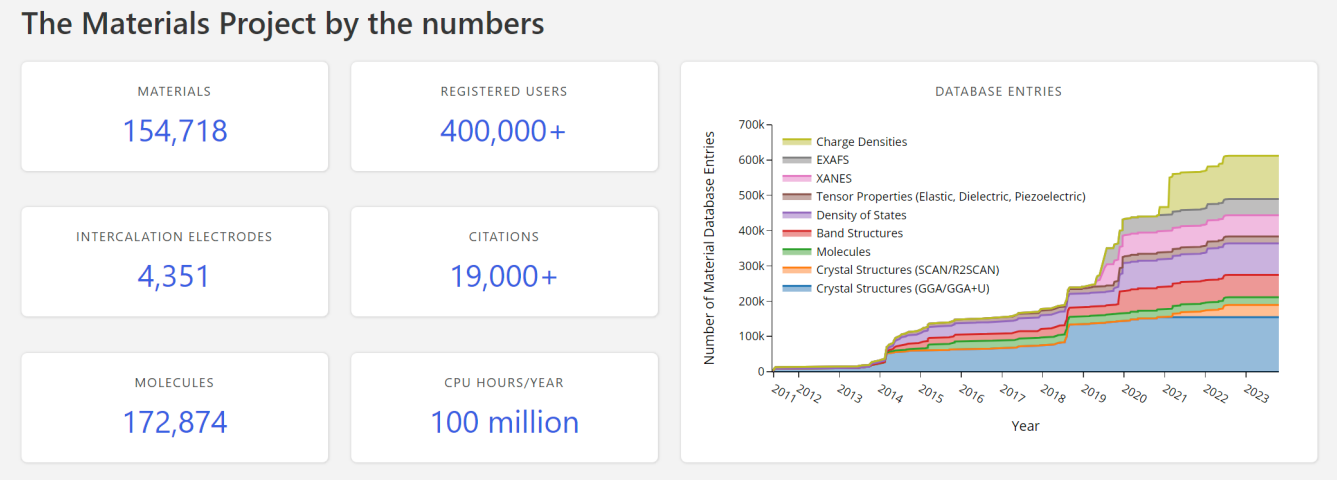
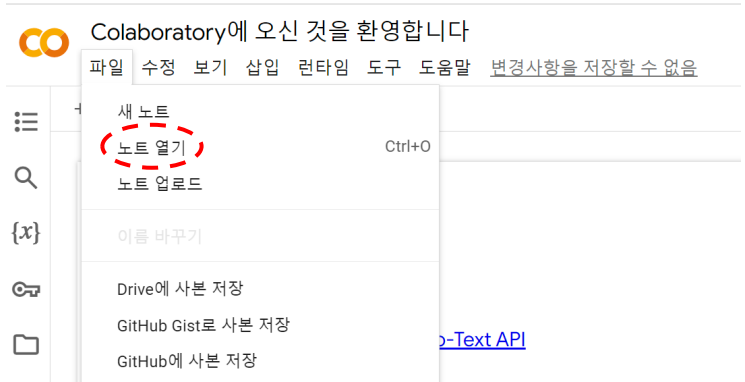
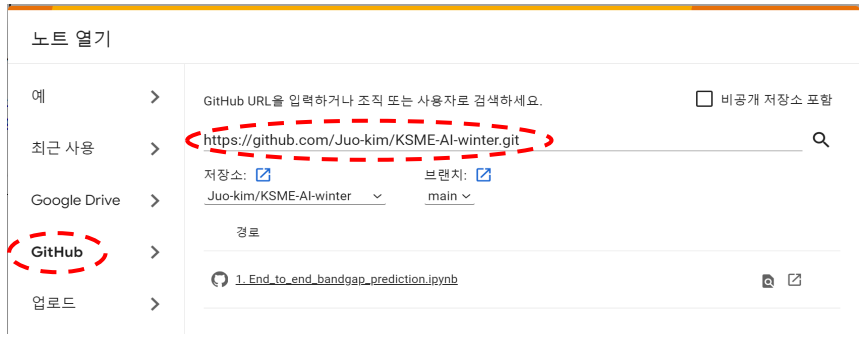
실습 자료 안내



1. Colab 접속 (https://colab.google/)
2. 파일 ➔ 노트 열기 ➔ GitHub ➔ URL 검색





- 15만여개 Inorganic 물질에 대한 DFT 계산 정보를 포함하고 있는 Materials Project 데이터 베이스 활용
- 물질의 전자 상태를 나타내는 값인 Bandgap 예측하는 모델 생성

URL : https://github.com/Juo-kim/KSME-AI-winter.git

KSME-AI-winter

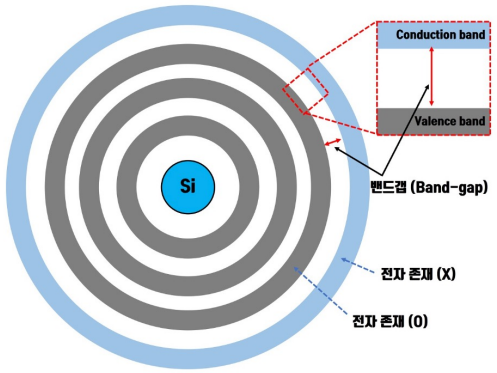
Juo Kim

CS-AI Lab
School of Mechanical Engineering
Soongsil University

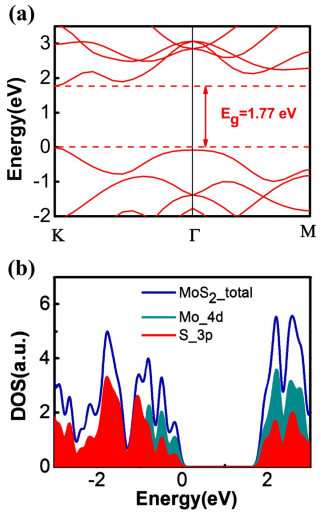
Contact : rlawndh14@gmail.com

Bandgap?

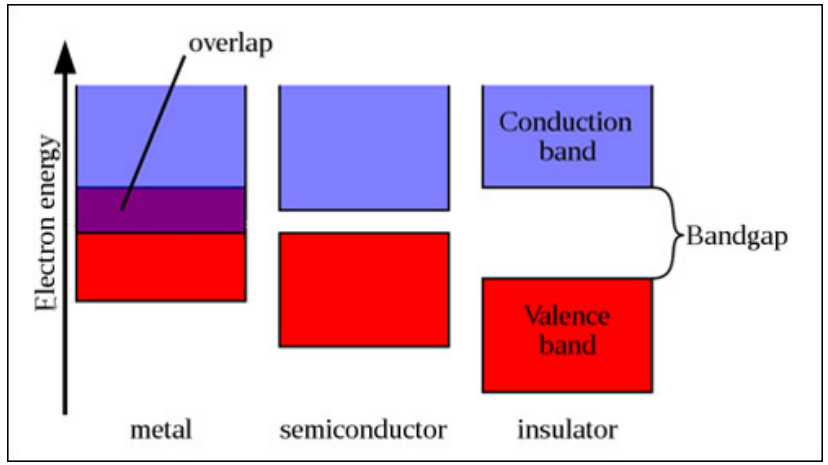
- Energy gap : 전자가 존재하고 있는 가장 높은 에너지 레벨 (Valence Band) 부터 전자가 존재하지 않는 가장 낮은 에너지레벨 (Conduction Band) 사이의 에너지 준위
- 물질이 외부에서 충분한 에너지를 공급받게 된다면, Valence band에 위치한 전자가 Conduction Band로 뛰어넘는 것이 가능함
- Bandgap의 크기에 따라 물질의 전기적 성질을 확인할 수 있음
- Metal = 0 eV, 0 eV < Semiconductor < 4 eV, Insulator ≥ 4 eV
- 전기적 성질을 알아보기 위해서 매우 중요하지만, Bandgap 계산을 위해 필요한 계산적, 실험적 cost가 큼



https://j-science.tistory.com/11



DOI: 10.1186/1556-276X-9-676

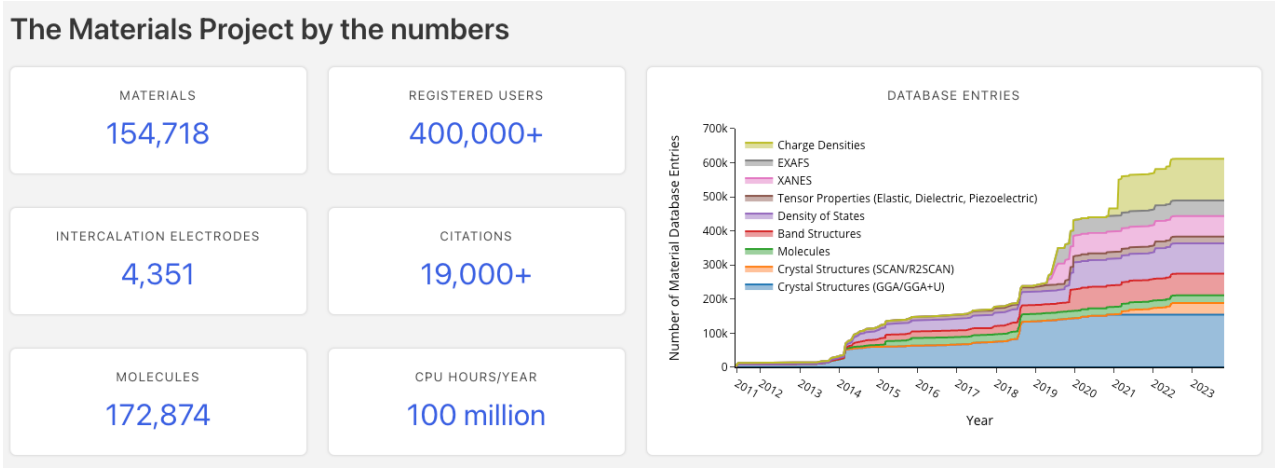


http://solarcellcentral.com/junction_page.html

Materials Project

- Open source Inorganic database 중 하나
- 154,718개의 Inorganic 물질에 대한 DFT 계산 기반의 물성 포함
- Bandgap, Energy above hull, Space group number, Crystal structure 등 많은 계산 결과

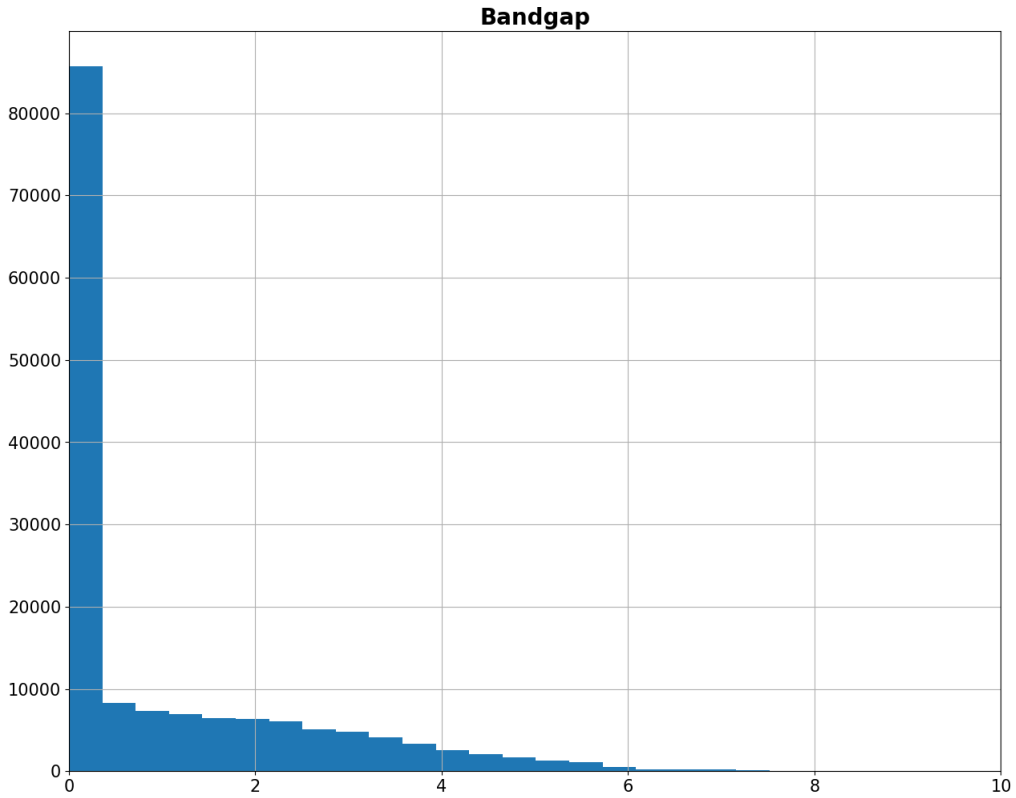
➔ Materials Project database로 bandgap을 예측할 수 있는 regression model을 만드는 것이 실습의 목표

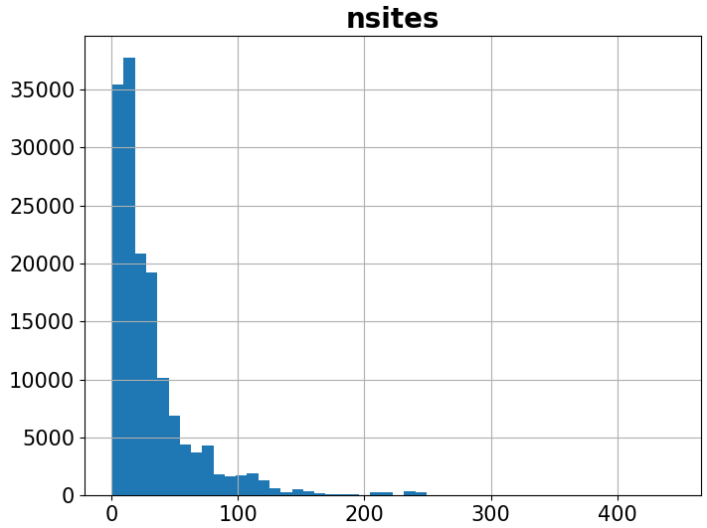


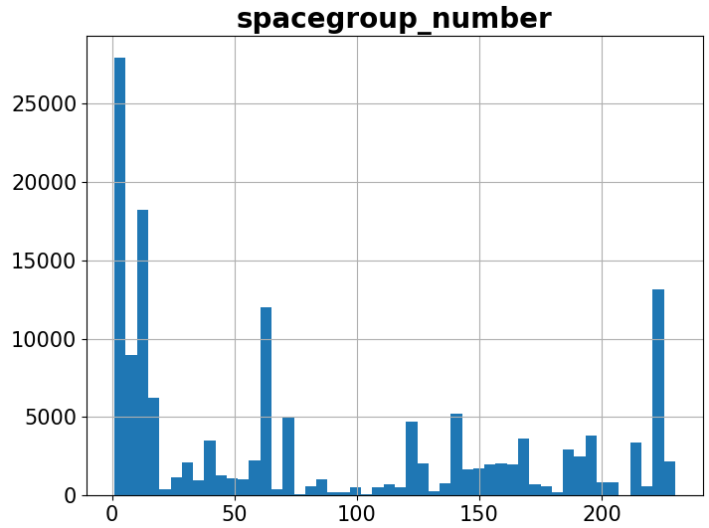
Chemical Descriptor

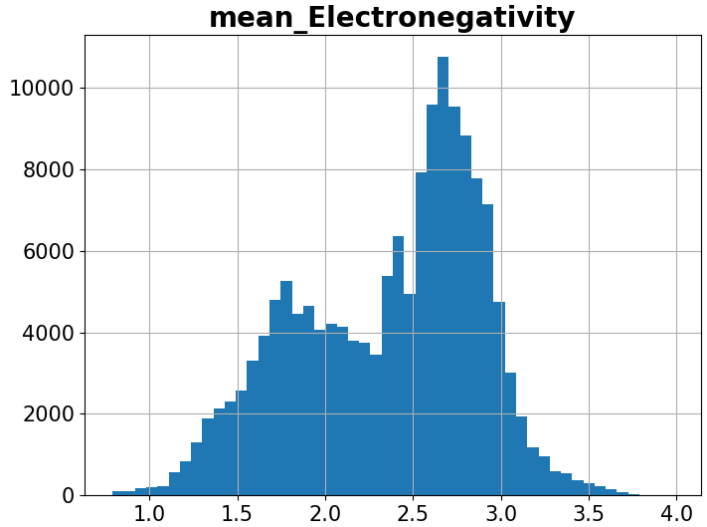
Meaning of chemical descriptors (CDs)	Number of attributes		Attributes
Stoichiometric attributes	6		Ncomp, Comp_L2Norm, Comp_L3Norm, Comp_L5Norm, Comp_L7Norm, Comp_L10Norm
Elemental-property-based attributes	132	Mean (22)	Atomic Number, MendeleevNumber, AtomicWeight, MeltingT, Column, Row, CovalentRadius, Electronegativity, NsValence, NpValence, NdValence, NfValence, NValance, NsUnfilled, NpUnfilled, NdUnfilled, NfUnfilled, NUnfilled, GSvolume_pa, GSbandgap, GSmagmom, SpaceGroupNumber
		Range (22)	
		Dev (22)	
		Max (22)	
		Min (22)	
		Most (22)	
Valance orbital occupation attributes	4		frac_sValence, frac_pValence, frac_dValence, frac_fValence
Ionic compound attributes	3		CanFormIonic, MaxIonicChar, MeanIonicChar

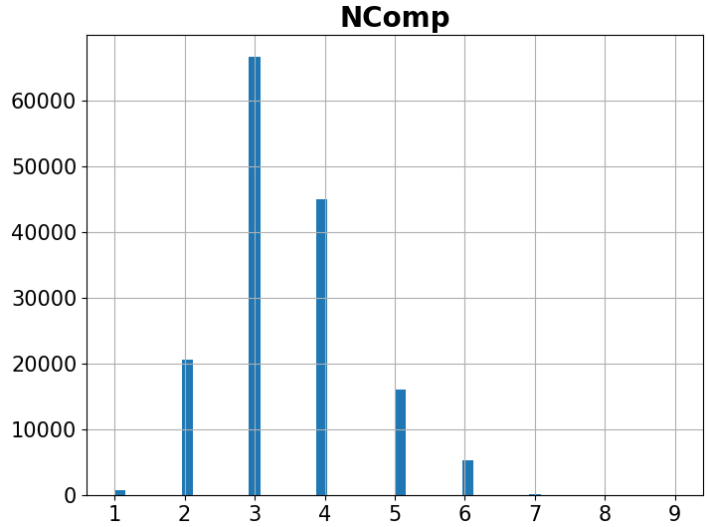
Data distribution



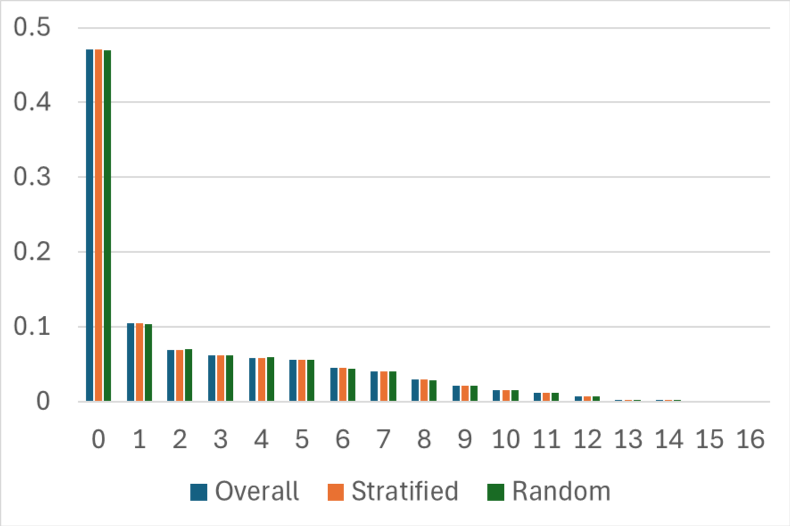


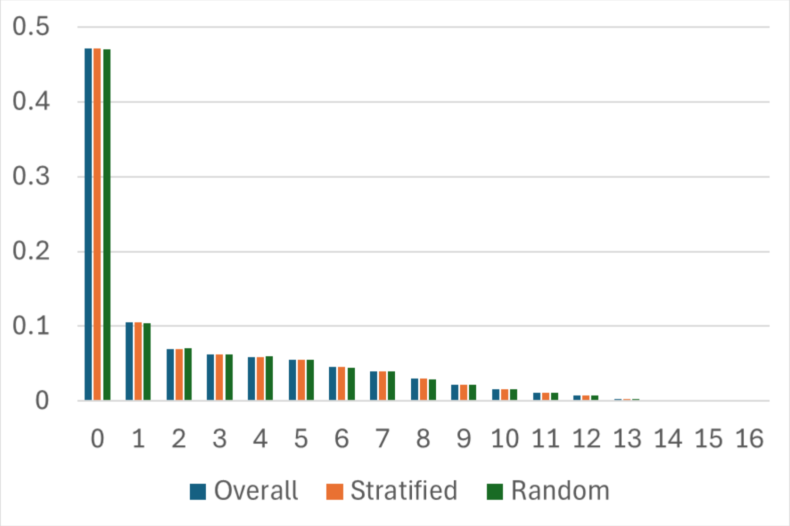


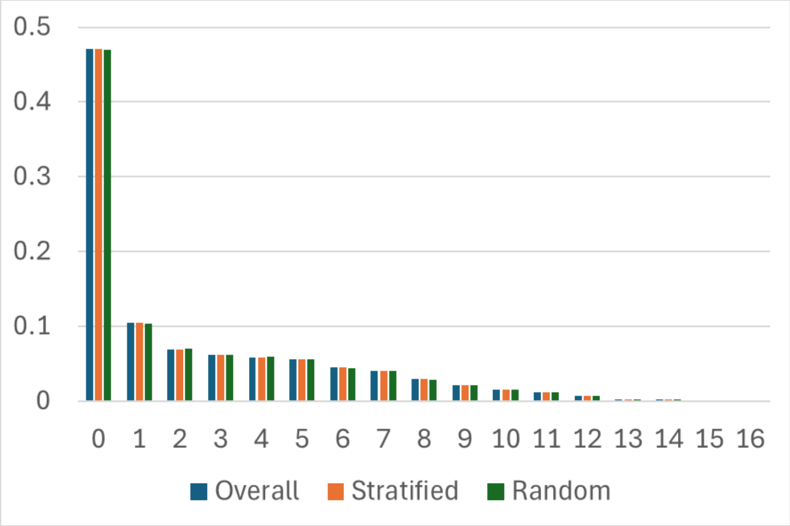


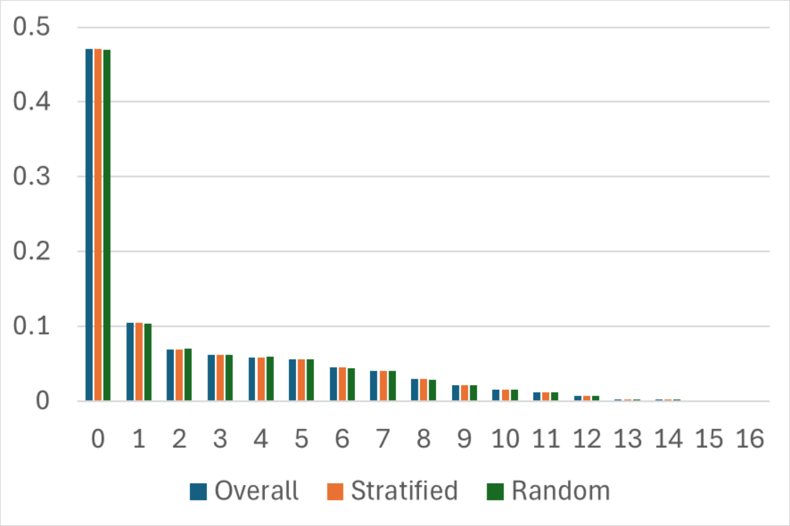


Random vs Stratified sampling









성능 평가 지표

	LR	DT	RF
R^2	0.488	0.666	0.666
MAE	0.807	0.419	0.419
RMSE	1.071	0.865	0.650

GdPO4	mp-1103324	3.2911	12	141	3	0.70710678	0.67354	0.66692688	0.66667829	0.66666679	18.5	56	15.1666667	64	8
GdPO4	mp-3735	2.613	24	14	3	0.70710678	0.67354	0.66692688	0.66667829	0.66666679	18.5	56	15.1666667	64	8

H2	mp-632291	8.8499	2	139	1	1	1	1	1	1	1	0	0	1	1
H2	mp-24504	8.0699	4	194	1	1	1	1	1	1	1	0	0	1	1
H2	mp-1066989	6.2332	4	65	1	1	1	1	1	1	1	0	0	1	1
H2	mp-730101	9.7197	8	19	1	1	1	1	1	1	1	0	0	1	1
H2	mp-1096977	0.2567	1	123	1	1	1	1	1	1	1	0	0	1	1
H2	mp-632250	7.4222	1	229	1	1	1	1	1	1	1	0	0	1	1
H2	mp-23907	7.4848	2	194	1	1	1	1	1	1	1	0	0	1	1
H2	mp-754417	0	1	191	1	1	1	1	1	1	1	0	0	1	1
H2	mp-634659	7.5517	1	225	1	1	1	1	1	1	1	0	0	1	1
H2	mp-850274	8.5338	2	139	1	1	1	1	1	1	1	0	0	1	1
H2	mp-1188177	7.3865	16	62	1	1	1	1	1	1	1	0	0	1	1
H2	mp-738409	8.1865	32	4	1	1	1	1	1	1	1	0	0	1	1
H2	mp-731827	8.1932	16	62	1	1	1	1	1	1	1	0	0	1	1
H2	mp-973783	8.8022	4	64	1	1	1	1	1	1	1	0	0	1	1
H2	mp-570752	6.6359	2	194	1	1	1	1	1	1	1	0	0	1	1
H2	mp-1181265	1.5178	16	62	1	1	1	1	1	1	1	0	0	1	1
H2	mp-632172	9.3289	2	139	1	1	1	1	1	1	1	0	0	1	1

algorithm	mean mae	std mae	mean rmse	max max_error
coGN	**0.1559**	0.0017	0.3956	7.3352
coNGN	**0.1697**	0.0035	0.4271	7.9674
ALIGNN	**0.1861**	0.0030	0.4635	7.4756
MegNet (kgcnn v2.1.0)	**0.1934**	0.0087	0.4715	7.8821
DimeNet++ (kgcnn v2.1.0)	**0.1993**	0.0058	0.4720	14.0169
Finder_v1.2 structure-based version	**0.2193**	0.0012	0.4989	7.6676
MODNet (v0.1.12)	**0.2199**	0.0059	0.4525	7.5685

https://matbench.materialsproject.org/

Thank you