the KRR. We also analyze the role of variables for obtaining a high prediction accuracy. Next, a comparison between the $T_C$ value estimated by KRR and the experimentally observed $T_C$ value, for the new compounds is studied. Through a systematic analysis, we show that insights into the actuating mechanisms of physical phenomena can be obtained through the determination of a set of physically meaningful parameters by employing machine learning algorithms.

2. Methodology

2.1. Data collection and representation

We collected experimental data of 101 binary compounds consisting of transition metals and rare-earth metals from the Atomwork database of NIMS [11, 12], including the crystal structure of the compounds and their observed $T_C$ values. Our task was to develop a model for estimating the $T_C$ of a new compound from the training data of known compounds. For this purpose, one of the most important steps is the choice of an appropriate data representation that reflects the knowledge of the application domain, i.e., a model of the underlying physics. On the other hand, a data representation that derives a good estimation model may imply the discovery of the underlying physics. To represent the structural and physical properties of each binary compound, we use a combination of 21 variables. We divide all the 21 variables into three groups. The first and second categories contain descriptive variables that describe the atomic properties of the transition-metal ($T$) and rare-earth ($R$) constituents. The properties were as follows: (1) atomic number ($Z_T, Z_R$); (2) covalent radius ($r_{covT}, r_{covR}$); (3) first ionization ($IP_T, IP_R$); and (4) electronegativity ($\chi_T, \chi_R$). In addition, descriptive variables related to the magnetic properties were included: the (5) total spin quantum number ($S_{3d}, S_{4f}$); (6) total orbital angular momentum quantum number ($L_{3d}, L_{4f}$); and (7) total angular momentum ($J_{3d}, J_{4f}$).

The selection of these variables originates from the physical consideration that the intrinsic electronic and magnetic properties will determine the $3d$ orbital splitting at transition-metal sites. To capture the effect of the $4f$ electrons better, especially the strong spin-orbit coupling
effect, we add three additional variables to describe the properties of the constituent rare-earth metal ions, (15) the projection of the total magnetic moment onto the total angular moment \((J_{4f}g_j)\), and (16) the projection of the spin magnetic moment onto the total angular moment \((J_{4f}(1-g_j))\) of the 4\(f\) electrons. The selection of these features is based on the physical consideration that the magnitude of the magnetic moment will determine the \(T_C\). It has been well established that information related to the crystal structure is very valuable for understanding the physics of binary compounds with transition metals and rare-earth metals. Therefore, we design a third group with structural variables that roughly represent the structural information at the transition metal and rare-earth metal sites, which are (17) the concentration of the transition metal \((C_T)\), and (18) the concentration of the rare-earth metal \((C_R)\). Note that if we use the atomic percentage for the concentration, the two quantities are not independent. Therefore, in this work, we measured the concentrations in units of atoms/\(\text{Å}^3\). This unit is more informative than the atomic percentage, as it contains information on the constituent atomic size. As a consequence, \((C_T)\) and \((C_R)\) were not completely dependent on each other. Other structure variables were also added: the mean radius of the unit cell (19) between two rare-earth elements \(r_{RR}\), (20) between two transition metal elements \(r_{TT}\), and (21) between transition and rare-earth elements \(r_{TR}\). We set the experimentally observed \(T_C\) as the target variable.

2.2. Kernel regression analysis

To learn a function for predicting the values of \(T_C\) of compounds from data represented by using vectors of the variables, we apply the KRR technique [13], which has recently been applied successfully in many materials science studies [14, 15, 16]. In our study, we apply the KRR with five kernel functions, the forms of which are defined as follows:

Cosine kernel:
\[
k(x_i, x_j) = \frac{x_i \cdot x_j}{\|x_i\| \cdot \|x_j\|}
\]

Linear kernel:
\[
k(x_i, x_j) = x_i \cdot x_j + c
\]

Polynomial kernel:
\[
k(x_i, x_j) = (\alpha x_i \cdot x_j + c)^d
\]

Gaussian kernel:
\[
k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)
\]

Laplacian kernel:
\[
k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{\sigma}\right)
\]

where \(\|x_i - x_j\| = \sum_{a=1}^{m} |x_{ia} - x_{ja}|\); \(x_i \cdot x_j = \sum_{a=1}^{m} x_{ia}x_{ja}\); and \(m\) is the number of dimensions; \(\sigma, \alpha\) are the variance of the Gaussian/Laplacian and the hyper parameter of polynomial kernel functions, respectively; \(d\) is the polynomial order of the polynomial kernel—we set \((d = 3)\) in all our experiments; \(c\) is a constant.

For a given new data instance \(x_\ast\), the predicted property \(\hat{f}(x_\ast)\) is expressed as the weighted sum of the kernel functions:
\[
\hat{f}(x_\ast) = \sum_{i=1}^{N} c_i k(x_\ast, x_i)
\]

where \(N\) is the number of training data points, and the weighting coefficients \(c_i\) for the corresponding data point \(x_i\) are determined by minimizing
\[
\sum_{i=1}^{N} [\hat{f}(x_i) - y_i]^2 + \lambda \sum_{i=1}^{N} ||c_i||^2_2.
\]
The regularization parameters $\lambda$ and hyper parameters $(\sigma, \alpha)$ are chosen with the help of cross-validation[17, 18], i.e., by excluding some of the data instances during the training process and maximizing the coefficient of determination $R^2$:

$$R^2(y^{\text{predict}}_{\text{vld}}, y^{\text{observed}}_{\text{vld}}) = 1 - \frac{\sum_{j=1}^{p_{\text{vld}}} \left[ \hat{f}(x_j) - y_j \right]^2}{\sum_{j=1}^{p_{\text{vld}}} \left[ \bar{y} - y_j \right]^2}$$  (8)

where $p_{\text{vld}}$ is the number of validation data points used to evaluate how the predicted values agree with the actual observed values, and $\bar{y}$ is the average of the validation set used to compare the values predicted for the excluded compounds with the known observed values. By employing the cross-validation framework, the evaluation of the prediction accuracy is carried out by using data that are not involved in the training process to avoid overfitting the training data. In this study, we use the maximized $R^2$ as a measure of prediction accuracy. To obtain a good estimation of the prediction accuracy, we carried out KRR with leave-one-out cross-validation using the collected data, of which $T_C$ is considered as a response variable, and the other variables are considered as descriptive variables. It should be noted that in KRR, the kernel metric represents a formulation to measure the similarity between data instances in an abstract description space [19]. Therefore, KRR analyses with different kernel metrics will provide a better understanding of the data set.

2.3. Kernel regression-based variable evaluation

To find the most appropriate set of variables for the prediction of $T_C$ and the hidden knowledge regarding the determination of the values of $T_C$ of the compounds, we utilize an exhaustive search to evaluate all possible variable combinations [20, 21, 22] to identify and remove irrelevant and redundant variables [23, 24, 25]. KRR models with various kernel types described previously are applied as the predictors. For a given variable combination, we determine the regularization parameters $\lambda$ and other hyperparameters corresponding to the specific kernel setting in each experiment (detail in Section 2.2) to maximize the prediction accuracy for $T_C$ of the KRR model. By training the KRR models for all combinations of the variables, we can obtain new data for all possible sets of variables. By using the variables in that set, the KRR model can achieve the best prediction accuracy. To analyze this kind of data, we focus on two research objectives: (1) to find the set of variables that yield the best prediction accuracy for the prediction of $T_C$, (2) to find the variables that are relevant for the prediction of $T_C$.

One can easily presume that incorporating irrelevant variables into the KRR model may impair its prediction accuracy. Therefore, we define the prediction ability $PA(S)$ of a set $S$ by the maximum prediction accuracy that the KRR model can achieve by using the variables in a subset $s$ of $S$ as follows:

$$PA(S) = \max_{\forall s \subset S} R^2_s; \quad s_{PA} = \arg \max_{\forall s \subset S} R^2_s,$$  (9)

where $R^2$ is the value of the coefficient of determination $R^2$ [26] achieved by the KRR using a variable set $s$ as the independent variables. $s_{PA}$ is the variable subset of $S$ that yields the prediction model having the maximum prediction accuracy.

2.4. Variable relevance analysis

On the basis of Eq.(9), we can evaluate the relevance [27, 28] of a variable for the prediction of $T_C$ using the expected reduction in the prediction ability resulting from the removal of this variable from the full set of variables. Let $D$ be a full set of variables, $d_i$ a variable, and $D_i = D - \{d_i\}$ the full set of variables after removing the variable $d_i$. The degree of the relevance of variables can be formalized as follows:
1. Strong relevance: a variable is strongly relevant if and only if

\[ PA(D) - PA(D_i) = \max_{s \in D} R_s^2 - \max_{s \in D_i} R_s^2 > 0. \]  

Among the strongly relevant variables, a variable that causes a larger reduction in the prediction ability when it is removed can be considered to be a strong variable. The degree of relevance of a strongly relevant variable can be computationally estimated by using the leave-one-out approach, i.e., by leaving out a variable in the currently considered variable set for the KRR analysis and evaluating the extent to which the prediction accuracy is impaired.

2. Weak relevance: a variable is weakly relevant if and only if

\[ PA(D) - PA(D_i) = \max_{s \in D} R_s^2 - \max_{s \in D_i} R_s^2 = 0 \]

and

\[ \exists D'_i \subset D_i \text{ s.t } PA([d_i, D'_i]) - PA(D'_i) > 0. \]  

It is clearly seen from Eq. (11) that estimation of the degree of relevance for the weakly relevant variables cannot be carried out in a straightforward manner, as with the strongly relevant variables. Weakly relevant variables are variables that are relevant for prediction, but they can be replaced by other variables.

2.5. Ensemble learning for prediction

In this step, we select a set of variable combinations, that yields top highest \( R^2 \) score KRR models, and consider as valuable variable combinations for capturing the \( T_C \) phenomenon. We
Table 1: Prediction accuracy of ensemble learning model with different kernel matrices

| Kernel | Cosine kernel | Linear kernel | Polynomial kernel | Gaussian kernel | Laplacian kernel |
|--------|---------------|---------------|------------------|----------------|-----------------|
| PA     | 0.572         | 0.569         | 0.982            | 0.982          | 0.973           |
| sPA    | ZR, Jf, ZT,  | rRTR, S3d,   | ZR S4f, L3d,     | ZR, χT, J3d,   | χR, Jf (1−gj), |
|        | rconv, Tr,   | rRR, rTT,    | C3d, rRR         | J3d, rRR       | ZR, C3d, rRTR,  |
|        | CR            | C3d           | C3d              | CT, CR         | C3d             |

then apply a technique from ensemble learning [29, 30, 31] to integrate all these KRR models to build a prediction model for predicting the TC of new compounds.

In statistical theory, the prediction error of a model in the prediction of the target variable y for a new data instance x can be decomposed into three factors: bias, variance, and irreducible factors [29], as follows:

\[
E[(y - \hat{f}(x))^2] = (Bias[\hat{f}(x)])^2 + Var[\hat{f}(x)] + \sigma_{irr}^2
\]  

(12)

with Bias and Var are :

\[
Bias[\hat{f}(x)] = E[f(x) - \hat{f}(x)]
\]  

(13)

\[
Var[\hat{f}(x)] = E[\hat{f}(x)^2] - (E[\hat{f}(x)])^2
\]  

(14)

where \( f(x) \) is an actual function associated with the new data points \( y = f(x) + \epsilon \), the noise \( \epsilon \) has zero mean and variance \( \sigma_{irr} \); \( \hat{f}(x) \) is the estimated function (or predicted values of discrete points x).

The irreducible factor \( \sigma_{irr} \) depends on the information of the target variable stored in the prediction variables, which were used for describing the training data instances. It should be noted that increasing the number of training data points does not guarantee that the prediction ability of the learning model will improve. This error can only be reduced by improving the similarity with the training data or by improving the designed descriptive variable. On the other hand, the reducible factors, bias and variance, depend on the nature of the modeling method, i.e., the KRR in these experiments. In particular, the variance factor only measured the movement of the KRR around its mean. In this research, we empirically investigate the probability distribution of the predicted values of the target variable for new data instances by utilizing an ensemble learning and bootstrap aggregating–bagging [32] algorithm. In more detail, we vary both the combination of descriptive variables and the training set by sampling with replacement from the original data uniformly. By doing so, we can indirectly observe the behavior of both the bias and the irreducible factors. An prediction model for predicting the TC of new compounds can be constructed from the obtained distribution of prediction values.

3. Results and discussion

3.1. Kernel regression-based variable evaluation

To investigate the scientific connection [20] between the variables and the actuation mechanisms of the physical phenomenon of TC in these bimetal systems, we examine whether TC can be predicted by using the designed variables of the compounds. A screening was conducted for all possible variable combinations, \( 2^{21} - 1 = 2,097,151 \) that respectively derive the same number of prediction models. As described in section 2.2, a given kernel metric formula associates with a method to measure the similarity between compounds. Therefore, five kernel metrics- cosine,
linear, Gaussian, polynomial, and Laplacian are all implemented and analyzed in this section. Leave-one-out cross-validations [17, 18] is performed to evaluate the prediction accuracy of these models.

The prediction abilities $PA$ (equation 9) of the designed descriptive variables set for different kernel-type predictors are summarized in the table 1. The highest $PA - R^2$ score 0.982 is achieved by two models deriving from variable combinations $s_{PA} = \{Z_R, \chi_T, J_{3d}, \gamma_{TR}, C_T, C_R\}$ with the Gaussian kernel and $s_{PA} = \{Z_R, S_{4f}, L_{3d}, J_{3d}, \gamma_{RR}, C_T, C_R\}$ with the polynomial kernel. The $PA$ value with the Laplacian kernel experiment achieves an $R^2$ score of 0.973, with its $s_{PA} = \{\chi_R, \chi_T, J_{4f} (1 - g_j), Z_T, r_{covT}, IP_T, S_{3d}, L_{3d}, J_{3d}, C_R\}$. In this screening result, there are 892,612 variable combinations associated with the Gaussian kernel; 284,649 variable combinations associated with the polynomial kernel and 1,317,193 variable combinations associated with the Laplacian kernel that yield regression models with $R^2$ scores exceeding 0.90. It is noted that, even with the same kernel metric, several regression models achieved a similar excellent prediction accuracy because the designed variables are not independent variables.

On the other hand, the $PA$ values with linear and cosine kernels were lower than 0.8, i.e., the $T_C$ variable cannot be predicted by our designed descriptive variable set with these kernel regression models.

Finally, the $PA$ analysis indicates that with all the designed variables, it is possible to accurately predict the values of $T_C$ of the rare-earth transition bimetal alloy compounds by using Gaussian, polynomial, and Laplacian KRR models with the designed variables. In the subsequent sections, we will discuss the method to improve the maximum prediction accuracy of the models as well as the physical meaning of strongly relevant variables.

### 3.2. Strongly relevant and weakly relevant features

Figure 2 shows the dependence of the highest prediction accuracy $PA$—red lines on the number of variables recruited in the Gaussian-polynomial-Laplacian kernel regression models. In general, the prediction accuracy in all the experiments reaches the highest value with the number of descriptive variables from 6 to 8, and then gradually decreases when the number of recruited variables increases. The small subset of the designed descriptive variables and the large number of high-accuracy models described above originate from the fact that the overuse of many weakly relevant variables [23, 24, 25] weakens the correlation between the similarity of the compounds, which is measured using the kernel of the variables, and the differences in their $T_C$ values.

Next, we evaluate the relevance of each variable for the prediction of $T_C$. We compare $PA(D)$ of the full set of variables $D$ and $PA(D - \{d_i\})$ for all the variables $d_i$. We found that most of the variables are weakly relevant, and the prediction accuracy does not significantly change, except in one case—when removing the variables of the concentration of the rare-earth metal $C_R$. It is clearly seen in Figure 2 that the absence of $C_R$ in the Gaussian and polynomial kernel model results in a dramatic decrease in the accuracy: $PA(D) < PA(D - \{C_R\})$; therefore, $C_R$ is surely assigned as a strongly relevant variable in terms of the prediction of $T_C$ (see Section 2.4).

This result is consistent with the understanding so far that the values of $T_C$ of binary alloys consisting of 3$d$ transition-metal and 4$f$ rare-earth metals are mainly determined by the magnetic interaction in the transition-metal sublattice. In terms of molecular field theory [33], we have

$$T_C \sim n_{TT} M_T^2 / 3k_B,$$

where $M_T$ and $n_{TT}$ are the magnetization and molecular field coefficients of the transition-metal sublattice, respectively. Both $M_T$ and $n_{TT}$ strongly depend on $C_R$. The dependencies of $M_T$ and $n_{TT}$ on $C_R$ are different in compounds with different combinations of rare-earth metal (R) and transition metal (T). In Co-based and Ni-based compounds, $M_T$ and $n_{TT}$ tend to decrease when $C_R$ increases. This leads to a rapid decrease in $T_C$ with an increase in $C_R$. For example, in
Figure 2: Dependence of the best prediction accuracy on the number of variables in the Gaussian(a) - polynomial(b) - Laplacian(c) kernel regression model. The red line represents the maximum prediction ability \( PA(D) \) of the full descriptive variable set \( D \) with respect to the different numbers of variables. The other lines represent the trend of \( PA(D - \{d_i\}) \) by removing variable \( \{d_i\} \) from \( D \) in the same manner. The significant decrease of \( PA(D - \{C_R\}) \) shows that the concentration of the rare-earth element \( C_R \) is strongly relevant to \( T_C \).

Gd–Co compounds, \( T_C \) decreases from 1,404 K to 143 K with the increase in the concentration of Gd from 0% to 75%. In contrast, in R–Fe compounds, \( M_T \) and \( n_{TT} \) tend to increase with increasing \( C_R \). This leads to an increase in \( T_C \) with increasing \( C_R \). Indeed, in Gd–Fe compounds, \( T_C \) rapidly increases from 0 K to 827 K with the increase in the concentration of Gd from 0% to 33%. Detailed correlation between \( C_R \) and \( T_C \) in different transition metal-based compound is shown in Figure 1c.

3.3. Prediction of \( T_C \) for new compounds
In this section, we discuss the ability of ensemble learning using the obtained KRR models in the prediction of \( T_C \) for new compounds. The test set of new compounds includes five Fe–based compounds. The \( T_C \) of two of these compounds has been determined experimentally: SmFe\(_{12}\) – 555(K)[34], YFe\(_{12}\) – 483(K)[35]. No \( T_C \) information is available for NdFe\(_{12}\), GdFe\(_{12}\), and DyFe\(_{12}\).

From the discussion in Section 3.1, we know that \( C_R \) is the most strongly relevant to \( T_C \). The \( C_R \) values of the five compounds in the test set are approximately 0.006 atoms/A\(^3\), which is much lower than the smallest \( C_R \) value for all the other Fe–based compounds (which is 0.0075 atoms/A\(^3\) of Fe\(_{17}\) compounds–based) (see Figure 1c). The only compound having a similar \( C_R \) value is LaCo\(_{13}\) with \( C_R \) of 0.0055 atoms/A\(^3\). Further, the compounds in the test set are recently synthesized and have a crystal structure significantly different from that of all the other compounds in the data set. Therefore, the prediction for the \( T_C \) of these compounds could be seen closer to an extrapolative than an interpolative prediction problem.

Figure 3 shows the predicted value distributions for all test compounds (black line histogram) and the observed values (red dash line). It is easy to recognize that the distributions of the predicted \( T_C \) for these test compounds can be approximated by a mixture of three separate Gaussian distributions. We represent these distributions using red, blue and green, in increasing order of their mean values. From this Gaussian distribution decomposition, we can suggest that at least three distinguish functions \( \hat{f}(\mathbf{x}) \) can be regressed from the observed data to model the \( T_C \) phenomenon. Further, since the functions are learned from sub-groups of all the data set
Figure 3: Predicted value distribution of test compounds by using a bagging model constructed from top–5 highest prediction accuracy kernel regression model with Gaussian and Laplacian kernels. The distributions appear as a mixture of several Gaussian distribution components. This serves as evidence to show that there are a number of functions that generate such components. The black dashed lines show the predicted value obtained by taking the average of all possible values. The red dashed lines show the observed values.

by the bagging method, we can suggest that there are at least three sub-groups of compounds corresponding to these three models.

Next, we analyze the relation of the experimentally observed $T_C$ of the two test compounds SmFe$_{12}$ and YFe$_{12}$ to their predicted $T_C$ distribution. It is obvious that both the experimentally observed values correspond to the green Gaussian distribution, i.e., the largest $T_C$ group (see Figure 3a, 3b). The mean value of these green Gaussian distributions is close to the observed value of 535.1 K, as compared with 555.0 K of SmFe$_{12}$, and 443.7 K as compared with 483.0 K of YFe$_{12}$. From the fact that all the five test compounds are Fe$_{12}$–based compounds with the same crystal structure, we can infer that the experimental values of the remaining three compounds also correspond to the green distribution. Therefore, we predict that the $T_C$ of NdFe$_{12}$, GdFe$_{12}$, and DyFe$_{12}$ are 488.5 K, 568.5 K, and 444.7 K, respectively. There is also the potential for further experimental study on the $T_C$ of these compounds.

Lastly, all the above investigations show that further research is required to extract hidden functions in data as well as an alternative method of integrating predicted results from ensemble
learning algorithms.

4. Conclusion
In this study, we analyzed the $T_C$ target variable of binary alloys consisting of 3d transition-metal and 4f rare-earth elements by a machine learning technique called kernel regression-based variable evaluation. Twenty-one descriptive variables were designed assuming completely obtained information of the target variable. Multiple kernel regression with different kernel types: cosine, linear, Gaussian, polynomial, and Laplacian kernels were implemented as predictor models. All possible descriptive variable combinations were generated to construct the corresponding prediction models. The prediction accuracy of $T_C$ was maximized with different combinations of the designed descriptive variables with Gaussian, polynomial, and Laplacian kernel types. Further investigation the result of the screening prediction ability process shows that the rare-earth concentration is the most strongly relevant descriptive variable. Finally, we discuss the utilization of an ensemble model constructed by top highest prediction accuracy and a bagging algorithm to predict a test set of compounds. The final predicted $T_C$ values for NdFe$_{12}$, GdFe$_{12}$, and DyFe$_{12}$ are 488.5 K, 568.5 K, and 444.7 K, respectively. This research also shows that there are at least three mechanisms that drive the $T_C$ phenomenon. It also highlights the need to extract these mechanisms in future work. To conclude, in this research, we demonstrated how the kernel regression-based model evaluation can be utilized for ascertaining the scientific connection between the variable and the actuation mechanisms of a physical phenomenon by using both descriptive and predictive analyses.

Acknowledgments
This work was partly supported by PRESTO and by the “Materials Research by Information Integration” Initiative (MI$^2$I) project of the Support Program for Starting Up Innovation Hub, both from the Japan Science and Technology Agency (JST), Japan; by the Elements Strategy Initiative Project under the auspices of MEXT; and also by MEXT as a social and scientific priority issue (Creation of New Functional Devices and High-Performance Materials to Support Next-Generation Industries; CDMSI) to be tackled by using a post-K computer.

References
[1] Sagawa M, Fujimura S, Togawa N, Yamamoto H and Matsura Y 1984 J. Appl. Phys. 55 2083–2087
[2] Xu B Liu Z A 2011 Computational Materials Science 50 841 – 846
[3] Wipf A 2012 Statistical Approach to Quantum Field Theory (Springer)
[4] PMyles H 1997 Introductory Solid State Physics (Second Edition) (CRC Press)
[5] Ruban A V, Khmelevskyi S, Mohn P and Johansson B 2007 Physical Review B 75
[6] Yousef S, Da G, Thanh N, Scotty B, R C J and Wanda A 2012 Phys. Rev. B 85 104104
[7] Yang S, Lach-hab M, Vaisman I I and Blaisten-Barojas E 2009 Phys. Chem. C 113 21721
[8] Hautier G, Fischer C C, Jain A, Mueller T and Ceder G 2010 Chem. Mater 22 3762
[9] C S J, Matthias R, Katja H, Miller K R and Kieron B 2012 Phys. Rev. Lett 108 253002
[10] Isayev O, Fourches D, Muratov E N, Osnes C, Rasch K, Tropscha A and Curtarolo S 2015 Chem. Mater 27 735
[11] Villars P, Berndt M, Brandenburg K, Cenzual K, Daams J, Hulliger F, Massalski T, Okamoto H, Osaki K, Prince A, Putz H and Iwata S 2004 Journal of Alloys and Compounds 367
[12] Xu Y, Yamazaki M and Villars P 2011 Ipn. J. Appl. Phys. 50
[13] Murphy K P (ed) 2012 Machine learning: a probabilistic perspective (MIT Press)
[14] Matthias R 2015 International Journal of Quantum Chemistry 115 1058–1073
[15] Botu V and Ramprasad R 2014 Int. J. Quant. Chem. 115 1074–1083
[16] Pilania G, Wang C, Jiang X, Rajasekaran S and Ramprasad R 2013 Scientific Reports 3 2810 EP –
[17] Stone M 1974 Journal of the Royal Statistical Society. Series B (Methodological) 36
[18] Picard R R and Cook R D 1984 Journal of the American Statistical Association 79 URL http://www.jstor.org/stable/2288403
[19] Murphy K P (ed) 2012 Chapter 14, Machine learning: a probabilistic perspective (MIT Press)
[20] Dam H C, Nguyen V C, Pham T L, Nguyen A T, Terakura K, Miyake T and Kino H 2018 *Journal of the Physical Society of Japan* 87 113801 (Preprint https://doi.org/10.7566/JPSJ.87.113801) URL https://doi.org/10.7566/JPSJ.87.113801

[21] Kohavi R and John G H 1997 *Artificial Intelligence* 97 273 – 324 ISSN 0004-3702 relevance URL http://www.sciencedirect.com/science/article/pii/S000437029700043X

[22] Liu H and Yu L 2005 *IEEE Transactions on Knowledge and Data Engineering* 17 ISSN 1041-4347

[23] Duangsoithong R and Windeatt T 2009 Relevance and redundancy analysis for ensemble classifiers *Machine learning and data mining in pattern recognition* ed Perner P (Springer Berlin Heidelberg) pp 206–220

[24] Almuallim H and Dietterich T 1991 Learning with many irrelevant features *The Ninth National Conference on Artificial Intelligence* (AAAI Press, Menlo Park) p 547552

[25] Biesiada J and Duch W 2007 *Computer Recognition Systems 2, Advances in Soft Computing, Springer, Berlin, Heidelberg* 45

[26] Kvalseth T O 1985 *The American Statistician* 39

[27] Yu L and Liu H 2004 *J. Mach. Learn. Res.* 5 1205–1224

[28] Visalakshi S and Radha V 2014 A literature review of feature selection techniques and applications: Review of feature selection in data mining 2014 IEEE International Conference on Computational Intelligence and Computing Research pp 1–6

[29] Hastie Trevor Tibshirani Robert F J 2009 *The Elements of Statistical Learning* (Springer)

[30] TG D 2000 *Multiple Classifier Systems, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg* 1857

[31] Zhou Z H (ed) 2012 *Ensemble Methods: Foundations and Algorithms* (Chapman & Hall CRC)

[32] Breiman L 1996 *Machine Learning* 24 123–140 ISSN 1573-0565 URL https://doi.org/10.1007/BF00058655

[33] PMyers H (ed) 1997 *Introductory Solid State Physics (Second Edition)* (CRC Press)

[34] Hirayama Y, Takahashi Y, Hirosawa S and Hono K 2017 *Scripta Materialia* 138 62 – 65 ISSN 1359-6462 URL http://www.sciencedirect.com/science/article/pii/S1359646217302737

[35] Suzuki H 2017 *AIP Advances* 7 056208 (Preprint https://doi.org/10.1063/1.4973799)