Estimation of Electronic Band Gap Energy Using Machine Learning

Sagar Prakash Barad & Sajag Kumar

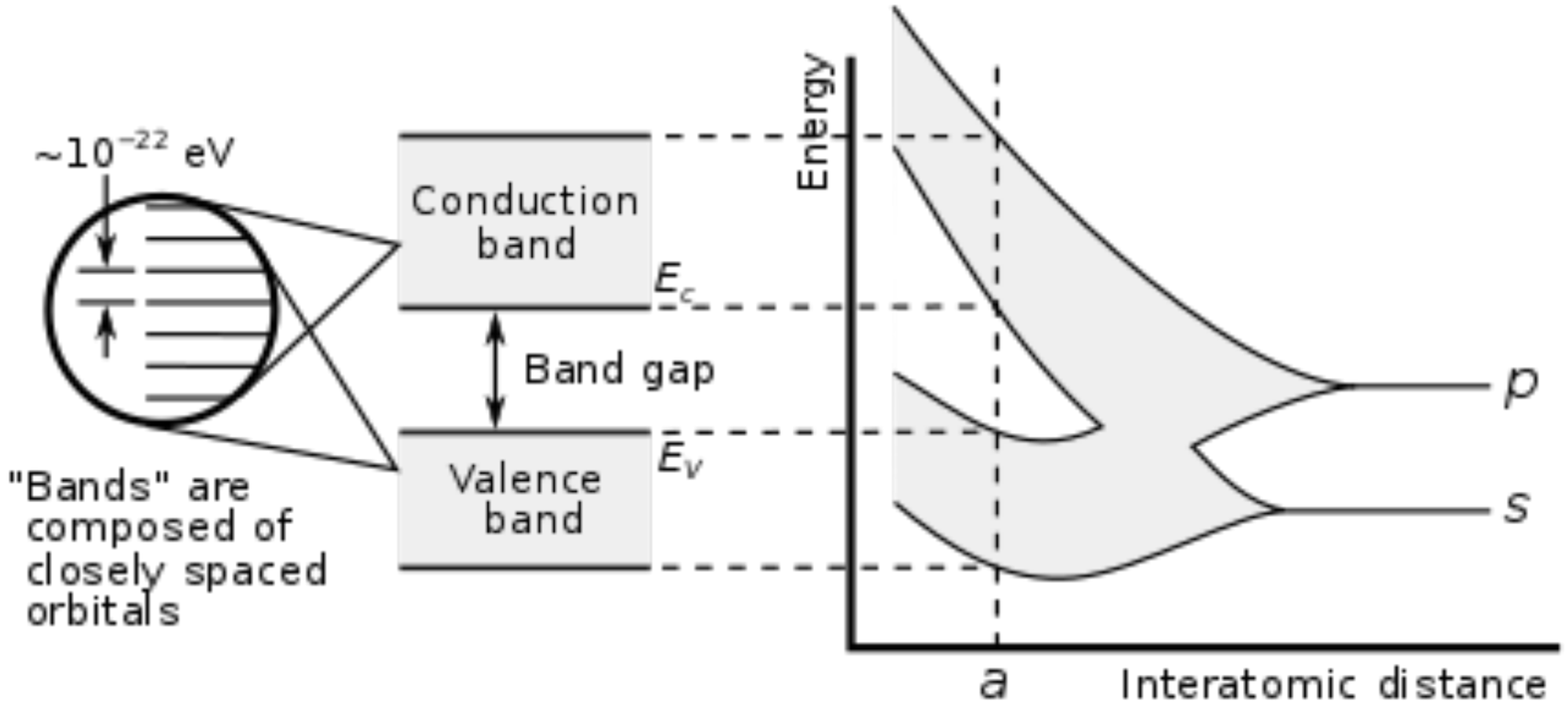


Figure credits - Band gap. (2022, November 11). In *Wikipedia*. https://en.wikipedia.org/wiki/Band_gap

Zhang Y, Xu W, Liu G, Zhang Z, Zhu J, Li M (2021) Bandgap prediction of two-dimensional materials using machine learning. PLoS ONE 16(8): e0255637. https://doi.org/10.1371/journal. pone.0255637.

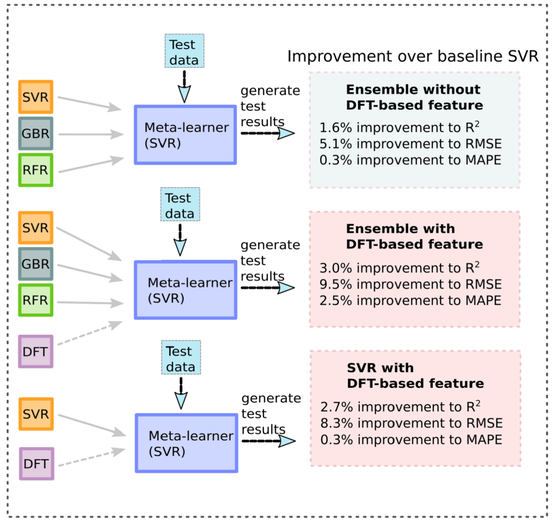
(Without 'gap_nosoc')

Models	Parameters	MAE	RMSE	R^2
SVR	C = 50, epsilon = 0.2 gamma = 50,kernel = rbf	0.10	0.41	0.75
GBDT	n_estimators = 21000, max_depth = 21, min_samples_split = 5, max_features = 0.8, learning_rate = 0.001	0.12	0.24	0.92
RF	n_estimators = 15000, max_depth = 20, min_samples_split = 5,max_features = 0.8, min_samples_leaf = 3	0.10	0.27	0.90
MLP	solver = adam, hidden_layer_sizes = (262,140,139,180), activation = tanh,alpha = 1e-8, tol = 1e-6, max_iter = 5000, learning_rate_init = 0.01	0.24	0.43	0.73

(With 'gap_nosoc')

Models	Parameters	MAE	RMSE	R^2
SVR	C = 50, epsilon = 0.2 gamma = 50,kernel = rbf	0.11	0.17	0.95
GBDT	n_estimators = 21000, max_depth = 21, min_samples_split = 5, max_features = 0.8, learning_rate = 0.001	0.03	0.09	0.98
RF	n_estimators = 15000, max_depth = 20, min_samples_split = 5,max_features = 0.8, min_samples_leaf = 3	0.03	0.10	0.98
MLP	solver = adam, hidden_layer_sizes = (262,140,139,180), activation = tanh,alpha = 1e-8, tol = 1e-6, max_iter = 5000, learning_rate_init = 0.01	0.06	0.12	0.97

Kauwe, S.K., Welker, T. & Sparks, T.D. *Extracting Knowledge from DFT: Experimental Band Gap Predictions Through Ensemble Learning*. *Integr Mater Manuf Innov* 9, 213–220 (2020).



- Composition based.
- Low accuracy on individual models.
- Heterogeneous dataset.

- Used DFT based features.
- Big dataset.
- Used neural networks.

Midway Targets

1. Come up with a consistent set of features/classification scheme for different types of materials.

Done. But in a different way.

2. Implement classic machine learning algorithms to achieve the current state-of-the-art results.

Done.

3. Perform ensemble learning to improve upon the hitherto achieved results.

Ensemble methods implemented. But worse results achieved.

Expected Results

1. Performance of ML algorithms: as good as in case of specific types of materials.

Slightly worse.

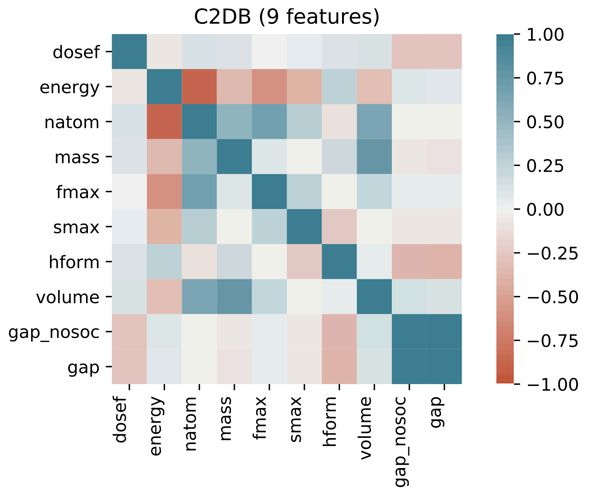
2. Ensemble learning: better results than the individual algorithms.

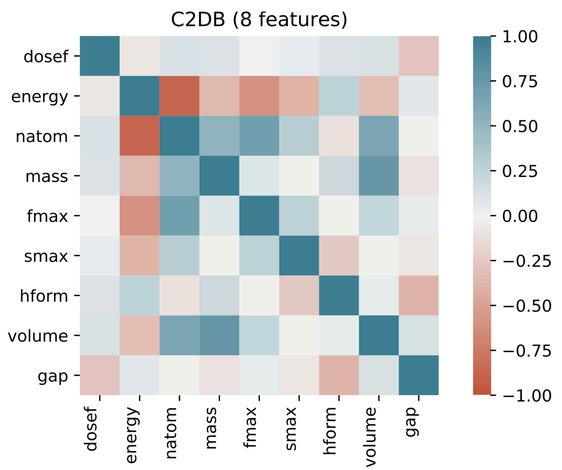
Nope.

DATASET(s)

C2DB

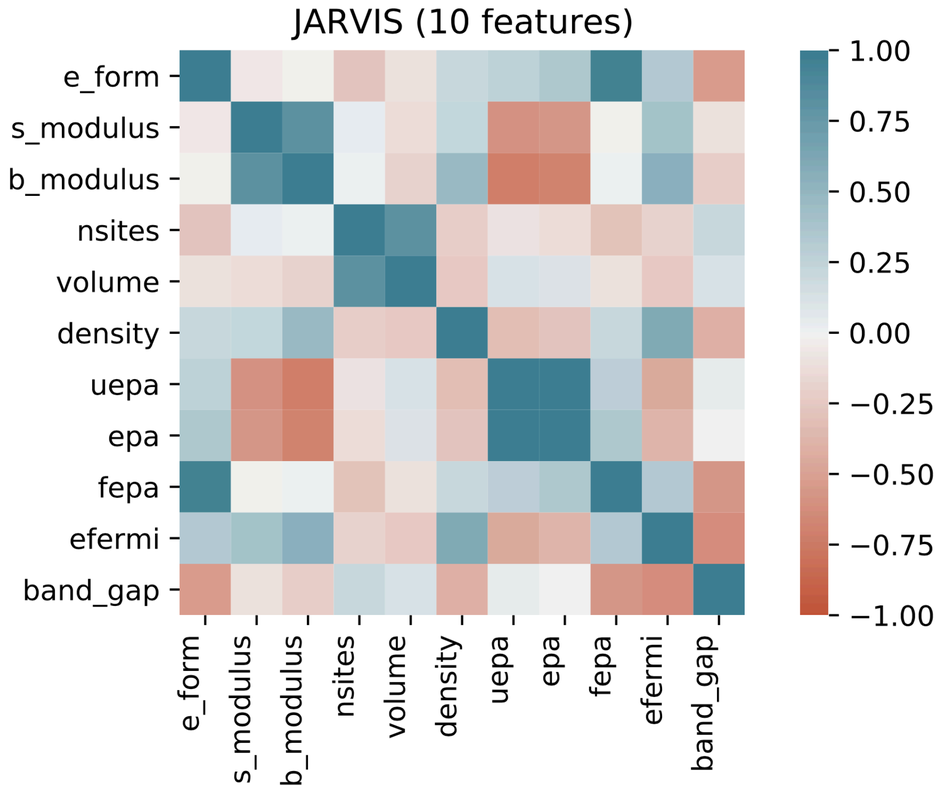
3129 rows and 8 (9) columns

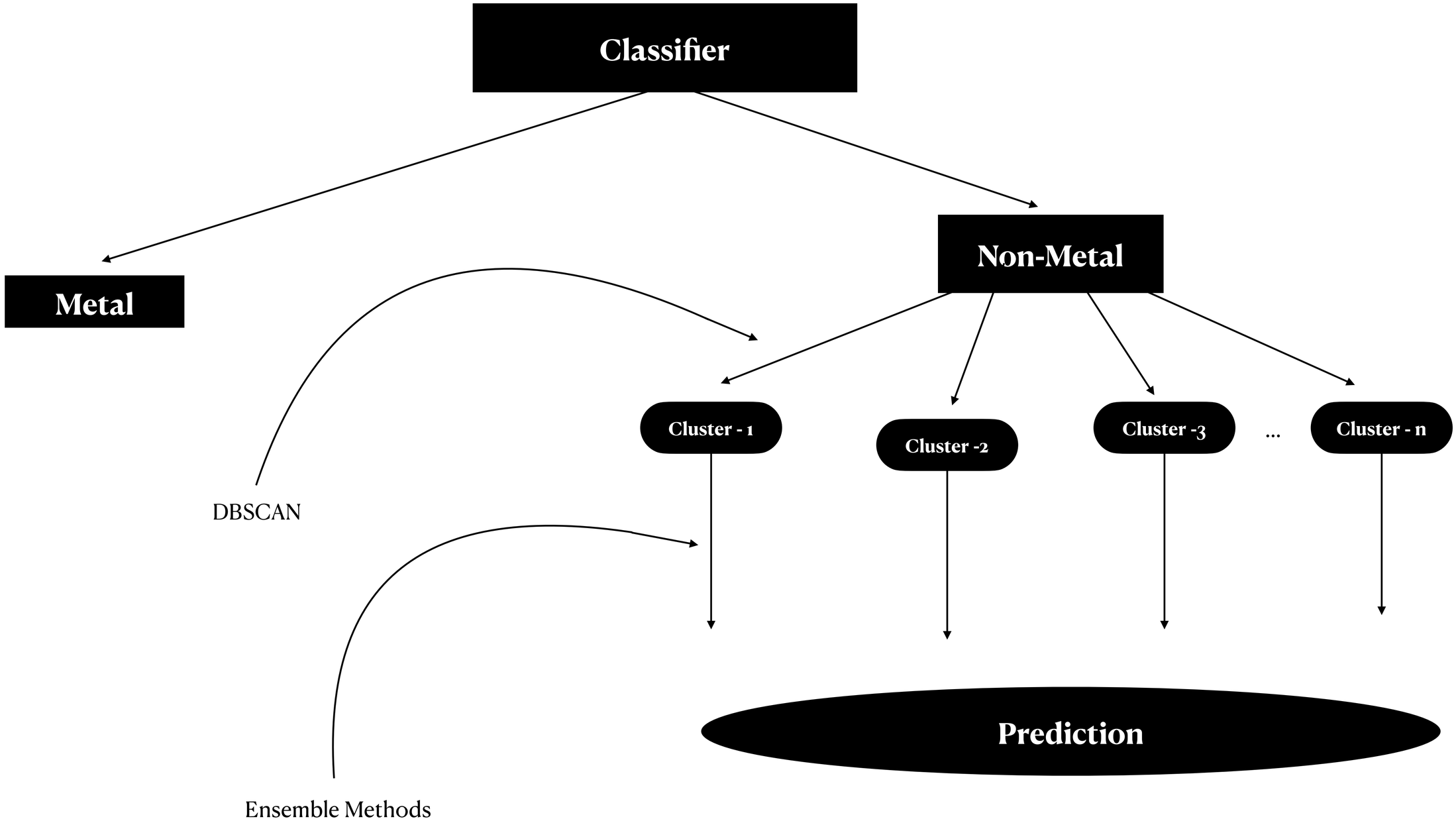




JARVIS

6590 rows and 10 columns




Classifier
Metal
Non-Metal
Cluster - 1
Cluster -2
Cluster -3
...
Cluster - n
DBSCAN
Ensemble Methods
Prediction

A New Material

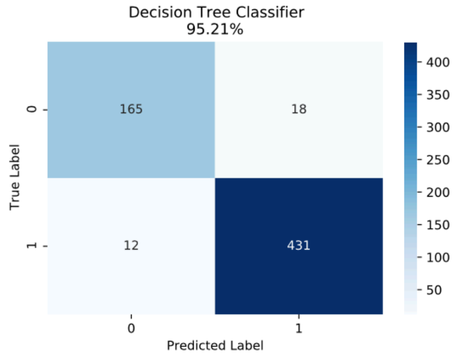
1. Classified as a metal or non-metal.

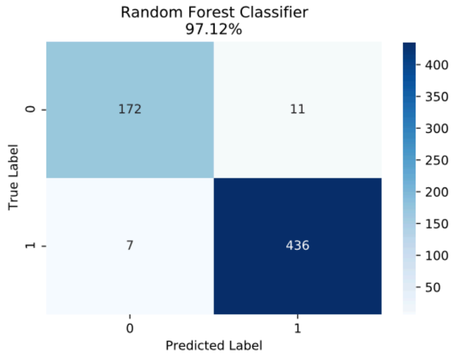
2. If classified as a non-metal, assigned to the corresponding cluster.

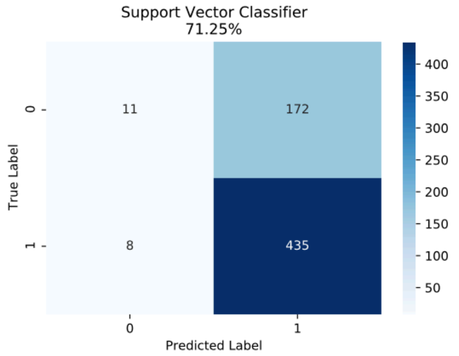
3. Fed to the model corresponding to the assigned cluster.

4. Estimated band gap.

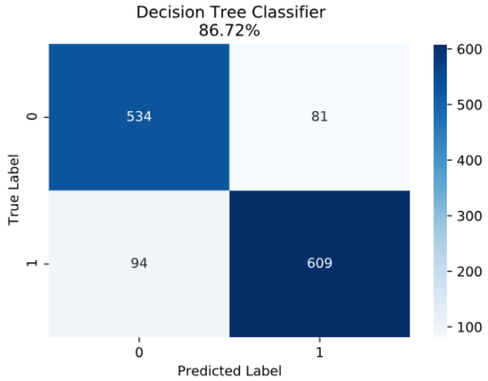
Classifier

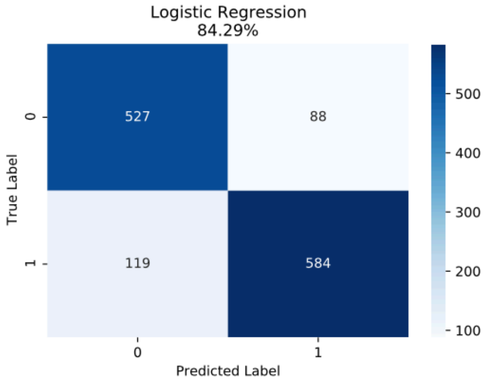


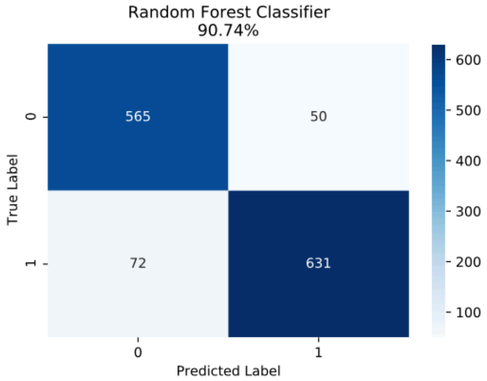


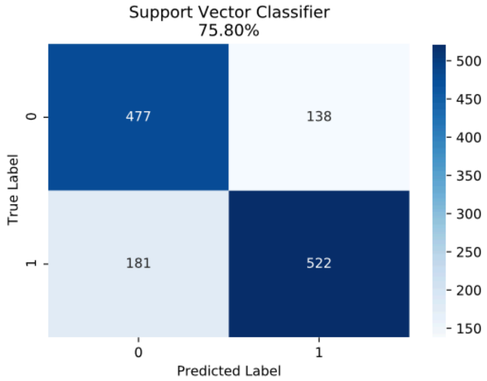


C2DB









JARVIS

Clustering

C2DB

7 Clusters

eps = *120* and *min_samples* = *1*

JARVIS

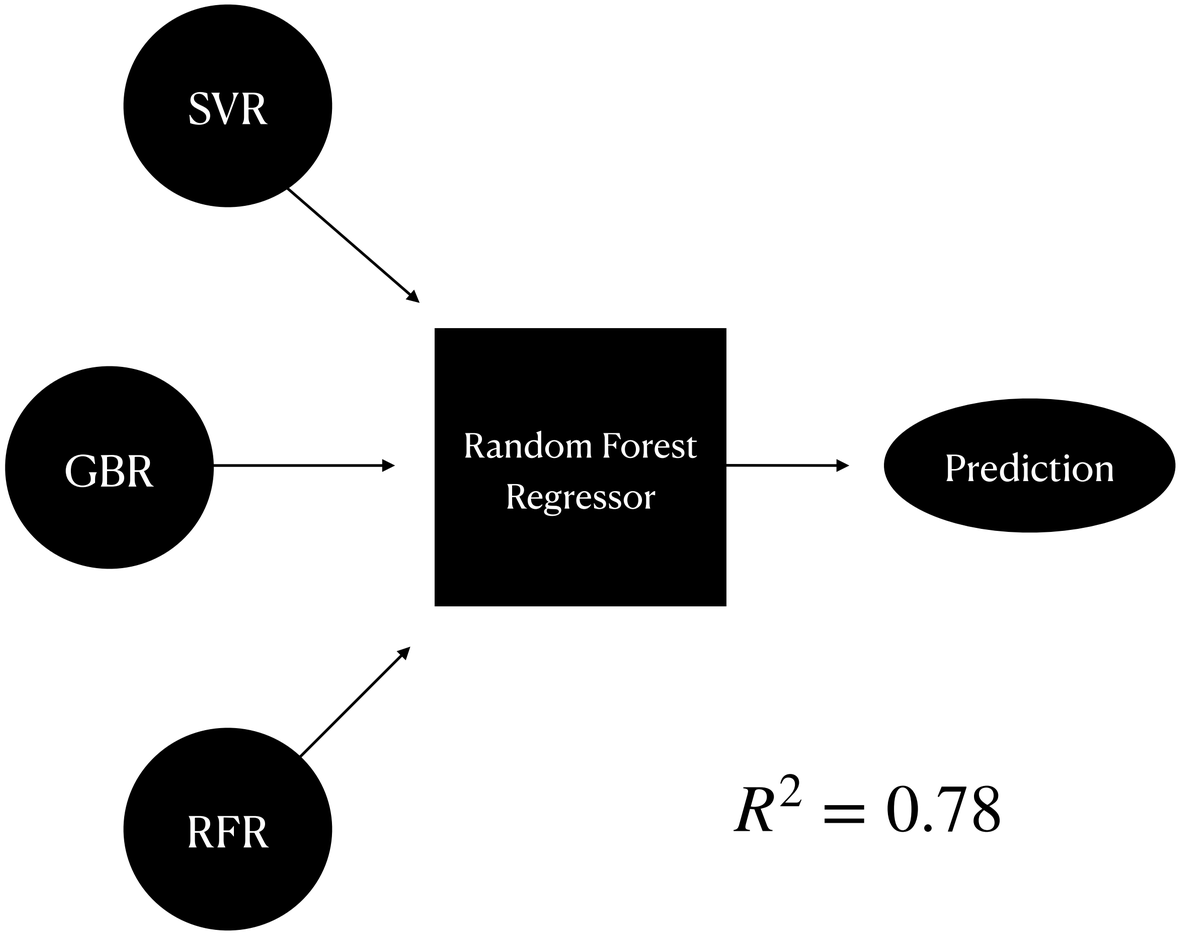
6 Clusters

eps = *100* and *min_samples* = *3*

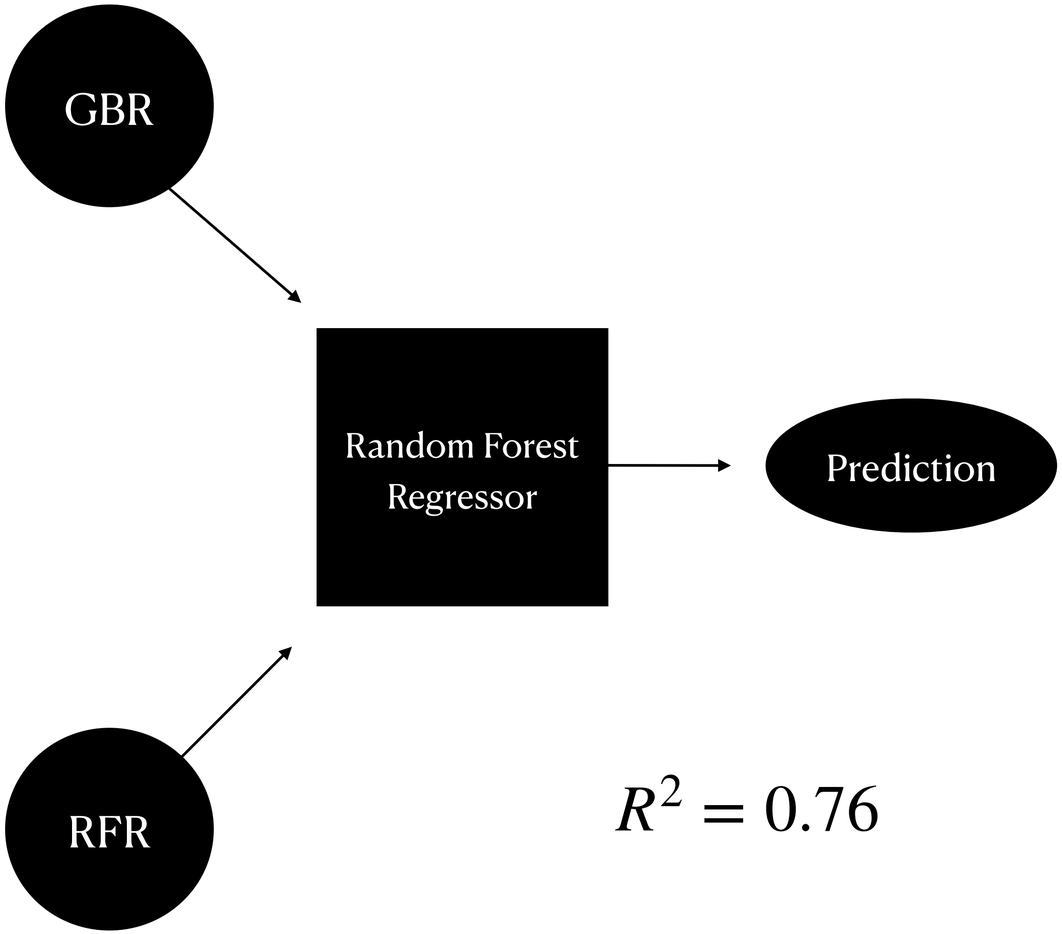
Regression

	C2DB			JARVIS		
	R^2	MAE	RMSE	R^2	MAE	RMSE
SVR	0.73	0.24	0.41	0.74	0.46	0.81
RF	0.90	0.10	0.25	0.82	0.88	1.28
GBDT	0.91	0.09	0.24	0.77	0.38	0.75

Ensembles



$R^2 = 0.78$



$R^2 = 0.76$

Future Plans

1. Improve the ensemble methods and train them on the clusters obtained through DBSCAN.
2. Further improve the metal non-metal classifier through ensemble methods.
3. Explore encoding methods to enable better learning of material properties.
4. Train deep learning networks on bigger datasets such as Materials Project and AFLOW.
5. Explore the use of graph neural networks and representation learning in learning materials properties

1. Yan K., Liu Y., Lin Y., & Ji S. (2022) *Periodic graph transformers for crystal material property prediction*. *36th Conference on Neural Information Processing Systems (NeurIPS 2022)*.

2. *Dong Y, Wu C, Zhang C, Liu Y, Cheng J, Lin J (2019) Bandgap prediction in configurationally hybridised graphene and boron nitride, npj Computational Materials (2019).*

3. Rosen I., Qu J., & Marks J. (2018) *Predicting electronic properties of materials*. *CS229 Report*.