Predicting Band Gap of Transition Metal Trihalides using Machine Learning

N. K. Nepal and M. P. Ghimire

Journal of Nepal Physical Society

Volume 9, Issue 2, December 2023

ISSN: 2392-473X (Print), 2738-9537 (Online)

Editor in Chief:

Dr. Hom Bahadur Baniya

Editorial Board Members:

Prof. Dr. Bhawani Datta Joshi
Dr. Sanju Shrestha
Dr. Niraj Dhital
Dr. Dinesh Acharya
Dr. Shashit Kumar Yadav
Dr. Rajesh Prakash Guragain

JNPS, **9** (2): 34-41 (2023)

DOI: https://doi.org/10.3126/jnphyssoc.v9i2.62287

Published by:

Nepal Physical Society
P.O. Box: 2934
Tri-Chandra Campus
Kathmandu, Nepal
Email: nps.editor@gmail.com

JNPS 9 (2): 34-41 (2023)
ISSN: 2392-473X (Print), 2738-9537 (Online)
Research Article *© Nepal Physical Society*
DOI: https://doi.org/10.3126/jnphyssoc.v9i2.62287

Predicting Band Gap of Transition Metal Trihalides using Machine Learning

N. K. Nepal and M. P. Ghimire*
Central Department of Physics, Tribhuvan University, Kirtipur-44613, Kathmandu, Nepal
*Corresponding Email: madhav.ghimire@cdp.tu.edu.np

Received: *3rd October, 2023;* ***Revised:*** *17th November, 2023;* ***Accepted:*** *21st December, 2023*

ABSTRACT

Machine Learning (ML) is a tool that finds pattern in data sets for analysis and prediction and is very useful in various sectors such as materials discovery, predicting material properties, medical science, and data science etc. From predicting lattice parameters, stability, band gap, and many other material properties with low cost and good accuracy, ML is becoming popular among material scientists. Recently, a new compound in a two-dimensional (2D) realm (transition metal trihalides) is getting immense interest due to its intriguing properties such as magnetic order in lower dimensions. In this work, we have selected transition metal trihalides to predict band gaps using ML. For the dataset, we have used some compound and elemental properties extracted from the materials project database and periodic table for the features set, and band gaps within the range of 0-3 eV for target variables. We performed the least absolute shrinkage and selection operator (LASSO) for feature selection and selected five features from a set of 24 features. After training and testing our data in four types of ML algorithms (support vector regression (SVR), kernel ridge regression (KRR), gradient boosted regression tree (GBRT), and random forest regression (RFR)), the GBRT model is found to be best with the lowest mean absolute error (MAE) < 0.27 eV and root mean squared error (RMSE) < 0.34 eV.

Keywords: Machine Learning, Gradient Boosted Regression Tree, Random Forest Regression, Kernel Ridge Regression, Support Vector Regression.

I. INTRODUCTION

Since the synthesis of Graphene [1], many 2D materials have gained immense research interest due to their promising applications in the field of spintronics, optoelectronics etc. Graphene, Transition Metal dichalcogenides (TMDs), MXenes, Hexagonal-Boron Nitrides, 2D Perovskites are most prominent materials having huge number of research conducted. Recently, new emerging 2D materials (Transition Metal Trihalides) are attracting researchers that has even shown intrinsic magnetic properties. After the discovery of intrinsic magnetism in $Cr_2Ge_2Te_6$ [2], and CrI_3 [3], research interest on transition metal trihalides has surged rapidly. Even many works on obtaining magnetic properties on layered types of materials are in increasing order. Transition metal trihalides (MX_3, M = transition metals, X = halogens) are a layered type of materials such that each of three layers are stacked in order of ABC with weak Van der Walls interaction between two layers [4]. The oxidation number of M is +3 and that for X is -1. These materials are mostly found in trigonal, hexagonal and monoclinic types of crystal systems. Furthermore, many research works have been carried out on the compound of type CrX_3, and VX_3, many compounds within this system are still unexplored, and many materials are yet to be discovered.

From empirical experiments to computational simulation, the advancement of tools and methods can be divided into four paradigms. These are also known as the four paradigms of material science [5–7]. The first paradigm, called empirical science, which includes different experimental works to find the properties of materials. These are costly and time consuming. The second paradigm is called model based theoretical science, where different

laws have been proposed to find material properties. The third paradigm of science is known as computational science. The advancement of computational tools and simulation based on complex theories such as density functional theory (DFT) has been a crucial development of tools in the field of material science. Due to these advancements, a new tool has been emerging in the field of material science, known as (big) data-driven science. In the fourth paradigm (data-driven science), ML is taking research in a new path for predicting different properties of novel materials. It includes discovering new materials as well as predicting properties of materials. Realizing about pros and cons of different tools, all other tools are somehow more costly, and time consuming than data-driven science. With the help of different statistical methods to analyze the huge amount of data, recent research has proven that ML can predict properties of different materials with low computational cost and less time with good accuracy. Rajan et al. [8] used kernel ridge regression (KRR), support vector regression (SVR), gaussian process regression (GPR), and Bootstrap Aggregating Regression algorithms to predict Band Gap of Fuctionalized MXene. Among these models, they have found that the GPR model predicts the band gap with the lowest root-mean-squared error (RMSE) of 0.14 eV. Furthermore, Zhang et al. [9] have worked to predict the band gap of Double Perovskites. They used the random forest regression model and found that the model has predicted band gap with model accuracy of 85.6% with a root mean square error of 0.64 eV.

Machine Learning (ML) is a science of programming that enables the machine (or computers) to act or work by analyzing the given data [11]. In particular, the ML algorithm analyzes the pattern of the provided data during training and predicts unseen data based on the model obtained during training [12]. ML further can be divided into supervised, unsupervised and reinforcement learning. Supervised learning includes learning using both features and labels [13], however, in unsupervised learning (such as clustering) there are no labels for each set of features. Furthermore, supervised learning is divided into mainly two types of learning [14]: Classification and Regression.

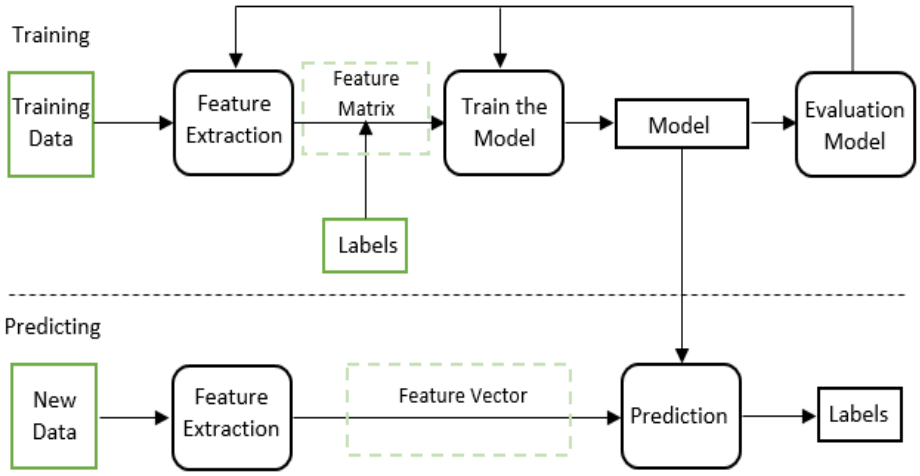

Fig. 1: Supervised Machine Learning Workflow. Adapted from [10].

Recently, discovery of van der waals (vdW) magnetic materials and finding 2D magnetic topological insulators has found to predict using ML [15,16], which also include MX_3 type of compound with various compound configurations. Rhone et al. have predicted magnetism and thermodynamical stability of $A_2B_2X_6$ type vdW materials combining both DFT and ML. [17] With the atomic properties as input and magnetic moment, magnetic excitation energy and the formation energy as target variables, they have predicted several thermodynamically stable magnetic vdW materials employing few ML model such as kernel ridge regression, extra trees regression, and neural network regression. However, up to our knowledge prediction of band

gap of such compound has not found to observed using ML. Prediction of band gap is essential as it provides materials electronic behaviors such as conductors, semiconductors, and insulators. In addition to magnetism, semi conducting properties is also essential to apply such vdW materials in spintronics applications. Here, we used ML to predict band gap of MX_3 compounds. We used several feature selection techniques such as Pearson's correlation coefficient and least absolute shrinkage and selection operator (LASSO) regression. After feature selection we used several ML algorithms to train and test the data, and find that GBRT model to be predicted with lowest mean absolute error and root mean squared error.

II. METHODS

A. Data Pre-processing and Feature Engineering

As ML provides prediction based on the data we have provided, the collection of data should be done from a standard repository. So for the data collection, we have used the Materials project database [18]. First, we have collected all the materials of the MX_3 system (M = Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Nb, Mo, Tc, Ru, Rh, Pd, Ag, Cd, Hf, Ta, W, Re, Os, Ir, Pt, Au, Hg and X= (F, Cl, Br, I). After, we included materials with band gaps within 0-3 eV range. Further, we have only selected compound that either belongs to the Trigonal, Hexagonal or Monoclinic Crystal systems. A total of 70 compounds, including compounds that have both or all of the crystal systems has been collected. All the compound properties of those materials were taken from the Materials Project Database. Furthermore, we also have included different properties of elements (such as atomic radius, Pauling's electronegativity etc) were obtained from the Periodic Table. Altogether, we have collected 24 properties as a feature set and band gap as target variables. The band gap we have used are Perdew-Burke-Ernzerhof (PBE) band gap obtained by using generalized-gradient approximation (GGA).

After the data collection, we worked on feature engineering. Feature selection is an important task during data pre-processing. The model with a large number of feature sets can learn from the noise of irrelevant feature sets, thus diminishing the performance of the model. It is also important because it can mitigate the computational cost of the modeling. Various statistical methods are known for different types of ML work.

We have used a LASSO regression [19] method for feature selection. This process aims to reduce the sum of squared differences (L2 norm) while simultaneously applying a regularization technique that encourages a reduction in the sum of mean absolute differences (L1 norm) on the coefficients (β)[8]

$$L(\beta) = \left|\left|y - X\beta\right|\right|_2^2 + \alpha\left|\left|\beta\right|\right|_1 \text{ ---------- (1)}$$

where, $\left|\left|\ldots.\right|\right|_1$ and $\left|\left|\ldots.\right|\right|_2$ are the L1 and L2 norms, and α is a parameter that controls the shrinkage.

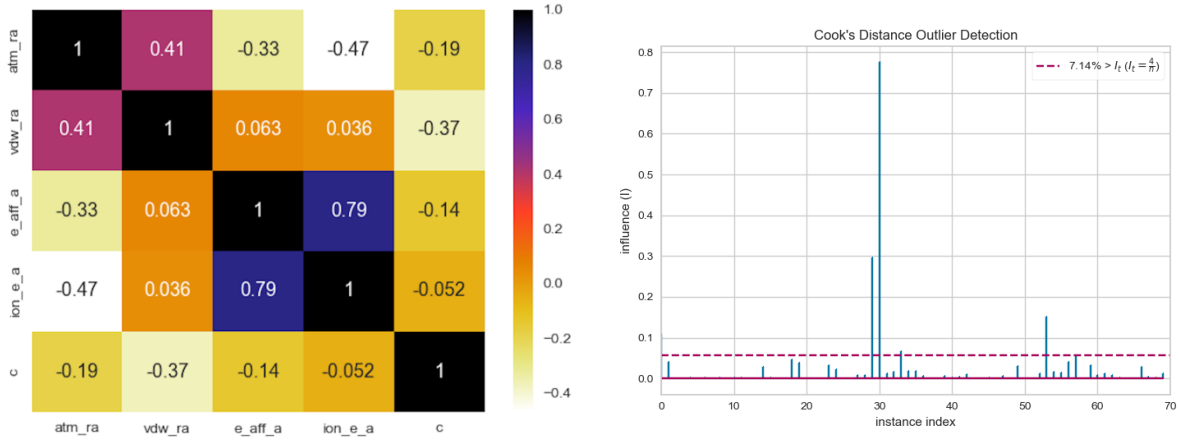

Fig. 2: The plot of Pearson's correlation exists between each set of features obtained after lasso and Cook's distance analysis (upper). Outlier analysis using Cook's distance (lower).

Setting α = 0.01, we have found 11 features with non-zero coefficients. Among those features, we have picked the top five features. With such, now our 24 features set has been reduced to 5 features

set. Furthermore, for the detection of outliers within data, we have calculated Cook's distance using a Yellowbrick Api [20]. By using a threshold value of 4/n, n is the total number of data samples, indicated by red dot lines in FIG.2. (lower), we have found 4 data samples as an outlier and hence removed from the data set. Cook's distance measures the combination of residuals and leverage of each data sample that might cost the model performance. Detecting and removing outliers will help to improve accuracy of ML algorithm. In addition, Pearson's correlation coefficient also has been calculated to find the correlation between two features in order to neglect features that have $|P| > 0.85$ (P = Pearson's correlation coefficients). Having highly correlated data won't improve the model accuracy but might create complexity while training the ML algorithm. Hence it is better to remove highly correlated inputs from the training data sets. Before LASSO regression, we found six such features and hence removed them, however, after LASSO and Cook's Distance, we did not find any $|P| > 0.85$ between the remaining five features which can be seen on heat map from FIG.2. FIG.3. is a Kernel Density Estimation (KDE) plot of standardized input feature variables and KDE plot of the target variable. The KDE plot here estimates the probability density of random variables using a Gaussian kernel function.

After all, we are left with 66 data samples and 5 features. As of now, our data has been prepared, however, we cannot feed the data set to our ML model directly. Many ML models need some kind of scaling of data that converts input variables belonging to a certain range. A standard scaling function found in the Sci-Kit Learn library has been used for the scaling of the data. Standard Scaling scaled the input data as, $X_{sc} = (x-u)/s$, where u is the mean of the training data sample and s is the standard deviation of the training data sample.

Furthermore, a sci-kit-learn train-test splitting module has been used to separate 85% of data for the training purpose and 15% of the data for testing purposes for LASSO regression. In addition, splitting 90% of data for the training set and 10% of data as a test set using the same module was used while fitting the model with all four algorithms.

B. ML Algorithms

For the model analysis, we have used four regression ML algorithms: kernel ridge regression (KRR), support vector regression (SVR), and gradient boosted regression tree (GBRT), and random forest regression (RFR). KRR [21–24] is a supervised ML algorithm which is a combination of the kernel method and ridge regression. For the prevention of overfitting during training of the model, if L2 regularization has been used, it is called ridge regression. The L2 regularization [25,26] is given by the formula $\text{argmin}_w\{||Xw - y||^2_2 + \lambda||w||^2_2\}$. Where, X is n by d matrix defined

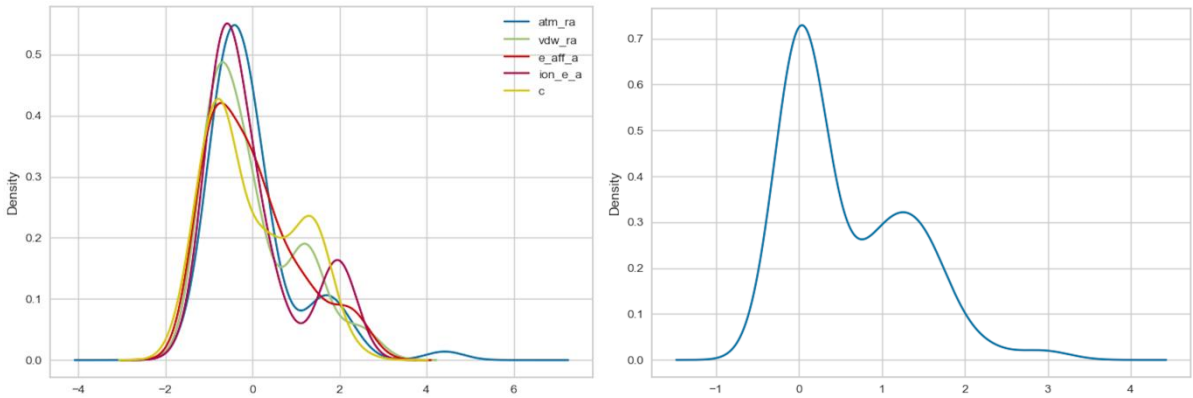

Fig. 3: Kernel Density Estimation(KDE) plot for the standardized feature space (upper) and for target variable (lower) using Gaussian Kernel.

by $X_{ij} = x_{ij}$ (n = number of data points, d = total number of features) and y $(y_1,y_2,....,y_n)$ is n-dimensional vector. λ, a positive parameter, which control weight, w, i.e., bigger the λ smaller the $||w||^2$. Now the kernel method using ridge regression maps the samples in a high dimensional space with nonlinear mapping and the model learns from the training. The predicted value of the target variable in the KRR model is obtained as the equation,

$$y = \sum_{i=1}^{n} \alpha_i \Phi\left(x_j, \hat{x}\right) \quad \text{---------- (2)}$$

After obtaining α, a n×1 unknown solution vectors, provided by solving the equation, $(K + \lambda nI)\alpha = y$, where K is a kernel matrix build from training data set as $K_{ij} = \Phi(x_i,x_j)$ and y is a n×1 regressand input vector corresponded to X [21]. Furthermore, talking about SVR, the main idea is to find f(x) that must have not greater than ε deviation from the actual target variables (y_i) [24,27]. It is a regression generalization of a Support Vector Machine (SVM) that forms a nonlinear regression function using the kernel method. In general, SVR optimizes the problem by finding a convex ε-insensitive loss function that must be minimized and hence finding the flattest tube that accumulates most of the training instances [28]. The main difference between KRR and SVR is the loss function used in these models. The ε-insensitive region around the function is called ε-tube. If the predicted values lie inside the tube, the loss is zero, otherwise, the loss is equal to the difference between the predicted value and the radius of ε tube [29].

Moreover, we also have used two ensemble types of regression models: GBRT and RFR. The ensemble method combines the results of the base estimators to improve generali- zability over one estimator. A GBRT [30,31] is a type of ensemble method belonging to a family of boosting methods. In this model, the prediction y_i for input x_i is given by the equation below.

$$y_i = F_m(x_i) = \sum_{m=1}^{M} h_m(x_i) \quad \text{---------- (3)}$$

where h_m are estimators known as weak learners, and that M refers to an n-estimator parameter. This boosting algorithm is also known to be built in a greedy fashion as all other boosting algorithms. i.e.

$$F_m(x) = F_{m-1}(x) + h_m(x) \quad \text{---------- (4)}$$

Here the newly delivered tree h_m is geared up which will reduce a sum of losses L_m, given the preceding ensemble F_{m-1}.

Random Forest Regression [24,32,33] is another type of ensemble method belonging to the family of averaging methods. It fits a number of decision trees in a randomly obtained subset of data [9] and finally uses averaging to get more accuracy on prediction and to mitigate overfitting.

For all models, all the parameters were optimized using the grid search method found in the sci-kit learn library. We also have used 5-fold cross-validation and after five-fold cross-validation, we again searched the parameters for the respective algorithm and again fitted the ML model and hence predicted the band gap for a test set of data. For the ML model accuracy observation, we have calculated the, mean absolute error (MAE) and root mean squared error (RMSE).

III. RESULTS AND DISCUSSION

In this work, we have collected some elemental and compound properties. All the data were obtained from the Materials Project Database and a periodic table. After that, we worked on feature selection using LASSO regression. The starting number of feature (24 features) has been reduced to 5. After taking the top five features obtained by using LASSO regression, we have also eliminated outliers calculated using the cook's distance. We also calculated and plotted the Pearson correlation coefficient to observe the correlation between the two features. We have eliminated one of each correlated feature having correlation coefficients $|p| > 0.85$. Then, we used mainly four ML models, SVR, KRR, GBRT, and RFR. Among them, SVR and KRR are Kernel based regressions in which the main difference is the type of loss function. For the kernels, we have used grid search to choose the best between radial basis function (RBF) and polynomial (Poly). Further, other parameters have also been optimized using grid search, best kernels and parameters have been used for fitting the model. Such a process has been performed two times. First, we used five-fold cross-validation, and then we fitted the model for prediction. Before providing raw data as input, we first separated the data set into a train set of data and a test set of data, 90 % for training and 10 % for testing. Furthermore, We have Standardized the data using standard scaling as a data pre-processing. The other two models, GBRT, and RFR are ensemble types of ML models, which worked based on trees. Since they are tree-based models normalization of the data is not mandatory [34]. We have split the data as train and test set, 90 % of data for the training set and 10 % of data for the test set. We also optimized the parameters of GBRT and RFR using the grid-search method and then fitted the model with optimized parameters.

From the SVR model, we have predicted the band gap for the test set of data with a MAE of 0.37 eV and RMSE of 0.49 eV. While, the KRR model predicted the target variable with MAE < 0.31 eV, and RMSE < 0.38 eV. On comparing the performance obtained from two regression models that used kernel functions, we got better performance from KRR i.e. less MAE and less

RMSE from KRR model. Again, the GBRT model has predicted the band gap using a test set of data with MAE < 0.27 eV and RMSE < 0.34 eV. Finally, we have used RFR and found that this model has predicted band gap with MAE < 0.34 eV and RMSE < 0.41 eV. On comparing these two ensemble methods of ML, GBRT and RFR, better performance was observed from GBRT.

In summary, we have four models that have predicted band gaps with some errors, and the best result was obtained from GBRT Model. The corresponding feature importance plot using GBRT and RFR is shown in Figure 4. Further investigating the feature importance plot, we see that in both ensemble model, the importance of feature for prediction are almost same, however, that of electron affinity, atomic radius and van der waal radius of transition metal atom is higher in RFR model than in GBRT.

IV. CONCLUSIONS

Our research focuses on leveraging machine learning (ML) for predicting materials properties, showcasing a growing interest in this field. ML has proven instrumental for material scientists, accelerating the discovery of novel materials with remarkable electronic, magnetic, and notably superconducting critical temperatures, all achieved in less time with high accuracy. The key to achieving precise results lies in the availability of extensive and standardized datasets. While some researchers rely on high-throughput density functional theory (DFT) calculations to generate datasets for ML models, our approach involves extracting data from a widely recognized standard database.

To enhance model accuracy, various statistical methods have been employed in our work. Notably, the feature selection process, incorporating techniques such as LASSO regression, Pearson's correlation, and Cook's distance for outlier detection, has proven effective. Additionally, optimizing hyper-parameters is crucial for improving prediction capabilities. In our study, we employed a grid-search process to systematically test and identify the best parameters among all possibilities.

In predicting the band gap of MX_3, we evaluated four ML models with optimized hyper-parameters. Our methodology not only underscores the importance of robust datasets but also highlights the significance of meticulous feature selection, outlier detection, and hyper-parameter optimization in achieving accurate predictions for materials properties which provide good accuracy of prediction. TABLE I. shows a summary of the performance of ML models.

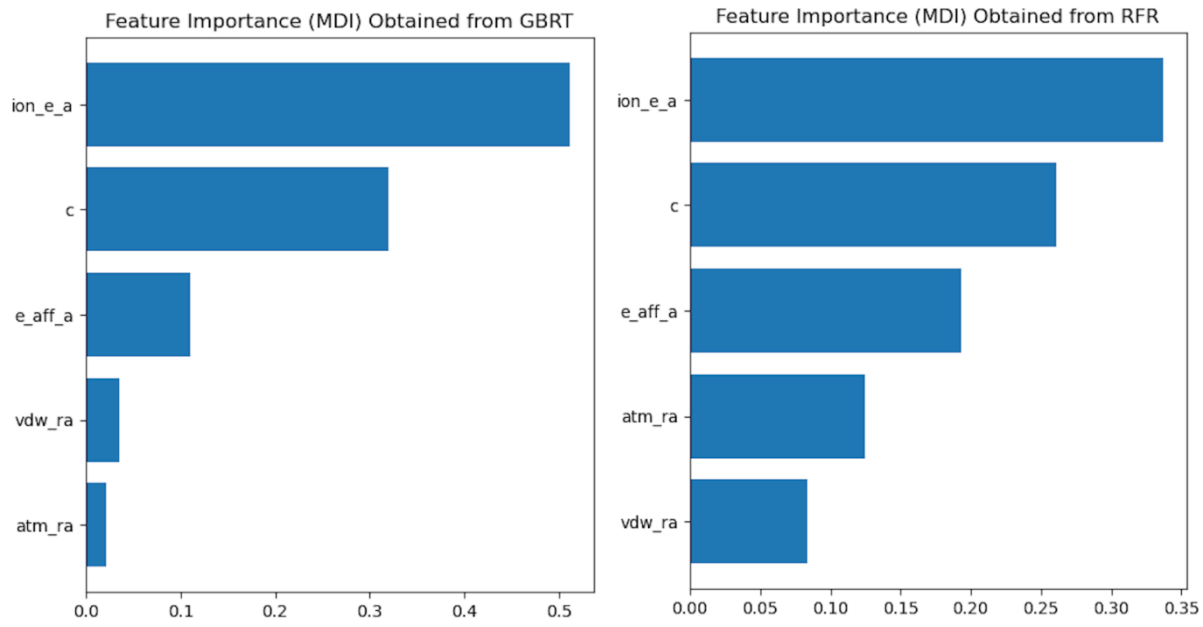

Fig. 4: Feature Importance plot Obtained using GBRT (upper) and Obtained using RFR (below).

Table 1: Table of MAE and RMSE obtained from SVR, KRR, GBRT, RFR models while predicting band gap for test set of data.

Models	MAE (eV)	RMSE (eV)
SVR	0.37	0.49
KRR	0.31	0.38
GBRT	0.27	0.34
RFR	0.34	0.41

ACKNOWLEDGMENTS

This work was supported by a grant from UNESCO-TWAS and the Swedish International Development Cooperation Agency (SIDA) with award numbers 21-377 RG/PHYS/ASG. The views expressed herein do not necessarily represent those of UNESCO-TWAS, SIDA, or its Board of Governors. M.P.G. and N.K.N. thanks University Grants Commission, Nepal for the collaborative research grant and fellowship. We acknowledge NVIDIA for providing the license package of Deep learning, a part of machine learning. Technical support from Dr. Rajendra Adhikari, Kathmandu University is gratefully acknowledged.

REFERENCES

[1] Novoselov, K. S.; Geim, A. K.; Morozov, S.V.; Jiang, D. E.; Zhang, Y.; Dubonos, S. V.; Grigorieva, I. V. and Firsov, A. A. Electric field effect in atomically thin carbon films. *Science*, **306** (5696): 666-669 (2004).

[2] Gong, C.; Li, L.; Li, Z.; Ji, H.; Stern, A.; Xia, Y.; Cao, T.; Bao, W.; Wang, C.; Wang, Y. and Qiu, Z. Q. Discovery of intrinsic ferromagnetism in two-dimensional van der Waals crystals. *Nature*, **546**(7657): 265-269 (2017).

[3] Huang, B.; Clark, G.; Navarro-Moratalla, E.; Klein, D. R.; Cheng, R.; Seyler, K. L.; Zhong, D.; Schmidgall, E.; McGuire, M. A.; Cobden, D. H. and Yao, W. Layer-dependent ferromagnetism in a van der Waals crystal down to the monolayer limit. *Nature*, **546**(7657): 270-273 (2017).

[4] Liu, P.; Lu, F.; Wu, M.; Luo, X.; Cheng, Y.; Wang, X. W.; Wang, W.; Wang, W. H.; Liu, H. and Cho, K. Electronic structures and band alignments of monolayer metal trihalide semiconductors MX 3. *Journal of Materials Chemistry C*, **5**(35): 9066-9071 (2017).

[5] Agrawal, A. and Choudhary, A. Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. *Apl Materials*, **4**(5): (2016).

[6] Wei, J.; Chu, X.; Sun, X.Y.; Xu, K.; Deng, H. X.; Chen, J.; Wei, Z. and Lei, M. Machine learning in materials science. *InfoMat*, **1**(3): 338-358 (2019).

[7] Schleder, G. R.; Padilha, A. C.; Acosta, C. M.; Costa, M. and Fazzio, A. From DFT to machine learning: recent approaches to materials science–a review. *Journal of Physics: Materials*, **2**(3): 032001 (2019).

[8] Rajan, A. C.; Mishra, A.; Satsangi, S.; Vaish, R.; Mizuseki, H.; Lee, K. R. and Singh, A. K. Machine-learning-assisted accurate band gap predictions of functionalized Mxene. *Chemistry of Materials*, **30**(12): 4031-4038 (2018).

[9] Zhang, J.; Li, Y. and Zhou, X. Machine-Learning Prediction of the Computed Band Gaps of Double Perovskite Materials. *arXiv preprint arXiv:2301.03372* (2023).

[10] Ghareeb, S.; Hussain, A. J.; Al-Jumeily, D.; Khan, W.; Al-Jumeily, R.; Baker, T.; Al Shammaa, A. and Khalaf, M. Evaluating student levelling based on machine learning model's performance. *Discover Internet of Things*, **2**(1): 3 (2022).

[11] Géron, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc." (2022).

[12] Zhou, Z. H. *Machine learning*. Springer Nature (2021).

[13] Mahesh, B. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, **9**(1): 381-386 (2020).

[14] Nasteski, V. An overview of the supervised machine learning methods. *Horizons. B*, **4**: 51-62 (2017).

[15] Rhone, T. D. Bhattarai, R. Gavras, H. Lusch, B. Salim, M. Mattheakis, M. Larson, D. T. Krockenberger, Y. and Kaxiras, E. Artificial Intelligence Guided Studies of van der Waals Magnets. *Advanced Theory and Simulations*, 2300019 (2023).

[16] Xu, H.; Jiang, Y.; Wang, H. and Wang, J. Discovering two-dimensional magnetic topological insulators by machine learning. *arXiv preprint arXiv:2306.14155* (2023).

[17] Rhone, T. D.; Chen, W.; Desai, S.; Torrisi, S. B.; Larson, D. T.; Yacoby, A., & Kaxiras, E. Data-driven studies of magnetic two-dimensional materials. *Scientific reports*, **10**(1): 15795 (2020).

[18] Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G. and Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL materials*, **1**(1): (2013).

[19] Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **58**(1): 267-288 (1996).

[20] http://www.scikit-yb.org/en/latest/

[21] You, Y.; Demmel, J.; Hsieh, C. J. and Vuduc, R. Accurate, fast and scalable kernel ridge regression on parallel and distributed systems. In *Proceedings of the 2018 International Conference on Supercomputing*, 307-317 (2018).

[22] Vovk, V. Kernel ridge regression. In *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*. Berlin, Heidelberg: Springer Berlin Heidelberg, 105-116 (2013).

[23] Welling, M. Kernel ridge regression. *Max Welling's classnotes in machine learning*, 1-3 (2013).

[24] Lewis Atwell, T.; Townsend, P. A. & Grayson, M. N. Machine learning activation energies of chemical reactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **12**(4): e1593 (2022).

[25] Cortes, C.; Mohri, M. and Rostamizadeh, A. L2 regularization for learning kernels. *arXiv preprint arXiv:1205.2653* (2012).

[26] Van Laarhoven, T. L2 regularization versus batch and weight normalization. *ArXiv preprint arXiv:1706.05350* (2017).

[27] Smola, A. J. and Schölkopf, B. A tutorial on support vector regression. *Statistics and computing*, **14:** 199-222 (2004).

[28] Awad, M. and Khanna, R. *Efficient learning machines: theories, concepts, and applications for engineers and system designers.* Springer nature, p. 268 (2015).

[29] Drucker, H.; Burges, C. J.; Kaufman, L.; Smola, A. and Vapnik, V. Support vector regression machines. *Advances in neural information processing systems*, **9** (1996).

[30] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. and Vanderplas, J. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, **12:** 2825-2830 (2011).

[31] Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics:* pp.1189-1232 (2001).

[32] Segal, M. R. Machine learning benchmarks and random forest regression (2004).

[33] Biau, G. Analysis of a random forests model. *The Journal of Machine Learning Research*, **13:** pp.1063-1095 (2012).

[34] Li, W.; Wang, Z.; Xiao, X.; Zhang, Z.; Janotti, A.; Sanguthevar, R. and Medasani, B. Predicting band gaps and band-edge positions of oxide perovskites using DFT and machine learning. *arXiv preprint arXiv:2008.12412* (2020).