# Predicting Band Gap of Carbon and Nitrogen-Based Photocatalysts Using Machine Learning

Pouya Pishkar[1,*]
[1] School of Metallurgy and Materials Engineering, College of Engineering, University of Tehran, Tehran, Iran

**ABSTRACT**

*The increasing need for sustainable energy solutions has driven the development of photocatalytic materials with tailored band gap properties. In this study, a Random Forest Regressor model was developed to predict the band gap of materials incorporating carbon and nitrogen. The model utilized a dataset comprising 3626 materials with features such as density, energy above the hull, magnetic ordering, and structural parameters, with data sourced from Materials Project. The predictive performance of the model was evaluated using metrics including MAE = 0.450 eV, RMSE = 0.677 eV, and a $R^2$ = 0.813, indicating strong predictive capability. Feature importance analysis revealed that magnetic ordering and density were the most influential factors, contributing 22% and 16.9%, respectively, to band gap predictions. A correlation heatmap further highlighted the relationships among material properties, with density showing a strong negative correlation (-0.98) with band gap values. The findings demonstrate the effectiveness of machine learning in accurately predicting band gaps, overcoming the limitations of traditional computational methods, and enabling the rapid identification of promising photocatalysts. This approach significantly accelerates the discovery of materials for applications in water splitting, $CO_2$ reduction, and environmental remediation, supporting the transition to sustainable energy solutions.*

***Keywords:*** **Photocatalysts, Band gap prediction, Machine learning, Carbon and nitrogen-based materials, Random forest regressor**

## 1. *INTRODUCTION*

*The demand for efficient and sustainable energy solutions has intensified the search for advanced photocatalytic materials capable of harnessing solar energy for applications such as water splitting [1], carbon dioxide reduction [2], and environmental remediation [3]. Photocatalysts, particularly those with tunable band gaps, are central to this endeavor as they enable the absorption of light and the generation of electron-hole pairs essential for driving chemical reactions [4] to [6]. While a broad range of materials, including oxides [7], nitrides [8], and carbon-based compounds [9], have been explored, the development of hybrid systems combining multiple material types has emerged as a powerful strategy to overcome the limitations of single-component photocatalysts [10].*

*The energy gap of a semiconductor, known as the band gap, defines the energy required to move an electron from the valence band to the conduction band. Accurately measuring this energy is vital for understanding and predicting the photophysical and photochemical behaviors of semiconductors. This value is particularly important in discussions surrounding the photocatalytic capabilities of these materials [11].*

*Carbon- and nitrogen-containing compounds have garnered significant interest due to their unique electronic and structural properties, offering opportunities for tailoring band gap energies to match specific photocatalytic applications. These materials, often inspired by their natural abundance and environmental compatibility, provide a versatile platform for designing next-generation photocatalysts [11] [12]. Researchers have focused on modifying their chemical composition, nanostructures, and defect states to achieve optimal light absorption and catalytic performance, thereby opening pathways for more efficient solar-driven reactions [13] to [15].*

*Determining the properties of these materials through experimental or computational approaches can be a lengthy process. However, the development of faster computing capabilities has enabled the use of density-functional-theory-based (DFT) methods to efficiently calculate band gaps within a practical timeframe [16] to [17]. Despite this, fundamental gaps derived using the local-density or generalized-gradient approximations (LDA or GGA) tend to be underestimated. To address this, the GW method, which is based on many-body perturbation theory [18], can provide more accurate band gap estimations. Unfortunately, such approaches are computationally intensive and often too costly for practical application [16]. Recently, statistical learning has gained traction as a valuable tool for predicting the structures [19] and properties [20] to [23] of various material classes. These techniques can effectively estimate properties like paramagnetic or ferromagnetic behavior [24] to [25], density of states [26], band gaps [27], toxicity [28], light absorption [29], and drug loading capacity [30] within feasible timeframes [16].*

*To address these challenges, machine learning (ML) has emerged as a powerful tool, offering rapid and accurate predictions based on material properties. By leveraging datasets of known materials and their band gaps, ML models can identify patterns and relationships that are not readily apparent through conventional methods [31].*

*In this study, we develop a Random Forest Regressor model to predict the band gap of photocatalysts containing carbon, nitrogen, and potential metallic components. The model is trained on a dataset incorporating various physicochemical and structural features of these materials, aiming to evaluate the influence of metallic inclusions on band gap predictions. Additionally, we employ advanced visualization techniques to interpret the results and uncover the contributions of different material properties.*

*This work aims to demonstrate the feasibility and effectiveness of ML-driven approaches in accelerating the discovery and design of novel photocatalysts, particularly those with complex compositions containg carbon and nitrogen. By reducing the reliance on experimental and computationally expensive methods, this study contributes to the broader effort of advancing sustainable technologies for energy and environmental applications.*

## 2. *MATHEMATICAL BACKGROUND*

*To assess the predictive performance of each model on the test dataset, we utilized several metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Explained Variance ($R^2$) [32].*

*In statistics, the MAE quantifies the average magnitude of errors between paired observations that represent the same phenomenon. It provides a straightforward measure of prediction accuracy by averaging the absolute differences between predicted and actual values* [33].

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - x_i| \qquad (1)$$

*where $y_i$ is the prediction and $x_i$ the true value and N is the total number of data points. In statistics, the MSE, is a metric used to evaluate an estimator's performance. It represents the average of the squared differences between estimated values and the actual values, providing insight into the accuracy of predictions or estimates* [34] *to* [35].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}i)^2 \qquad (2)$$

*In this context, $y_i$ represents the i-th observed value, while $\hat{y}_i$ refers to the corresponding predicted value. Additionally, n denotes the total number of observations in the dataset.*

*The RMSE is the square root of the mean of the squared differences between observed and predicted values. These differences are termed residuals when calculated on the data sample used for estimation, as they reference the model's estimates. When computed on out-of-sample data (i.e., the full dataset compared to true values rather than estimates), they are referred to as errors or prediction errors. [36].*

$$RMSE=\sqrt{MSE} \qquad (3)$$

*In statistics, the coefficient of determination, commonly symbolized as $R^2$, indicates the percentage of variability in the dependent variable that can be accounted for by the independent variable(s). It serves as a measure of how well the model explains the observed data [37].*

$$R^2 = 1 - \frac{RSS}{TSS} \qquad (4)$$

*Where $R^2$ represents the coefficient of determination. RSS stands for the sum of squares of residuals. TSS refers to the total sum of squares.*

## 3. *RANDOM FOREST REGRESSOR MODEL*

*Based on the previous work by Priyanga et al [38], it was found that the Random Forest model outperformed other models in this type of prediction, achieving an accuracy percentage of 91%.*

*Random Forest is an ensemble method introduced by Breiman in 2001[39], which consists of multiple decision trees. As suggested by its name, the construction of a random forest is done in a random manner. It is built from numerous decision trees that are uncorrelated with one another. The method employs random sampling with replacement, which involves two key processes: data random sampling and feature random sampling. In data random sampling, random subsets are selected from the dataset, while in feature random sampling, a random selection of features is made from the available set. Importantly, no pruning is applied to the individual trees, allowing each tree to grow freely. Random Forest is versatile, capable of addressing both classification and regression problems. In classification tasks, the final output is determined by a majority vote from all the trees, while in regression tasks, the result is the average of the predictions from all the trees [40].*

## 4. *SETTING UP THE ENVIRONMENT*

*The main libraries used in the code are pandas, numpy**,** sklearn**,** matplotlib, and seaborn. Pandas is used for data manipulation and analysis, offering powerful tools for handling data frames, cleaning, and transforming data [41].*

*Numpy is essential for numerical computing, providing support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions [42].*

*Scikit-learn (sklearn) is a machine learning library that offers a wide range of algorithms for data modeling, preprocessing, and evaluation, including models for regression, classification, and clustering [42].*

*Matplotlib and Seaborn are visualization libraries; Matplotlib is used for creating static, animated, and interactive plots, while Seaborn builds on Matplotlib to provide more aesthetically pleasing and complex visualizations, particularly for statistical data [43].*

## 5. *MODEL EVALUATION*

### 5.1 Actual vs. Predicted Band Gap

*The actual vs. predicted chart illustrates the relationship between the actual band gap values and the predicted band gap values generated by the model [44]. The points on the chart represent photocatalytic material compositions as predicted by the model. The line represents the ideal prediction, where the predicted values perfectly match the actual values [44].*

*As shown in Figure 1a, most points lie close to the red line, especially within the range of medium band gaps. This indicates that the model performs well in predicting mid-range values. At very small or very large ranges (both ends of the x and y axes), some points deviate more significantly from the red line, indicating higher errors in these intervals.*

*In the small band gap domain (band gap <1 eV), there is slightly greater scatter among the points. This suggests that the model faces challenges in accurately predicting compositions with small band gaps. In the large band gap domain (band gap > 4 eV), there are fewer data points in this range, and some points exhibit relatively high errors. This may be due to a lack of sufficient training data for this domain.*

*In some cases, particularly in higher ranges, the model tends to underestimate the predicted values compared to the actual ones. This type of error could stem from improper data normalization or insufficient diversity in the features.*

### 5.2. Residual distribution

*This chart illustrates the distribution of residuals, which are the differences between the actual values and the predicted values by the model. The residual distribution is approximately symmetric, indicating that the model generally has balanced errors for higher and lower values (Figure 1b). This behavior suggests that the model does not exhibit systematic bias in predicting the band gap. The mean of the residuals is close to zero, confirming that the model's positive and negative errors tend to cancel each other out on average. Toward the tails of the distribution, there is slight skewness toward positive values, indicating that the model occasionally underestimates the actual values. Some residuals fall within very high or very low ranges, suggesting that the model has high errors for certain specific combinations, which could be due to excessive complexity in these combinations or insufficient information.*

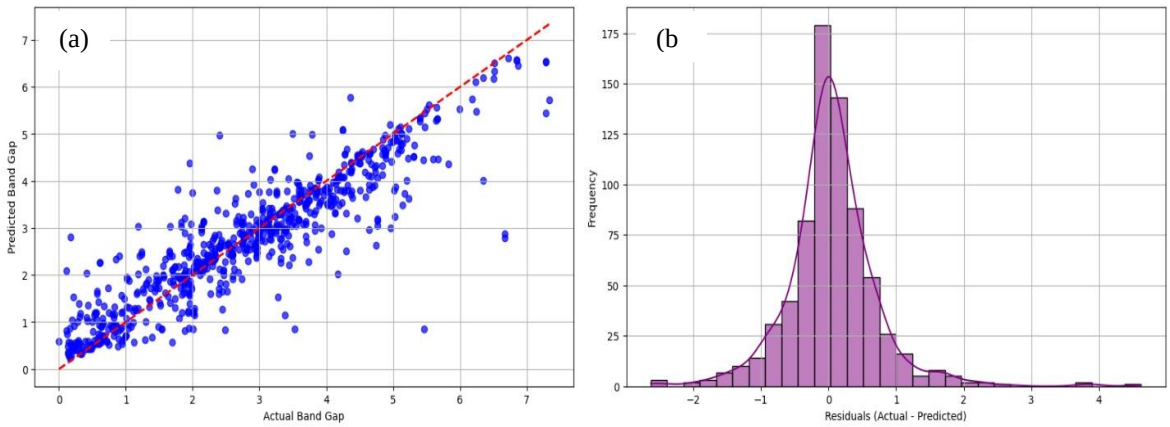


***Fig. 1.*** *a) Actual vs. predicted plot.and b) residual distribution plot.*

## 6. *DATA INSIGHTS AND FEATURE ANALYSIS*

### 6.1 Feature importance

*This chart illustrates the relative importance of each feature in predicting the Band Gap value, calculated using the Random Forest model. As shown in Figure 2a Magnetic ordering_NM was identified as the most significant feature, contributing 22% to the prediction, indicating that the magnetic properties of materials directly influence their electronic behavior and band structure.*

*The second and third most important features are Density and Energy Above Hull, accounting for 16.9% and 16.3% of the importance, respectively. Density, as a key structural parameter, reflects atomic packing and its impact on the material's electron affinity. Meanwhile, Energy Above Hull represents the material's thermodynamic stability, playing a critical role in determining phase stability and semiconductor behavior.*

*Other features, such as Formation Energy, Volume, and Space Group Number, have moderate effects and serve as complementary factors in predicting the Band Gap. Table 1 summarizes a portion of this datas.*

*These findings highlight that the magnetic and structural properties of materials, particularly atomic density and phase stability, have significant impacts on predicting their semiconductor characteristics. This information can be instrumental in designing new materials and optimizing electronic properties for specific applications, such as solar cells [45] or electronic components [46].*

**Table 1.** *A summarize of feature importance datas*

| Feature | Importance |
|---|---|
| Magnetic Ordering_NM | 0.225554 |
| Density | 0.169101 |
| Energy Above Hull | 0.163208 |
| Formation Energy | 0.105137 |
| Total Megnetization | 0.046802 |

**6.2 Distribution of Band gap**

*The data indicate that most materials have a Band Gap in the range of 1 to 3 eV. This range is typically associated with semiconductors that have widespread applications in photocatalysts [47], photovoltaics [48], and ultra-sensitivit sensors [49].*

*The concentration of data in this range suggests that the dataset is well-targeted for photocatalytic applications (Figure 2b).*

*The presence of materials with high Band Gaps (>3 eV) indicates the existence of strong insulators [50]. These materials can be useful for applications such as energy storage [51] or the fabrication of specialized electronic devices [52].*

*Materials with Band Gaps below 1 eV are likely metals or semimetals that are suitable for applications in solar cells [53] or sensing [54].*

*The number of materials with very small (<1 eV) or very large (>4 eV) Band Gaps is relatively low. This imbalance could affect the model's performance in predicting these extreme values.*

*The mean Band Gap is 2.67 eV with a standard deviation of 1.58 eV, indicating significant diversity in the dataset, although most data points are concentrated around the mean.*

*Approximately 50% of the data lies within the range of 1.47 eV to 3.79 eV, encompassing many materials with Band Gaps in this range.*

*There are data points that significantly deviate from the main distribution. These outliers may correspond to unusual materials or measurement errors.*

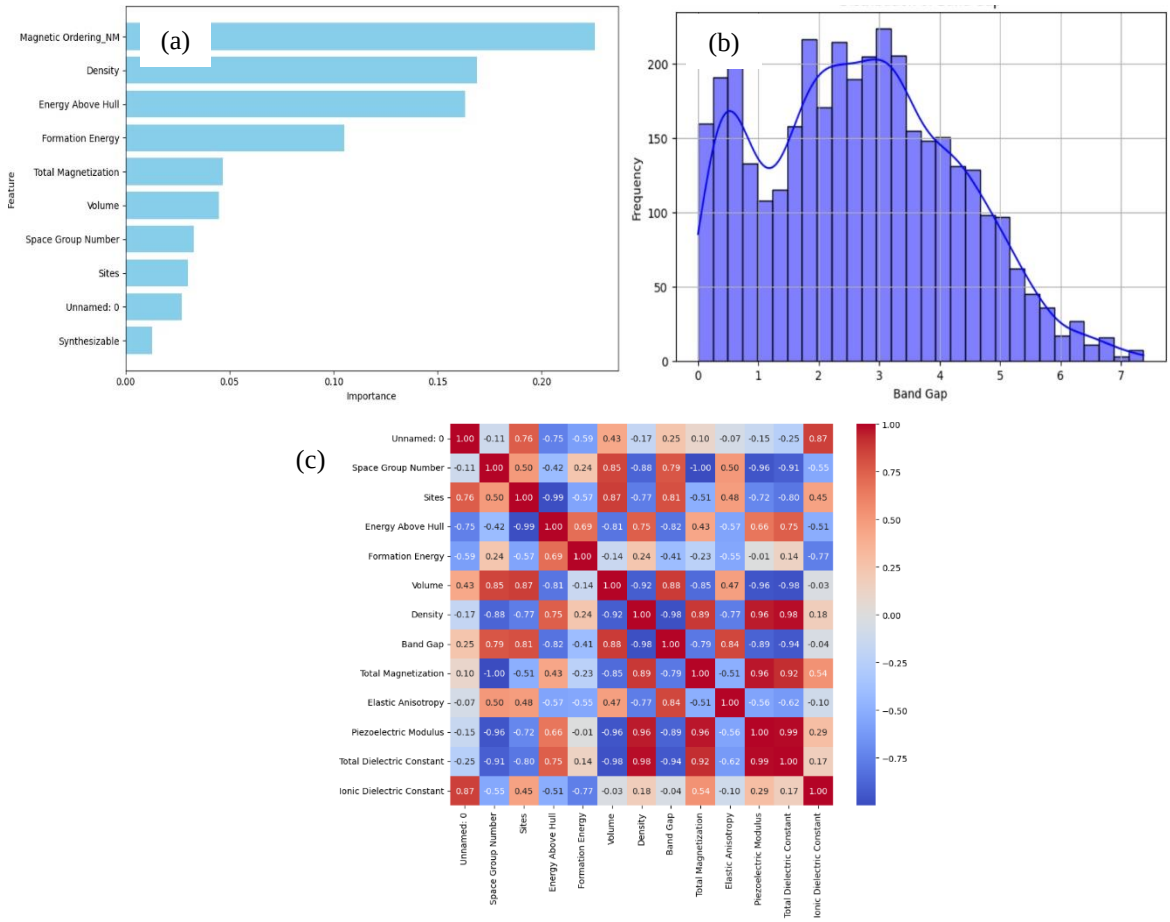


***Fig2.*** *a) Feature importance plot, b) distribution of band gap plot and c) heatmap plot*

### 6.3 Heatmap

*The correlation heatmap illustrates the significant relationships among various properties and the band gap, calculated using the Pearson Correlation Coefficient [55]. The correlation coefficient r for two random variables X and Y is defined as:*

$$r = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E(X)^2} \cdot \sqrt{E(X^2) - E(X)^2}} \qquad (5)$$

*where E is the expected value operator. Some properties, such as Energy Above Hull and Density, exhibit a negative correlation with the band gap, meaning that an increase in these properties leads to a decrease in the band gap (Figure 2c). In contrast, properties like Elastic Anisotropy show a positive correlation with the band gap, indicating that a reduction in these properties is associated with a decrease in the band gap. Properties such as the Ionic Dielectric Constant exhibit a weak correlation with the band gap, suggesting they do not play a direct role in determining its value.*

*The correlation between Density and the Band Gap, at a value of -0.98, represents a very strong and negative relationship between these two variables. In other words, this implies that as the density of the material increases, the band gap significantly decreases. Additionally, Density shows a strong positive*

*correlation with the Total Dielectric Constant (+0.98), underscoring the critical role of density in determining the electronic and structural properties of materials. On the other hand, Volume has a positive correlation with the band gap (+0.88), indicating that an increase in the material's volume generally results in an increase in the band gap.*

*Certain properties, such as the Ionic Dielectric Constant, have a limited impact on the band gap (correlation coefficient of -0.04) and may be less significant in subsequent modeling stages.*

*Overall, this analysis sheds light on the physical and chemical relationships between material structure and the band gap, highlighting that properties such as Density and Energy Above Hull have a substantial impact on the band gap. This information can be particularly useful in designing new materials, especially photocatalytic materials.*

## 7. *PERFORMANCE OF THE MODEL IN PREDICTING BAND GAP*

*The Random Forest model achieved a coefficient of determination, explaining over 81% of the variance in the band gap data, which indicates a very good performance.*
*The MAE = 0.450, MSE = 0.458, and RMSE = 0.677 demonstrate the model's acceptable accuracy in predicting band gap values.*

*Analysis of the residuals shows that the model's errors are generally normally distributed, with a mean close to zero, indicating no systematic bias in the model. However, deviations are observed in very small (band gap < 1 eV) and very large (bsnd gsp > 4 eV) band gap values, highlighting the challenges the model faces in predicting these ranges.*

*Adding more data for materials with unusual band gap values can improve the balance of the dataset and enhance the model's performance. Additionally, a detailed examination of materials with outliers in band gap values could help eliminate unreliable data or identify new and unique compounds.*

## 8. *CONCLUSION*

*This study used a Random Forest Regressor to predict band gaps of carbon and nitrogen-based photocatalysts with high accuracy ($R^2$ = 0.813, MAE = 0.450 eV). Magnetic ordering (22%) and density (16.9%) were the most significant features. The actual vs. predicted plot showed most points closely aligning with the ideal line, though deviations were noted for extreme band gap values. Residual distribution plots revealed balanced errors with a slight skew toward underestimations in higher ranges. Feature importance analysis highlighted magnetic ordering and density as key factors, supported by the strong negative correlation (-0.98) between density and band gap in the heatmap. The band gap distribution plot showed most materials falling between 1 and 3 eV, suitable for photocatalytic applications, while outliers indicated dataset imbalances. The study confirms the high effectiveness of the Random Forest Regressor in accurately predicting band gaps of carbon and nitrogen-based photocatalysts, showcasing its potential to revolutionize material design and accelerate the development of efficient, sustainable energy solutions.*

***Data and code availability***

*The predicted materials data from the Materials Project database, along with the codes developed in this study, can be accessed at https://github.com/PouyaPishkar/BandGap.git*

***Conflict of interest***

*The authors declares that they have no conflicts of interest.*

**REFERENCES**

[1] Wu C, Xue S, Qin Z, Nazari M, Yang G, Yue S, et al. Making g-C3N4 ultra-thin nanosheets active for photocatalytic overall water splitting. Appl Catal B Environ 2021;282:119557. https://doi.org/10.1016/j.apcatb.2020.119557.

[2] Tjandra AD, Huang J. Photocatalytic carbon dioxide reduction by photocatalyst innovation. Chinese Chem Lett 2018;29:734–46. https://doi.org/10.1016/j.cclet.2018.03.017.

[3] Rinke P, Delaney K, García-González P, Godby RW. Photocatalytic reactors for Environmental Remediation: A review. Phys Rev A - At Mol Opt Phys 2004;70:228–36. https://doi.org/10.1103/PhysRevA.70.063201.

[4] Low J, Yu J, Jaroniec M, Wageh S, Al-Ghamdi AA. Heterojunction Photocatalysts. Adv Mater 2017;29:1–20. https://doi.org/10.1002/adma.201601694.

[5] Chen S, Takata T, Domen K. Particulate photocatalysts for overall water splitting. Nat Rev Mater 2017;2:1–17. https://doi.org/10.1038/natrevmats.2017.50.

[6] Yuan J, Li H, Wang G, Zhang C, Wang Y, Yang L, et al. Adsorption, isolated electron/hole transport, and confined catalysis coupling to enhance the photocatalytic degradation performance. Appl Catal B Environ 2022;303:120892. https://doi.org/10.1016/j.apcatb.2021.120892.

[7] Jing L, Zhou W, Tian G, Fu H. Surface tuning for oxide-based nanomaterials as efficient photocatalysts. Chem Soc Rev 2013;42:9509–49. https://doi.org/10.1039/c3cs60176e.

[8] Wang W, Tadé MO, Shao Z. Nitrogen-doped simple and complex oxides for photocatalysis: A review. Prog Mater Sci 2018;92:33–63. https://doi.org/10.1016/j.pmatsci.2017.09.002.

[9] Cao S, Yu J. Carbon-based H2-production photocatalytic materials. J Photochem Photobiol C Photochem Rev 2016;27:72–99. https://doi.org/10.1016/j.jphotochemrev.2016.04.002.

[10] Tran PD, Wong LH, Barber J, Loo JSC. Recent advances in hybrid photocatalysts for solar fuel production. Energy Environ Sci 2012;5:5902–18. https://doi.org/10.1039/c2ee02849b.

[11] Redinger A, Siebentritt S. Optical Properties and Electronic Structure of Amorphous Germanium. Copp Zinc Tin Sulfide-Based Thin-Film Sol Cells 2015;627:363–86. https://doi.org/10.1002/9781118437865.ch16.

[12] Yin M, Tan X, Wang K, Li H, Fan D, Wang Z, et al. Potential application in the photocatalysis field of a carbon nitrogen material and methods to improve photocatalytic efficiency. Diam Relat Mater 2024;144:110963. https://doi.org/10.1016/j.diamond.2024.110963.

[13] Wang CY, Maeda K, Chang LL, Tung KL, Hu C. Synthesis and applications of carbon nitride (CNx) family with different carbon to nitrogen ratio. Carbon N Y 2022;188:482–91. https://doi.org/10.1016/j.carbon.2021.12.027.

[14] Sakaushi K, Antonietti M. Carbon- and nitrogen-based porous solids: A recently emerging class of materials. Bull Chem Soc Jpn 2015;88:386–98. https://doi.org/10.1246/bcsj.20140317.
[15] Cheng M, Xiao C, Xie Y. Photocatalytic nitrogen fixation: The role of defects in photocatalysts. J Mater Chem A 2019;7:19616–33. https://doi.org/10.1039/c9ta06435d.
[16] Rajan AC, Mishra A, Satsangi S, Vaish R, Mizuseki H, Lee KR, et al. Machine-learning-assisted accurate band gap predictions of functionalized mxene. Chem Mater 2018;30:4031–8. https://doi.org/10.1021/acs.chemmater.8b00686.
[17] Jung JY, Park JH, Jeong YJ, Yang KH, Choi NK, Kim SH, et al. Self-Consistent Equations Including Exchange and Correlation Effects. Korean J Physiol Pharmacol 2006;10:289–95.
[18] Aryasetiawan F, Gunnarsson O. The GW method. Reports Prog Phys 1998;61:237–312. https://doi.org/10.1088/0034-4885/61/3/002.
[19] AlQuraishi M. Machine learning in protein structure prediction. Curr Opin Chem Biol 2021;65:1–8. https://doi.org/10.1016/j.cbpa.2021.04.005.
[20] Ben Chaabene W, Flah M, Nehdi ML. Machine learning prediction of mechanical properties of concrete: Critical review. Constr Build Mater 2020;260:119889. https://doi.org/10.1016/j.conbuildmat.2020.119889.
[21] Seko A, Hayashi H, Nakayama K, Takahashi A, Tanaka I. Representation of compounds for machine-learning prediction of physical properties. Phys Rev B 2017;95. https://doi.org/10.1103/PhysRevB.95.144110.
[22] Ward L, Agrawal A, Choudhary A, Wolverton C. A general-purpose machine learning framework for predicting properties of inorganic materials. Npj Comput Mater 2016;2:1–7. https://doi.org/10.1038/npjcompumats.2016.28.
[23] Heid E, Greenman KP, Chung Y, Li SC, Graff DE, Vermeire FH, et al. Chemprop: A Machine Learning Package for Chemical Property Prediction. J Chem Inf Model 2024;64:9–17. https://doi.org/10.1021/acs.jcim.3c01250.
[24] Katsikas G, Sarafidis C, Kioseoglou J. Machine Learning in Magnetic Materials. Phys Status Solidi Basic Res 2021;258. https://doi.org/10.1002/pssb.202000600.
[25] Jung SG, Jung G, Cole JM. Machine-Learning Prediction of Curie Temperature from Chemical Compositions of Ferromagnetic Materials. J Chem Inf Model 2024;64:6388–409. https://doi.org/10.1021/acs.jcim.4c00947.
[26] Fung V, Ganesh P, Sumpter BG. Physically Informed Machine Learning Prediction of Electronic Density of States. Chem Mater 2022;34:4848–55. https://doi.org/10.1021/acs.chemmater.1c04252.
[27] Olsthoorn B, Geilhufe RM, Borysov SS, Balatsky A V. Band Gap Prediction for Large Organic Crystal Structures with Machine Learning. Adv Quantum Technol 2019;2:1–12. https://doi.org/10.1002/qute.201900023.
[28] Cavasotto CN, Scardino V. Machine Learning Toxicity Prediction: Latest Advances by Toxicity End Point. ACS Omega 2022;7:47536–46. https://doi.org/10.1021/acsomega.2c05693.
[29] Xia Y, Wang G, Lv Y, Shao C, Yang Z. Prediction of light absorption properties of organic dyes using machine learning technology. Chem Phys Lett 2024;836:141030. https://doi.org/10.1016/j.cplett.2023.141030.
[30] Liu X, Wang Y, Yuan J, Li X, Wu S, Bao Y, et al. Prediction of the Ibuprofen

Loading Capacity of MOFs by Machine Learning. Bioengineering 2022;9. https://doi.org/10.3390/bioengineering9100517.
[31] Wei J, Chu X, Sun XY, Xu K, Deng HX, Chen J, et al. Machine learning in materials science. InfoMat 2019;1:338–58. https://doi.org/10.1002/inf2.12028.
[32] Zhang Y, Xu W, Liu G, Zhang Z, Zhu J, Li M. Bandgap prediction of two-dimensional materials using machine learning. PLoS One 2021;16:1–12. https://doi.org/10.1371/journal.pone.0255637.
[33] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30:79–82. https://doi.org/10.3354/cr030079.
[34] Bickel, Peter J.; Doksum KA. If we use quadratic loss, our risk function is called the mean squared error (MSE). Math. Stat. Basic Ideas Sel. Top. Vol. I (Second ed.)., 2015, p. 20.
[35] Gareth, James; Witten, Daniela; Hastie, Trevor; Tibshirani R. An Introduction to Statistical Learning: with Applications in R. Springer; 2021.
[36] Hyndman RJ, Koehler AB. and Business Statistics Another Look at Measures of Forecast Accuracy Another look at measures of forecast accuracy. Int J Forecast 2005;22:679–688.
[37] Glantz, Stanton A.; Slinker BK. Primer of Applied Regression and Analysis of Variance. McGraw-Hill; 1990.
[38] G SP, Mattur MN, Nagappan N, Rath S, Thomas T. Prediction of nature of band gap of perovskite oxides (ABO3) using a machine learning approach. J Mater 2022;8:937–48. https://doi.org/10.1016/j.jmat.2022.04.006.
[39] Breiman L. Random forests Mach Learn 45 (1): 5–32. 2001.
[40] Guo Z, Lin B. Machine learning stability and band gap of lead-free halide double perovskite materials for perovskite solar cells. Sol Energy 2021;228:689–99. https://doi.org/10.1016/j.solener.2021.09.030.
[41] McKinney W. pandas: a Foundational Python Library for Data Analysis and Statistics. Python High Perform Sci Comput 2011:1–9.
[42] Lavanya A, Gaurav L, Sindhuja S, Seam H, Joydeep M, Uppalapati V, et al. Assessing the Performance of Python Data Visualization Libraries: A Review. Int J Comput Eng Res Trends 2023;10:28–39. https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0104.
[43] Sial AH, Yahya S, Rashdi S. Comparative Analysis of Data Visualization Libraries Matplotlib and Seaborn in Python. Int J Adv Trends Comput Sci Eng 2021;10:277–81. https://doi.org/10.30534/ijatcse/2021/391012021.
[44] Sjah WS, Rahman B, Hindarto D, Wedha ABPB. Diagnostic on Car Internal Combustion Engine through Noise. SinkrOn 2023;8:1128–39. https://doi.org/10.33395/sinkron.v8i2.12392.
[45] Liang Y, Feng D, Wu Y, Tsai ST, Li G, Ray C, et al. Highly efficient solar cell polymers developed via fine-tuning of structural and electronic properties. J Am Chem Soc 2009;131:7792–9. https://doi.org/10.1021/ja901545q.
[46] Mantooth BA, Weiss PS. Fabrication, assembly, and characterization of molecular electronic components. Proc IEEE 2003;91:1785–802. https://doi.org/10.1109/JPROC.2003.818320.
[47] Wang Y, Silveri F, Bayazit MK, Ruan Q, Li Y, Xie J, et al. Bandgap Engineering of

Organic Semiconductors for Highly Efficient Photocatalytic Water Splitting. Adv Energy Mater 2018;8:1–10. https://doi.org/10.1002/aenm.201801084.

[48] Singh S, Kumar S, Dwivedi N. Band gap optimization of p-i-n layers of a-Si:H by computer aided simulation for development of efficient solar cell. Sol Energy 2012;86:1470–6. https://doi.org/10.1016/j.solener.2012.02.007.

[49] Feng S, Lin Z, Gan X, Lv R, Terrones M. Doping two-dimensional materials: Ultra-sensitive sensors, band gap tuning and ferromagnetic monolayers. Nanoscale Horizons 2017;2:72–80. https://doi.org/10.1039/c6nh00192k.

[50] Serhan M, Sprowls M, Jackemeyer D, Long M, Perez ID, Maret W, et al. Storing Energy in Plastics: A Review on Conducting Polymers & Their Role in Electrochemical Energy Storage. AIChE Annu Meet Conf Proc 2019;2019-Novem. https://doi.org/10.1039/x0xx00000x.

[51] Meng YS, Arroyo-De Dompablo ME. First principles computational materials design for energy storage materials in lithium ion batteries. Energy Environ Sci 2009;2:589–609. https://doi.org/10.1039/b901825e.

[52] Ren, F., & Zolper JC. Wide energy bandgap electronic devices. World Scientific.; 2003.

[53] Carey GH, Abdelhady AL, Ning Z, Thon SM, Bakr OM, Sargent EH. Colloidal Quantum Dot Solar Cells. Chem Rev 2015;115:12732–63. https://doi.org/10.1021/acs.chemrev.5b00063.

[54] Hildebrandt N, Spillmann CM, Algar WR, Pons T, Stewart MH, Oh E et al. Energy Transfer with Semiconductor Quantum Dot Bioconjugates: A Versatile Platform for Biosensing, Energy Harvesting, and Other Developing Applications. Chem Rev 2016;acs.chemre. https://doi.org/10.1021/acs.chemrev.6b00030.

[55] Rushdi MA, Yoshida S, Watanabe K, Ohya Y. Machine learning approaches for thermal updraft prediction in wind solar tower systems. Renew Energy 2021;177:1001–13. https://doi.org/10.1016/j.renene.2021.06.033.