Electronic Supplementary Material (ESI) for Journal of Materials Chemistry A.

Supporting Information

Machine Learning-Based High Throughput Screening for Nitrogen Fixation on Boron-Doped Single Atom Catalysts

Mohammad Zafari, Deepak Kumar, Muhammad Umer, Kwang S. Kim*

[1]Center for Superfunctional Materials, Department of Chemistry, Ulsan National Institute of Science and Technology (UNIST), 50 UNIST-gil, Ulsan 689-798, Korea

*kimks@unist.ac.kr

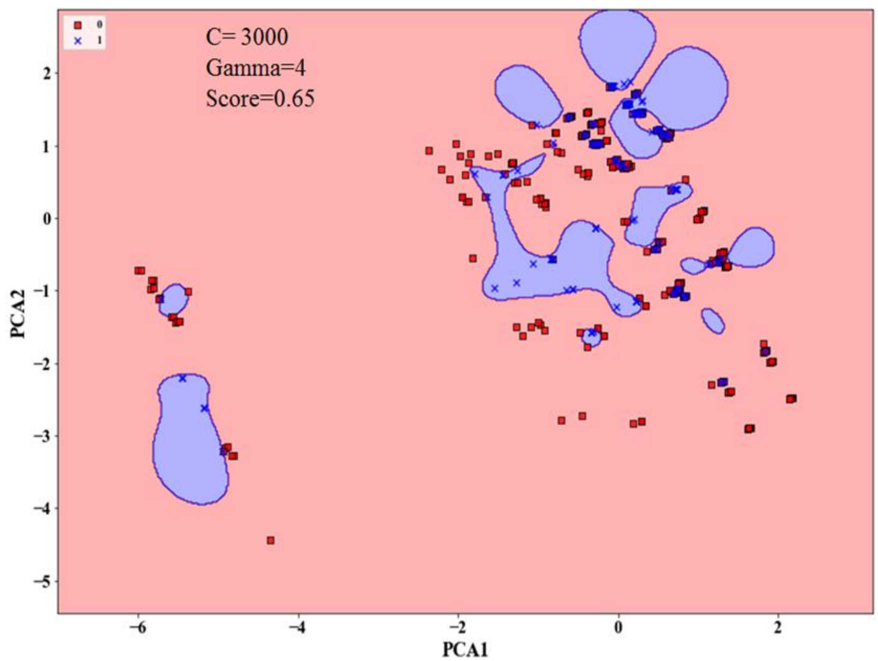

Fig. S1 Support Vector Machine (SVM) predictions based on the train samples, PCA2 versus PCA1. C (=3000), and Gamma (=4) are tuning parameters (blue: eligible, red: ineligible, red squares & blue crosses: training data).

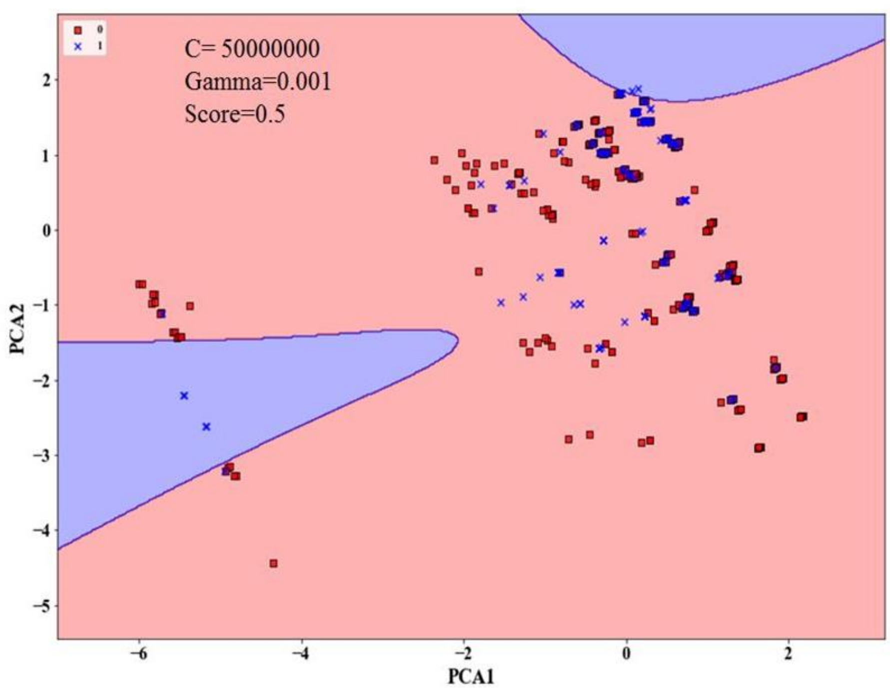

Fig. S2 Support Vector Machine (SVM) predictions based on the train samples, PCA2 versus PCA1. C (=5×10^7), and Gamma (=0.001) are tuning parameters (blue: eligible, red: ineligible, red squares & blue crosses: training data).

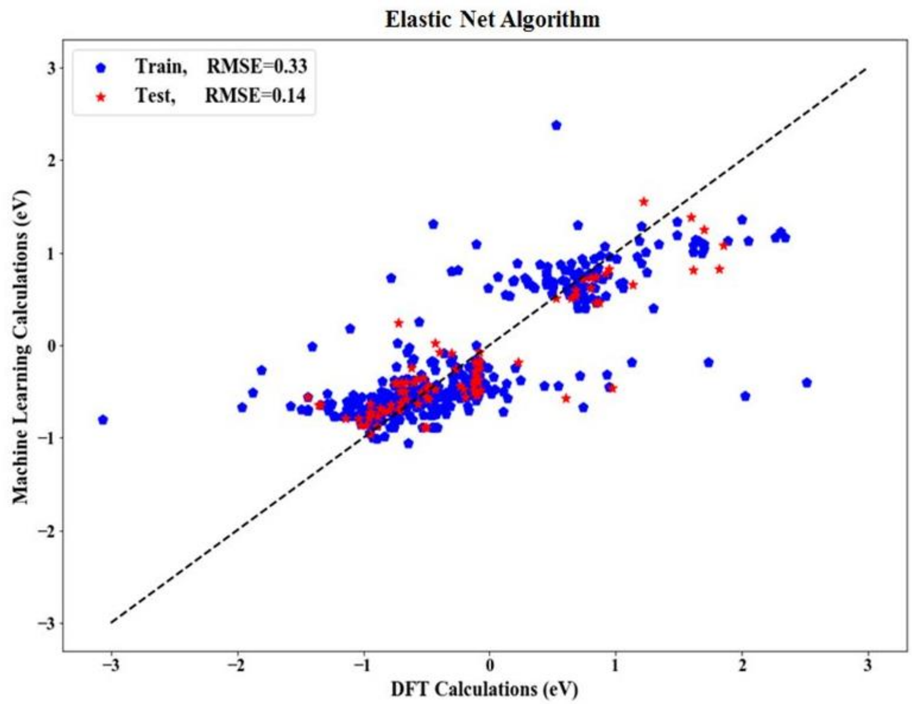

Fig. S3 Prediction performance (Elastic Net Algorithm) plot between DFT-Calculations and Machine-Learning Outputs.

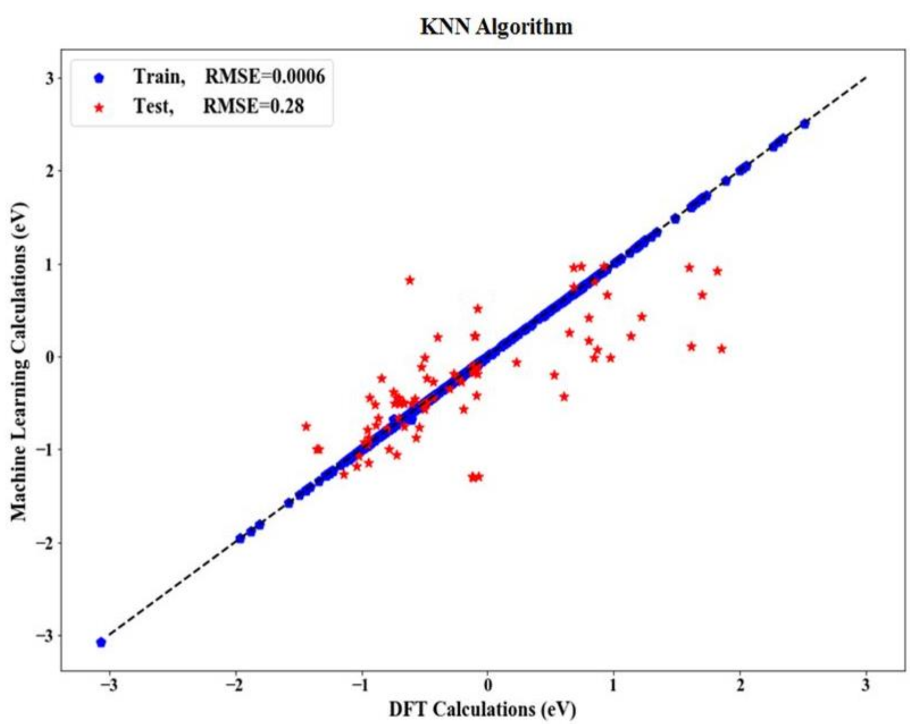

Fig. S4 Prediction performance (KNN Algorithm) plot between DFT-Calculations and Machine-Learning Outputs.

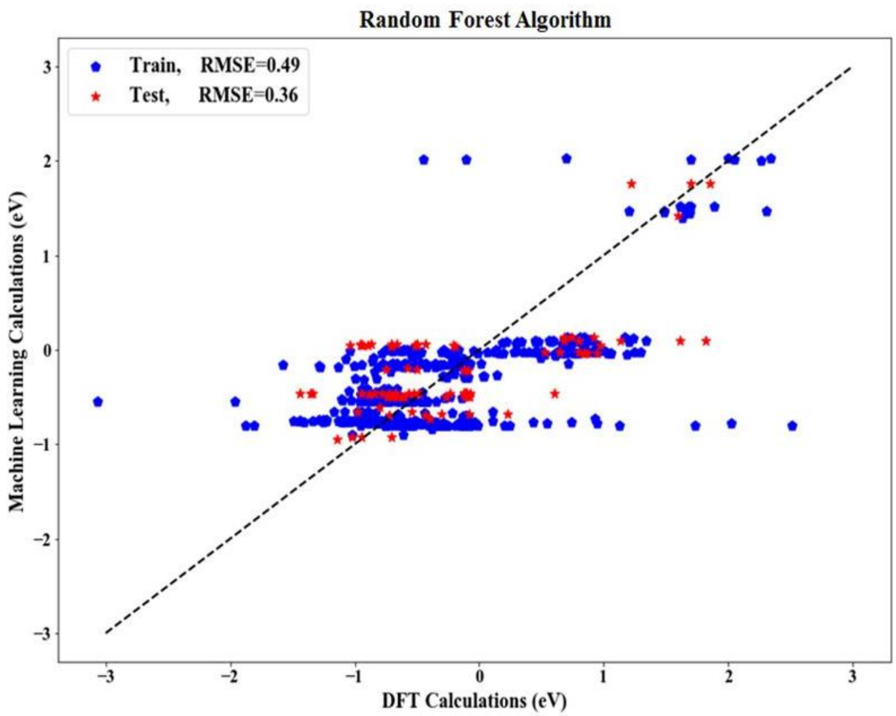

Fig. S5 Prediction performance (Random Forest Algorithm) plot between DFT-Calculations and

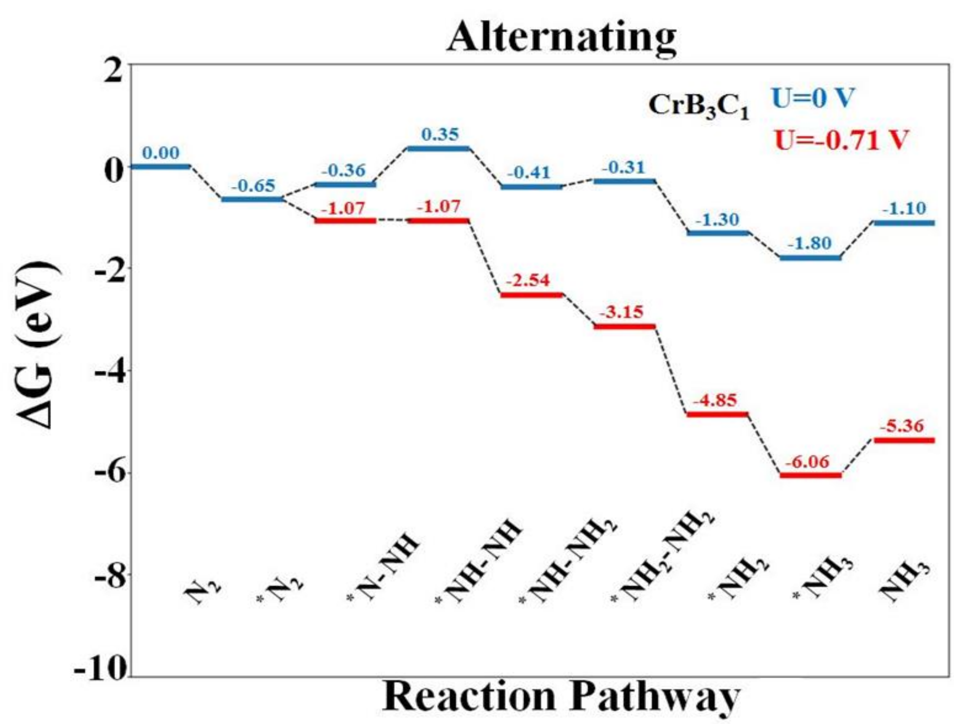

Fig. S6 Diagram of Free energy calculated (DFT) for NRR via Alternating mechanism for CrB_3C_1 at zero and applied potentials. The blue and red curves depict free energy changes for NRR at 0 and applied potentials.

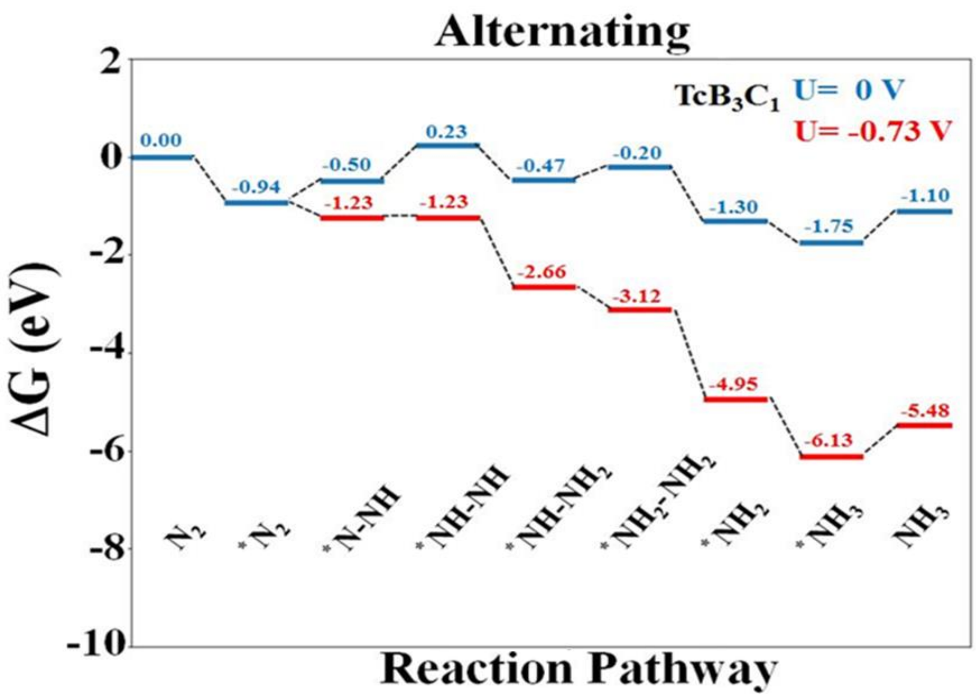

Fig. S7 Diagram of Free energy calculated (DFT) for NRR via Alternating mechanism for TcB_3C_1 at zero and applied potentials. The blue and red curves depict free energy changes for NRR at 0 and applied potentials.

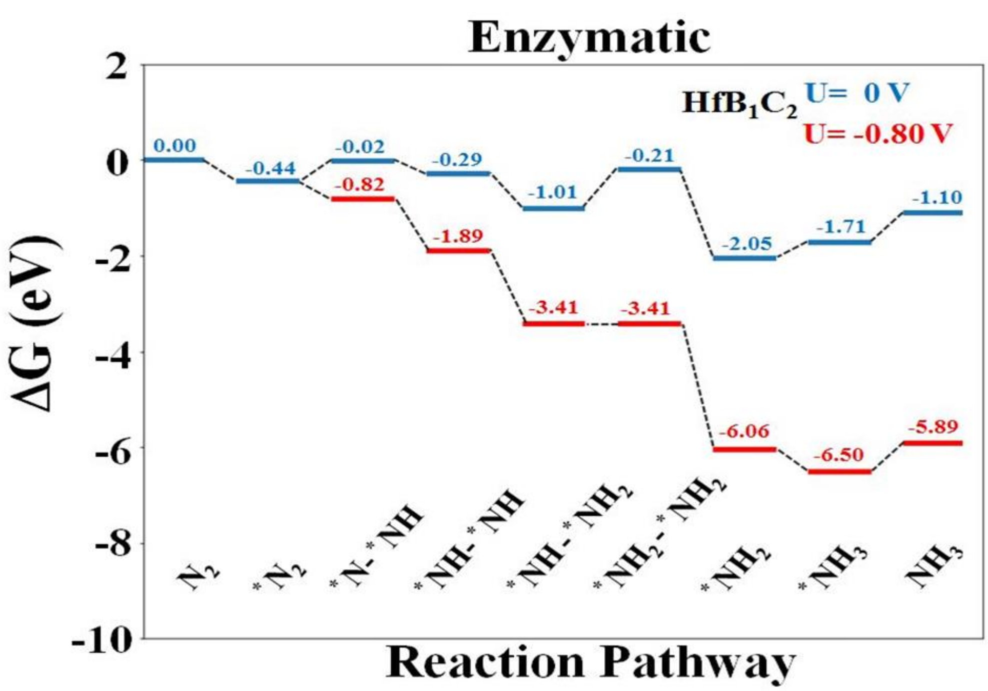

Fig. S8 Diagram of Free energy calculated (DFT) for NRR via Enzymatic mechanism for HfB_1C_2 at zero and applied potentials. The blue and red curves depict free energy changes for NRR at 0 and applied potentials.

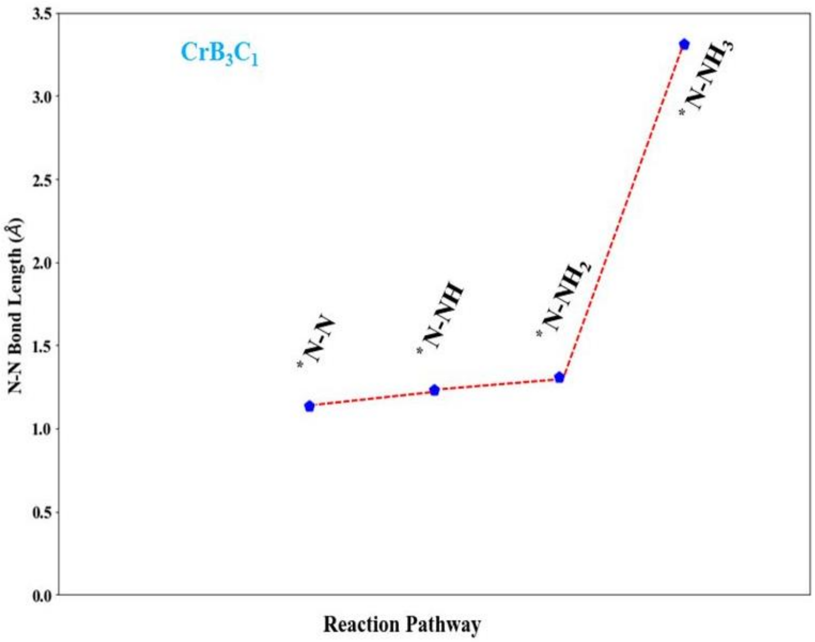

Fig. S9 N-N bond length for CrB_3C_1 at different reaction pathway via distal mechanism for end-on adsorption. The bond distance during different-steps reduction of N_2 to NH_3 indicates gradual increase in bond length before hydrogenation of NH_2 to NH_3. After that, there is a large increment in bond length up to 3.31 Å. Therefore, the stretching effect in bond length of adsorption species is comparable with that of hydrogenation.

Table S1. Adsorption energies (E_a) calculated by the following equation for the most promising candidates: $E_a = E_{G\text{-}TM} - E_G - E_{TM}$, where the $E_{G\text{-}TM}$, E_G and E_{TM} are the energies of (B-doped) graphene supported SACs, graphene and the transition metal atom ($E(M_{bulk})/N$), respectively.

System	Adsorption Energy (eV)
TcB_3C_1	**-0.21**
TcB_2C_1	**+0.9**
MoB_2C_1	**+1.21**
MoB_1C_2	**+0.62**
ReB_1C_2	**+0.50**
CrB_3C_1	**-0.15**
HfB_1C_2	**-0.42**

Supplemental Note1

For preparing data to feed DNN, we first optimized $*N_2$ structures, then we investigated the first criteria (ΔE_{N2}<-0.5 eV). After that, among the qualified catalysts from first round, we checked the next step with $*N_2H$ structures ($\Delta G_{N2\text{-}N2H}$<0.55 eV). In the same way, for those catalysts which met the requirement of the second round, the third criteria ($\Delta G_{NH2\text{-}NH3}$ < 0.7 eV) was tested. After the third round we calculated the $\Delta G_{NH3\text{-}Desorbed}$ for all catalysts (the label of $\Delta G_{NH3\text{-}Desorbed}$ is the same as third criteria) . The labelling of catalysts in some cases is shown as follows:

Catalysts	label
ΔE_{N2}<-0.5 eV	1
($\Delta G_{N2\text{-}N2H}$<0.55 eV	1
$\Delta G_{NH2\text{-}NH3}$ < 0.7 eV	1

Catalysts	label
ΔE_{N2}<-0.5 eV	0
($\Delta G_{N2\text{-}N2H}$<0.55 eV	0
$\Delta G_{NH2\text{-}NH3} \geq$ 0.7 eV	0

Catalysts	label
ΔE_{N2}<-0.5 eV	0
($\Delta G_{N2\text{-}N2H} \geq$0.55 eV	0

Catalysts	label
$\Delta E_{N2} \geq$-0.5 eV	0
($\Delta G_{N2\text{-}N2H}$<0.55 eV	0

Supplemental Note2

For training a neural network, some parameters need to be determined (number of layers, input data and corresponding target data, loss function, optimizer, activation function). For classification of two classes, most of the times binary_crossentropy is the best selection for loss function[1]. In fact, loss function compares how close the predictions are to the targets. In this step, we need some validation process to see the feedback of changing parameters on the performance of network. Therefore, we divided training data into two parts (training + validation), as shown in Figs. S10-S11. In this study we tested different configurations of neural network using the binary_crossentropy loss function and the Adam optimizer, and finally reached to the best performance of two hidden layers composed of 10 neurons in each layer.

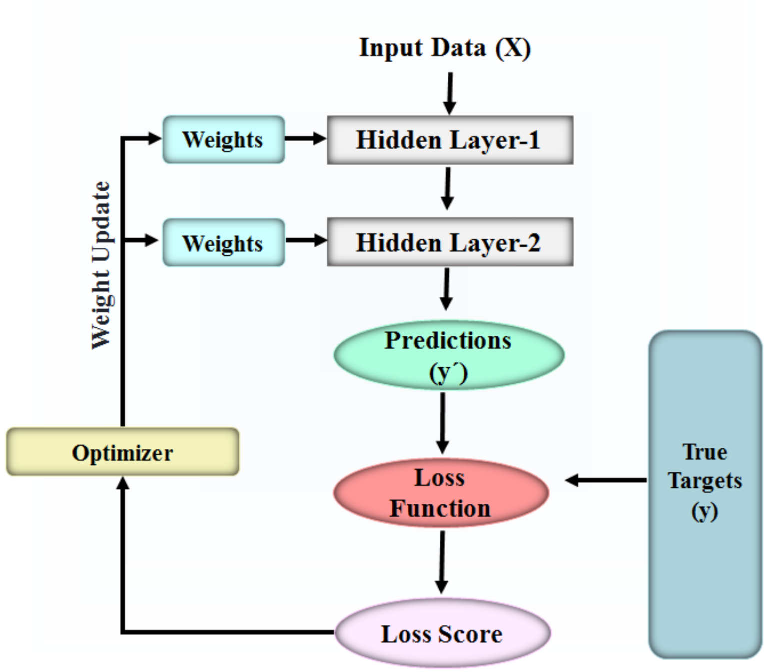

Fig. S10 Relation between loss function, optimizer, layers, input data and targets.

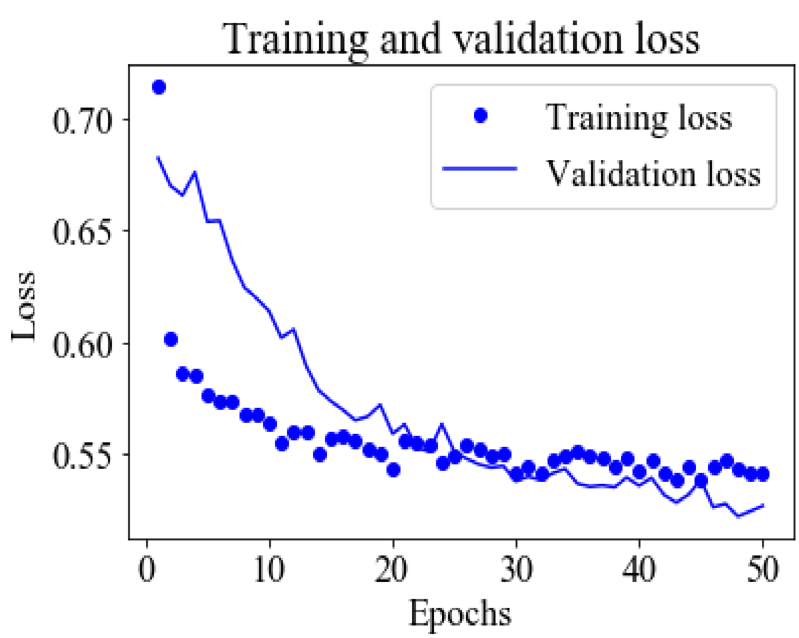

Fig. S11 Training and validation process (as an example) for selecting hyper parameters in the DNN.

Supplemental Note3

Here, we designed two separate models for prediction of catalytic activity (DNN) and free energy (LGBM). All of the training data are based on DFT results. For prediction of new compounds, it is better to use DNN first to search eligible catalysts, when the probability of a catalyst by DNN is less than 0.5, while in the second step the LGBM can be used to investigate the intermediate steps (to confirm whether the catalyst is active for NRR). In addition, we can prioritize the candidates for further computational and experimental work if both DNN and LGBM models simultaneously agree with eligibility of the given catalyst.

Table S2. Adsorption energy and free energy calculated through DFT and machine learning by Light Gradient Boosting Machine (LGBM) model for the data set. Here, we predicted ΔE_{N2}, ΔG_{N2-N2H}, $\Delta G_{NH2-NH3}$, $\Delta G_{NH3-Desorbed}$, by using the relaxed geometries of adsorbed species of $*N_2$, *N-NH (*N-*NH), $*NH_2$, and $*NH_3$, respectively, as the input to the model. N_2 molecules may be adsorbed to the surface by side-on or end-on configurations. We used LGBM to predict adsorption energy and free energy for data set. Many of non-eligible catalysts are removed from data set which are not presented in this table.

M	Structure	Adsorption species	DFT (eV)	LGBM (eV)	Type of data
V	$@B_4$	$^*N_2$ (end-on)	-1.000218	-1.0159378	Training
V	$@B_4$	$^*N_2$ (side-on)	-0.821077	-0.8555085	Training
Cu	$@B_4$	$^*N_2$ (end-on)	-0.497864	-0.4659206	Training
Cu	$@B_4$	$^*N_2$ (side-on)	-0.198403	-0.2730153	Training
Tc	$@B_4$	$^*N_2$ (end-on)	-1.341411	-1.2383999	Training
Tc	$@B_4$	$^*N_2$ (side-on)	-1.337701	-1.1268206	Training
Ru	$@B_4$	$^*N_2$ (end-on)	-1.235572	-1.1128087	Training
Ru	$@B_4$	$^*N_2$ (side-on)	-0.670788	-0.8909382	Training
Rh	$@B_4$	$^*N_2$ (end-on)	-0.479402	-0.6203168	Test
Rh	$@B_4$	$^*N_2$ (side-on)	-0.478874	-0.4058364	Test
Mo	$@B_4$	$^*N_2$ (end-on)	-1.226218	-1.2262215	Training
Mo	$@B_4$	$^*N_2$ (side-on)	-1.25054	-1.146023	Training
Au	$@B_4$	$^*N_2$ (end-on)	-0.108358	-0.1283405	Test
Au	$@B_4$	$^*N_2$ (side-on)	-0.111285	0.03220641	Test
Ag	$@B_4$	$^*N_2$ (end-on)	-0.116215	-0.1705582	Training
Ag	$@B_4$	$^*N_2$ (side-on)	-0.119417	-0.1539989	Training
Ir	$@B_4$	$^*N_2$ (end-on)	-0.605715	-0.4735573	Test
Ir	$@B_4$	$^*N_2$ (side-on)	-0.256806	-0.2912756	Training

M	Structure	Adsorption species	DFT (eV)	LGBM (eV)	Type of data
Pt	@B_4	$^*N_2$ (end-on)	-0.104224	-0.1067798	Training
Pt	@B_4	$^*N_2$ (side-on)	-0.120812	-0.0176758	Training
Pd	@B_4	$^*N_2$ (end-on)	-0.103645	-0.1447766	Training
Pd	@B_4	$^*N_2$ (side-on)	-0.117795	-0.286953	Training
Fe	@B_4	$^*N_2$ (end-on)	-1.132714	-0.9339176	Training
Fe	@B_4	$^*N_2$ (side-on)	-0.607004	-0.6754785	Training
Cr	@B_4	$^*N_2$ (end-on)	-0.863219	-0.8911311	Training
Cr	@B_4	$^*N_2$ (side-on)	-0.602066	-0.5678835	Training
Co	@B_4	$^*N_2$ (end-on)	-0.941514	-0.701653	Training
Co	@B_4	$^*N_2$ (side-on)	-0.41994	-0.5286075	Training
Ti	@B_4	$^*N_2$ (end-on)	-0.950535	-0.9292745	Test
Ti	@B_4	$^*N_2$ (side-on)	-0.814084	-0.8265996	Training
Sc	@B_4	$^*N_2$ (end-on)	-0.502265	-0.8080674	Test
Sc	@B_4	$^*N_2$ (side-on)	-0.464887	-0.4837888	Training
Ni	@B_4	$^*N_2$ (end-on)	-0.807857	-0.5317666	Training
Ni	@B_4	$^*N_2$ (side-on)	-0.851868	-0.3822635	Training
Mn	@B_4	$^*N_2$ (end-on)	-1.576677	-1.0122456	Training
Mn	@B_4	$^*N_2$ (side-on)	-1.284201	-0.760364	Training
Fe	@B_1C_3	$^*N_2$ (end-on)	-0.747741	-0.6611832	Training
Fe	@B_1C_3	$^*N_2$ (side-on)	-0.747258	-0.5205443	Test
Cu	@B_1C_3	$^*N_2$ (end-on)	-0.123166	-0.2781421	Training
Cu	@B_1C_3	$^*N_2$ (side-on)	-0.117671	-0.1413658	Training
Cr	@B_1C_3	$^*N_2$ (end-on)	-0.692366	-0.692168	Training
Cr	@B_1C_3	$^*N_2$ (side-on)	-0.327632	-0.5515291	Training
Co	@B_1C_3	$^*N_2$ (end-on)	-0.695231	-0.6646342	Training
Co	@B_1C_3	$^*N_2$ (side-on)	-0.695873	-0.5239952	Training
V	@B_1C_3	$^*N_2$ (end-on)	-0.916042	-0.8432707	Training
V	@B_1C_3	$^*N_2$ (side-on)	-0.672002	-0.7430636	Training
Ti	@B_1C_3	$^*N_2$ (end-on)	-0.864574	-0.8167785	Training
Ti	@B_1C_3	$^*N_2$ (side-on)	-0.663943	-0.8421664	Training
Sc	@B_1C_3	$^*N_2$ (end-on)	-0.533671	-0.5715778	Training
Sc	@B_1C_3	$^*N_2$ (side-on)	-0.348092	-0.4764701	Training
Ni	@B_1C_3	$^*N_2$ (end-on)	-0.521583	-0.4259715	Training
Ni	@B_1C_3	$^*N_2$ (side-on)	-0.518044	-0.2853326	Training
Mn	@B_1C_3	$^*N_2$ (end-on)	-0.6984	-0.5567578	Training
Mn	@B_1C_3	$^*N_2$ (side-on)	-0.699283	-0.4161188	Training
Tc	@B_1C_3	$^*N_2$ (end-on)	-0.891531	-0.8030027	Training

M	Structure	Adsorption species	DFT (eV)	LGBM (eV)	Type of data
Tc	@B_1C_3	$^*N_2$ (side-on)	-0.420858	-0.7312744	Training
Ru	@B_1C_3	$^*N_2$ (end-on)	-0.785748	-0.4674722	Test
Ru	@B_1C_3	$^*N_2$ (side-on)	-0.319123	-0.2875943	Training
Rh	@B_1C_3	$^*N_2$ (end-on)	-0.549367	-0.1946892	Training
Rh	@B_1C_3	$^*N_2$ (side-on)	-0.403817	-0.4120977	Training
Mo	@B_1C_3	$^*N_2$ (end-on)	-0.991205	-0.8221385	Training
Mo	@B_1C_3	$^*N_2$ (side-on)	-0.508124	-0.6725524	Training
Au	@B_1C_3	$^*N_2$ (end-on)	-0.129956	-0.1982397	Training
Au	@B_1C_3	$^*N_2$ (side-on)	-0.101619	-0.1899756	Training
Ag	@B_1C_3	$^*N_2$ (end-on)	-0.138025	-0.3249103	Training
Ag	@B_1C_3	$^*N_2$ (side-on)	-0.143702	-0.1139261	Training
Ir	@B_1C_3	$^*N_2$ (end-on)	-0.736302	-0.7377137	Test
Ir	@B_1C_3	$^*N_2$ (side-on)	-0.739402	-0.6369148	Training
Pt	@B_1C_3	$^*N_2$ (end-on)	-0.12661	-0.1329639	Training
Pt	@B_1C_3	$^*N_2$ (side-on)	-0.077972	-0.0414999	Training
Pd	@B_1C_3	$^*N_2$ (end-on)	-0.298097	-0.2395456	Training
Pd	@B_1C_3	$^*N_2$ (side-on)	-0.284541	-0.2129542	Training
Nb	@B_1C_3	$^*N_2$ (end-on)	-0.983636	-0.8115615	Training
Nb	@B_1C_3	$^*N_2$ (side-on)	-0.790763	-0.7421544	Training
Fe	@B_2C_2	$^*N_2$ (end-on)	-0.720428	-0.6899339	Training
Fe	@B_2C_2	$^*N_2$ (side-on)	-0.726116	-0.5168883	Training
Cu	@B_2C_2	$^*N_2$ (end-on)	-0.286738	-0.314846	Training
Cu	@B_2C_2	$^*N_2$ (side-on)	-0.293195	-0.128625	Training
Cr	@B_2C_2	$^*N_2$ (end-on)	-0.870997	-0.6696211	Test
Cr	@B_2C_2	$^*N_2$ (side-on)	-0.512982	-0.4965756	Training
Co	@B_2C_2	$^*N_2$ (end-on)	-0.726129	-0.7122419	Training
Co	@B_2C_2	$^*N_2$ (side-on)	-0.724168	-0.5391964	Training
V	@B_2C_2	$^*N_2$ (end-on)	-0.947236	-0.873426	Training
V	@B_2C_2	$^*N_2$ (side-on)	-0.899742	-0.7482127	Training
Ti	@B_2C_2	$^*N_2$ (end-on)	-0.911039	-0.8702086	Training
Ti	@B_2C_2	$^*N_2$ (side-on)	-0.965747	-0.8099229	Training
Sc	@B_2C_2	$^*N_2$ (end-on)	-0.892202	-0.8150846	Test
Sc	@B_2C_2	$^*N_2$ (side-on)	-0.934404	-0.695761	Test
Ni	@B_2C_2	$^*N_2$ (side-on)	-0.089315	-0.1225582	Training
Mn	@B_2C_2	$^*N_2$ (end-on)	-0.70773	-0.4946726	Test
Mn	@B_2C_2	$^*N_2$ (side-on)	-0.227778	-0.321627	Training
Tc	@B_2C_2	$^*N_2$ (end-on)	-0.961985	-0.8473053	Training

M	Structure	Adsorption species	DFT (eV)	LGBM (eV)	Type of data
Tc	$@B_2C_2$	$^*N_2$ (side-on)	-0.396638	-0.6288679	Training
Ru	$@B_2C_2$	$^*N_2$ (end-on)	-0.902568	-0.9455043	Training
Ru	$@B_2C_2$	$^*N_2$ (side-on)	-0.831711	-0.8091706	Training
Rh	$@B_2C_2$	$^*N_2$ (end-on)	-0.438473	-0.4409522	Training
Rh	$@B_2C_2$	$^*N_2$ (side-on)	-0.298152	-0.1717117	Training
Mo	$@B_2C_2$	$^*N_2$ (end-on)	-1.044273	-1.0490837	Training
Mo	$@B_2C_2$	$^*N_2$ (side-on)	-0.56552	-0.806216	Test
Ir	$@B_2C_2$	$^*N_2$ (end-on)	-0.540939	-0.5106569	Training
Ir	$@B_2C_2$	$^*N_2$ (side-on)	-0.314594	-0.1854188	Training
Nb	$@B_2C_2$	$^*N_2$ (end-on)	-0.963462	-0.7999705	Training
Nb	$@B_2C_2$	$^*N_2$ (side-on)	-0.925248	-0.8192364	Training
Ru	$@B_3$	$^*N_2$ (end-on)	-0.758486	-0.7544998	Training
Ru	$@B_3$	$^*N_2$ (side-on)	-0.349564	-0.3506389	Training
Rh	$@B_3$	$^*N_2$ (end-on)	-0.592409	-0.5953809	Training
Rh	$@B_3$	$^*N_2$ (side-on)	-0.233761	-0.3801123	Test
Mo	$@B_3$	$^*N_2$ (end-on)	-0.956401	-0.9406695	Training
Mo	$@B_3$	$^*N_2$ (side-on)	-1.079739	-1.0161005	Training
Nb	$@B_3$	$^*N_2$ (end-on)	-0.931344	-0.9855581	Training
Nb	$@B_3$	$^*N_2$ (side-on)	-0.97345	-0.8517475	Training
Co	$@B_3$	$^*N_2$ (end-on)	-1.07579	-0.7155042	Training
Co	$@B_3$	$^*N_2$ (side-on)	-0.612986	-0.5424586	Training
Fe	$@B_3$	$^*N_2$ (end-on)	-1.0367	-0.7849935	Training
Fe	$@B_3$	$^*N_2$ (side-on)	-0.541886	-0.5742129	Training
Cr	$@B_3$	$^*N_2$ (end-on)	-1.278046	-0.811904	Training
Cr	$@B_3$	$^*N_2$ (side-on)	-0.235363	-0.547035	Training
Mn	$@B_3$	$^*N_2$ (end-on)	-0.775028	-0.6634271	Training
Mn	$@B_3$	$^*N_2$ (side-on)	-0.491635	-0.5328219	Training
Ti	$@B_3$	$^*N_2$ (end-on)	-0.854153	-0.7332229	Training
Ti	$@B_3$	$^*N_2$ (side-on)	-0.80685	-0.7435217	Training
Sc	$@B_3$	$^*N_2$ (end-on)	-0.759423	-0.7383252	Training
Sc	$@B_3$	$^*N_2$ (side-on)	-0.718239	-0.6151442	Training
Tc	$@B_3$	$^*N_2$ (end-on)	-0.720426	-0.9850534	Training
Tc	$@B_3$	$^*N_2$ (side-on)	-0.345986	-0.4995029	Training
Ag	$@B_3$	$^*N_2$ (side-on)	-0.196123	-0.4498607	Training
Ir	$@B_3$	$^*N_2$ (end-on)	-0.700367	-0.6965415	Training
Ir	$@B_3$	$^*N_2$ (side-on)	-0.248328	-0.3433868	Training
Ru	$@B_3$	$^*H$ ($^*NNH$)	0.755664	0.75106448	Training

M	Structure	Adsorption species	DFT (eV)	LGBM (eV)	Type of data
Rh	$@B_3$	$^*H$ ($^*NNH$)	1.199135	0.93696342	Training
Mo	$@B_3$	$^*H$ ($^*NNH$)	0.280396	0.61319388	Training
Mo	$@B_3$	$^*H$	0.221341	0.3788939	Training
Nb	$@B_3$	$^*H$ ($^*NNH$)	0.305483	0.36741332	Training
Co	$@B_3$	$^*H$ ($^*NNH$)	1.001644	1.054067	Training
Co	$@B_3$	$^*H$	0.912845	0.65012733	Training
Fe	$@B_3$	$^*H$ ($^*NNH$)	0.942856	0.89857472	Test
Fe	$@B_3$	$^*H$	0.648008	0.87782172	Training
Cr	$@B_3$	$^*H$ ($^*NNH$)	0.790749	0.47045331	Training
Mn	$@B_3$	$^*H$ ($^*NNH$)	0.859308	0.51669043	Training
Ti	$@B_3$	$^*H$ ($^*NNH$)	0.302132	0.2442453	Training
Ti	$@B_3$	$^*H$	-0.248222	0.12403973	Training
Sc	$@B_3$	$^*H$ ($^*NNH$)	0.650094	0.54508006	Training
Sc	$@B_3$	$^*H$	0.065872	0.17677679	Training
Tc	$@B_3$	$^*H$ ($^*NNH$)	-0.305566	0.33497123	Training
Ir	$@B_3$	$^*H$ ($^*NNH$)	1.18126	0.83305259	Training
Fe	$@B_2C_2$	$^*H$ ($^*NNH$)	0.89183	0.7571625	Training
Cr	$@B_2C_2$	$^*H$ ($^*NNH$)	0.676216	0.7005217	Test
Cr	$@B_2C_2$	$^*H$	0.445427	0.56902209	Training
Co	$@B_2C_2$	$^*H$ ($^*NNH$)	0.839587	0.76085514	Test
V	$@B_2C_2$	$^*H$ ($^*NNH$)	0.627531	0.81123388	Training
V	$@B_2C_2$	$^*H$	0.581105	0.47158309	Training
Ti	$@B_2C_2$	$^*H$ ($^*NNH$)	0.811334	0.65351749	Training
Ti	$@B_2C_2$	$^*H$	0.356113	0.49165811	Training
Sc	$@B_2C_2$	$^*H$ ($^*NNH$)	0.699237	0.60945986	Training
Sc	$@B_2C_2$	$^*H$	0.161077	0.35477973	Training
Mn	$@B_2C_2$	$^*H$ ($^*NNH$)	0.746829	0.64705364	Training
Tc	$@B_2C_2$	$^*H$ ($^*NNH$)	0.612719	0.45367668	Training
Ru	$@B_2C_2$	$^*H$ ($^*NNH$)	1.055706	0.83842123	Training
Mo	$@B_2C_2$	$^*H$ ($^*NNH$)	0.847197	1.00656038	Training
Mo	$@B_2C_2$	$^*H$	0.450224	0.41795381	Training
Ir	$@B_2C_2$	$^*H$ ($^*NNH$)	0.949113	1.05241743	Training
Nb	$@B_2C_2$	$^*H$ ($^*NNH$)	0.687895	0.71480732	Training
Nb	$@B_2C_2$	$^*H$	0.612994	0.42829042	Training
Fe	$@B_1C_3$	$^*H$ ($^*NNH$)	0.852283	0.80190907	Training
Co	$@B_1C_3$	$^*H$ ($^*NNH$)	0.940698	0.82507338	Training
V	$@B_1C_3$	$^*H$ ($^*NNH$)	0.682656	0.67052029	Training

M	Structure	Adsorption species	DFT (eV)	LGBM (eV)	Type of data
V	@B_1C_3	$^*H$	0.472932	0.07530153	Training
Ti	@B_1C_3	$^*H$ ($^*NNH$)	0.862735	0.6479502	Training
Ti	@B_1C_3	$^*H$	0.336476	0.4317505	Training
Sc	@B_1C_3	$^*H$ ($^*NNH$)	1.292182	1.05675859	Training
Ni	@B_1C_3	$^*H$ ($^*NNH$)	1.248965	1.16953687	Training
Mn	@B_1C_3	$^*H$ ($^*NNH$)	0.802307	0.61358248	Test
Tc	@B_1C_3	$^*H$ ($^*NNH$)	1.060522	0.68013068	Training
Ru	@B_1C_3	$^*H$ ($^*NNH$)	1.027323	0.72995224	Training
Rh	@B_1C_3	$^*H$ ($^*NNH$)	-0.783328	0.12052253	Training
Mo	@B_1C_3	$^*H$ ($^*NNH$)	0.868285	0.67288789	Test
Mo	@B_1C_3	$^*H$	0.191069	0.36241746	Training
Ir	@B_1C_3	$^*H$ ($^*NNH$)	0.826928	0.51982642	Training
Nb	@B_1C_3	$^*H$ ($^*NNH$)	0.760206	0.6084208	Training
Nb	@B_1C_3	$^*H$	0.524987	0.36589498	Test
Fe	@B_4	$^*H$ ($^*NNH$)	0.701617	0.77621441	Training
Fe	@B_4	$^*H$	0.455509	0.62474358	Training
Cr	@B_4	$^*H$ ($^*NNH$)	0.731528	0.67960969	Training
Co	@B_4	$^*H$ ($^*NNH$)	0.663902	0.77458016	Training
Co	@B_4	$^*H$	0.399146	0.54406347	Training
V	@B_4	$^*H$ ($^*NNH$)	0.921031	0.58636288	Training
V	@B_4	$^*H$	0.186881	0.30995329	Training
Ti	@B_4	$^*H$ ($^*NNH$)	0.752522	0.58562954	Training
Ti	@B_4	$^*H$	0.212065	0.42417866	Training
Sc	@B_4	$^*H$ ($^*NNH$)	0.753915	0.63067889	Training
Sc	@B_4	$^*H$	0.127752	0.34996765	Training
Ni	@B_4	$^*H$ ($^*NNH$)	0.844	0.75256526	Test
Mn	@B_4	$^*H$ ($^*NNH$)	0.671078	0.63828618	Training
Mn	@B_4	$^*H$	0.452579	0.36362496	Training
Tc	@B_4	$^*H$ ($^*NNH$)	0.5	0.48153376	Training
Tc	@B_4	$^*H$	0.732644	0.43351695	Training
Ru	@B_4	$^*H$ ($^*NNH$)	0.512593	0.66322693	Training
Ru	@B_4	$^*H$	0.633289	0.49790941	Training
Mo	@B_4	$^*H$ ($^*NNH$)	0.492489	0.47621652	Training
Mo	@B_4	$^*H$	0.636914	0.35757385	Training
Ir	@B_4	$^*H$ ($^*NNH$)	0.523147	0.97554552	Training
Mo	@B_3	$^*NH_2$	-0.298682	-0.2867435	Training
Nb	@B_3	$^*NH_2$	-0.043531	-0.3336936	Training

M	Structure	Adsorption species	DFT (eV)	LGBM (eV)	Type of data
Ti	@B_3	$^*NH_2$	-0.078456	-0.1641644	Test
Tc	@B_3	$^*NH_2$	-0.655914	-0.6080694	Training
Cr	@B_2C_2	$^*NH_2$	-0.427937	-0.3653963	Test
Ti	@B_2C_2	$^*NH_2$	-0.27896	-0.2187841	Training
Sc	@B_2C_2	$^*NH_2$	0.225954	-0.1207549	Test
Mo	@B_2C_2	$^*NH_2$	-0.466853	-0.4549157	Training
V	@B_1C_3	$^*NH_2$	-0.454832	-0.4119138	Training
Ti	@B_1C_3	$^*NH_2$	-0.299989	-0.3021247	Test
Rh	@B_1C_3	$^*NH_2$	-0.603351	-0.301082	Training
Mo	@B_1C_3	$^*NH_2$	-0.401029	-0.3915514	Test
Nb	@B_1C_3	$^*NH_2$	-0.429772	-0.3578357	Training
Fe	@B_4	$^*NH_2$	-0.633176	-0.4428916	Training
Co	@B_4	$^*NH_2$	-0.731427	-0.5367507	Training
V	@B_4	$^*NH_2$	-0.439097	-0.2480691	Training
Ti	@B_4	$^*NH_2$	-0.143877	-0.2400341	Training
Sc	@B_4	$^*NH_2$	-0.23422	-0.3008712	Training
Mn	@B_4	$^*NH_2$	-0.658221	-0.3396492	Training
Mo	@B_3	$^*NH_3$	2.016044	1.60740419	Training
Nb	@B_3	$^*NH_3$	1.35639	1.41307846	Training
Ti	@B_3	$^*NH_3$	1.419393	1.68125019	Training
Tc	@B_3	$^*NH_3$	1.199671	1.40323341	Training
Cr	@B_2C_2	$^*NH_3$	1.018217	1.018731403	Training
Ti	@B_2C_2	$^*NH_3$	1.115486	1.17361478	Training
Sc	@B_2C_2	$^*NH_3$	1.109081	1.20557543	Training
Mo	@B_2C_2	$^*NH_3$	2.04941	1.5528172	Training
V	@B_1C_3	$^*NH_3$	1.132718	1.13469621	Training
Ti	@B_1C_3	$^*NH_3$	1.19624	1.12104053	Training
Rh	@B_1C_3	$^*NH_3$	2.242327	2.66458588	Training
Nb	@B_1C_3	$^*NH_3$	1.20204	1.15208068	Training
Fe	@B_1C_3	$^*NH_3$	1.199061	1.11367671	Training
Co	@B_4	$^*NH_3$	1.082685	1.041678761	Training
V	@B_4	$^*NH_3$	1.157345	1.1725904	Training
Ti	@B_4	$^*NH_3$	1.193598	1.14144442	Training
Sc	@B_4	$^*NH_3$	1.17748	1.18414929	Training
Mn	@B_4	$^*NH_3$	2.309851	1.71558205	Training
Ru	@B_1C_2	$^*N_2$ (end-on)	-0.505406	0.00984743	Training
Ru	@B_1C_2	$^*N_2$ (side-on)	-0.218374	-0.1267919	Training

M	Structure	Adsorption species	DFT (eV)	LGBM (eV)	Type of data
Rh	$@B_1C_2$	$^*N_2$ (side-on)	-1.807935	0.05609718	Training
Ir	$@B_1C_2$	$^*N_2$ (end-on)	-0.621356	-0.3522685	Test
Ir	$@B_1C_2$	$^*N_2$ (side-on)	-0.620011	-0.446793	Training
Mn	$@B_1C_2$	$^*N_2$ (end-on)	-0.503778	-0.1605519	Test
Mn	$@B_1C_2$	$^*N_2$ (side-on)	-0.288755	-0.143707	Training
Mo	$@B_1C_2$	$^*N_2$ (end-on)	-0.875103	-0.3773182	Training
Mo	$@B_1C_2$	$^*N_2$ (side-on)	-0.755745	-0.3474363	Training
Ru	$@B_1C_2$	$^*H$ ($^*NNH$)	0.946284	0.21840093	Training
Ru	$@B_1C_2$	$^*H$	0.930935	0.25853283	Training
Ir	$@B_1C_2$	$^*H$ ($^*NNH$)	1.121251	0.84941229	Training
Ir	$@B_1C_2$	$^*H$ ($^*NNH$)	1.735182	1.08196423	Training
Mn	$@B_1C_2$	$^*H$ ($^*NNH$)	0.972829	0.12202056	Test
Mn	$@B_1C_2$	$^*H$	0.717517	0.17470221	Training
Mo	$@B_1C_2$	$^*H$ ($^*NNH$)	0.540865	-0.0666073	Training
Mo	$@B_1C_2$	$^*H$	0.431584	-0.2143828	Training
Ru	$@B_1C_2$	$^*NH_3$	0.712551	1.08611614	Training
Ru	$@B_1C_2$	$^*NH_2$	-1.403011	-0.7963459	Training
Rh	$@B_1C_2$	$^*NH_3$	2.016966	1.65013142	Training
Rh	$@B_1C_2$	$^*NH_2$	-1.106188	-0.5692282	Training
Ir	$@B_1C_2$	$^*NH_3$	1.224852	1.1573749	Test
Ir	$@B_1C_2$	$^*NH_2$	-0.555849	-0.3742803	Training
Mn	$@B_1C_2$	$^*NH_3$	1.212587	1.34354148	Training
Mn	$@B_1C_2$	$^*NH_2$	-0.718674	-0.674122	Test
Mo	$@B_1C_2$	$^*NH_3$	-0.45365	1.03222545	Training
Mo	$@B_1C_2$	$^*NH_2$	-0.106948	-0.5751585	Training
V	$@B_3C_1$	$^*N_2$ (end-on)	-0.95476	-1.0173142	Training
Ti	$@B_3C_1$	$^*N_2$ (end-on)	-0.838026	-0.8598879	Training
Tc	$@B_3C_1$	$^*N_2$ (end-on)	-1.404214	-1.3517295	Training
Sc	$@B_3C_1$	$^*N_2$ (end-on)	-0.716851	-0.564543	Training
Ru	$@B_3C_1$	$^*N_2$ (end-on)	-1.025172	-0.9704925	Training
Rh	$@B_3C_1$	$^*N_2$ (end-on)	-0.607028	-0.6177672	Training
Nb	$@B_3C_1$	$^*N_2$ (end-on)	-0.943928	-1.0229582	Test
Mo	$@B_3C_1$	$^*N_2$ (end-on)	-1.020302	-1.063355	Test
Mn	$@B_3C_1$	$^*N_2$ (end-on)	-0.795688	-0.8578654	Test
Ir	$@B_3C_1$	$^*N_2$ (end-on)	-0.705535	-0.7084944	Test
Fe	$@B_3C_1$	$^*N_2$ (end-on)	-0.916291	-0.8010509	Training
Cu	$@B_3C_1$	$^*N_2$ (end-on)	-0.381467	-0.4504112	Training

M	Structure	Adsorption species	DFT (eV)	LGBM (eV)	Type of data
Cr	@B_3C_1	$^*N_2$ (end-on)	-1.118521	-1.11846268	Training
Co	@B_3C_1	$^*N_2$ (end-on)	-1.044935	-0.8115021	Training
V	@B_2C_1	$^*N_2$ (end-on)	-0.927215	-0.8901141	Training
Ti	@B_2C_1	$^*N_2$ (end-on)	-0.833313	-0.9367848	Training
Tc	@B_2C_1	$^*N_2$ (end-on)	-1.208903	-1.0296607	Training
Sc	@B_2C_1	$^*N_2$ (end-on)	-0.646234	-0.8721462	Training
Ru	@B_2C_1	$^*N_2$ (end-on)	-0.943304	-0.9288406	Training
Rh	@B_2C_1	$^*N_2$ (end-on)	-0.897664	-0.6601952	Training
Nb	@B_2C_1	$^*N_2$ (end-on)	-0.893754	-0.9931795	Training
Mo	@B_2C_1	$^*N_2$ (end-on)	-1.237849	-1.1329331	Training
Mn	@B_2C_1	$^*N_2$ (end-on)	-0.98159	-0.7540752	Training
Ir	@B_2C_1	$^*N_2$ (end-on)	-1.083412	-0.9195269	Training
Fe	@B_2C_1	$^*N_2$ (end-on)	-1.168056	-0.7540752	Training
Cu	@B_2C_1	$^*N_2$ (end-on)	-0.543787	-0.2843048	Test
Cr	@B_2C_1	$^*N_2$ (end-on)	-0.888687	-0.8247868	Test
Co	@B_2C_1	$^*N_2$ (end-on)	0.10879	-0.6502175	Training
Ti	@B_3C_1	$^*H$ ($^*NNH$)	0.741129	0.79007017	Training
Tc	@B_3C_1	$^*H$ ($^*NNH$)	0.447443	0.62924851	Training
Sc	@B_3C_1	$^*H$ ($^*NNH$)	0.844086	0.73540953	Training
Ru	@B_3C_1	$^*H$ ($^*NNH$)	0.692418	0.71191056	Training
Rh	@B_3C_1	$^*H$ ($^*NNH$)	1.132483	0.98377478	Test
Nb	@B_3C_1	$^*H$ ($^*NNH$)	0.74893	0.68977424	Training
Mo	@B_3C_1	$^*H$ ($^*NNH$)	0.831604	0.62810154	Training
Mn	@B_3C_1	$^*H$ ($^*NNH$)	0.710472	0.78235185	Training
Ir	@B_3C_1	$^*H$ ($^*NNH$)	0.878967	0.8134254	Training
Fe	@B_3C_1	$^*H$ ($^*NNH$)	0.769725	0.87800357	Training
Cu	@B_3C_1	$^*H$ ($^*NNH$)	1.609481	1.25987323	Test
Cr	@B_3C_1	$^*H$ ($^*NNH$)	0.294546	0.31916071	Training
Co	@B_3C_1	$^*H$ ($^*NNH$)	0.79891	0.91891066	Test
V	@B_2C_1	$^*H$ ($^*NNH$)	0.707071	0.7984288	Training
Ti	@B_2C_1	$^*H$ ($^*NNH$)	0.684693	0.71608209	Test
Tc	@B_2C_1	$^*H$ ($^*NNH$)	0.544283	0.73005095	Training
Sc	@B_2C_1	$^*H$ ($^*NNH$)	0.806984	0.74180785	Training
Ru	@B_2C_1	$^*H$ ($^*NNH$)	0.730017	0.87354565	Training
Rh	@B_2C_1	$^*H$ ($^*NNH$)	1.81469	1.26304819	Test
Nb	@B_2C_1	$^*H$ ($^*NNH$)	0.67643	0.71334021	Training
Mo	@B_2C_1	$^*H$ ($^*NNH$)	0.551769	0.69142145	Training

M	Structure	Adsorption species	DFT (eV)	LGBM (eV)	Type of data
Mn	@B_2C_1	*H (*NNH)	0.920119	0.93558471	Test
Fe	@B_2C_1	*H (*NNH)	1.165185	0.960690	Training
Cu	@B_2C_1	*H (*NNH)	1.34116	1.28853439	Training
Cr	@B_2C_1	*H (*NNH)	0.738937	0.78012006	Test
Co	@B_2C_1	*H (*NNH)	1.233718	1.2133156	Training
Tc	@B_2C_1	$^*NH_2$	-0.67223	0.526969	Training
Tc	@B_2C_1	$^*NH_3$	0.598966	0.4156656	Training
Mo	@B_2C_1	$^*NH_2$	-0.305873	-0.215155	Training
Mo	@B_2C_1	$^*NH_3$	0.612333	0.605699	Training
Cr	@B_3C_1	$^*NH_2$	-0.515266	-0.405623	Training
Cr	@B_3C_1	$^*NH_3$	0.69587	0.6235855	Training
Tc	@B_3C_1	$^*NH_2$	-0.452333	0.342010	Training
Tc	@B_3C_1	$^*NH_3$	0.65122	0.550293	Training
Zr	@B_3C_1	$^*N_2$(end-on)	-0.88	-1.13639	Test
Zr	@B_3C_1	$^*N_2$(side-on)	-0.84	-0.73329	Test
Hf	@B_3C_1	$^*N_2$(end-on)	-1.08	-1.20669	Test
Hf	@B_3C_1	$^*N_2$(side-on)	-1.03	-1.02489	Test
Os	@B_3C_1	$^*N_2$(end-on)	-1.14	-1.21195	Test
Os	@B_3C_1	$^*N_2$(side-on)	-1.25	-1.28361	Test
Hf	@B_1C_2	$^*N_2$(end-on)	-0.89	-0.82991	Test
Hf	@B_1C_2	$^*N_2$(side-on)	-0.73	-0.72408	Test
Re	@B_1C_2	$^*N_2$(end-on)	-1.05	-0.98283	Test
Re	@B_1C_2	$^*N_2$(side-on)	-0.95	-0.73584	Test
Hf	@B_1C_3	$^*N_2$(end-on)	-0.98	-0.75475	Test
Hf	@B_1C_3	$^*N_2$(side-on)	-0.98	-0.65343	Test
Ta	@B_1C_3	$^*N_2$(end-on)	-1.16	-1.06265	Test
Ta	@B_1C_3	$^*N_2$(side-on)	-1.15	-0.93593	Test
W	@B_1C_3	$^*N_2$(end-on)	-1.09	-0.76413	Test
W	@B_1C_3	$^*N_2$(side-on)	-0.9	-0.81224	Test
Re	@B_1C_3	$^*N_2$(end-on)	-1.09	-0.9561	Test
Re	@B_1C_3	$^*N_2$(side-on)	-0.81	-0.86976	Test
Os	@B_1C_3	$^*N_2$(end-on)	-1.12	-0.90906	Test
Os	@B_1C_3	$^*N_2$(side-on)	-0.54	-0.71734	Test
Re	@B_1C_2	*H	0.47	0.51734	Test
Hf	@B_1C_2	*H	0.42	0.55173	Test
Re	@B_1C_2	$^*NH_2$	-0.54	-0.44173	Test
Hf	@B_1C_2	$^*NH_2$	0.34	0.43734	Test

M	Structure	Adsorption species	DFT (eV)	LGBM (eV)	Type of data
Re	@B_1C_2	$^*NH_3$	0.75	0.91734	Test
Hf	@B_1C_2	$^*NH_3$	0.61	0.41734	Test

Table S3. Adsorption energies on nitrogen-doped graphene SACs which were calculated with DFT (PBE-TS) towards the generalization of this model (Light Gradient Boosting Machine (LGBM)). ΔE_{N2}, was calculated using the relaxed geometries of adsorbed species of $*N_2$. N_2 molecules may be adsorbed to the surface by side-on or end-on configurations.

M	Structure	Adsorption species	DFT(eV)	LGBM (eV)	Type of data
Cu	@N_1C_3	$^*N_2$ (side-on)	-0.1108336	-0.1148755	Training
Cr	@N_1C_3	$^*N_2$ (end-on)	-0.705851	-0.5952254	Training
Cr	@N_1C_3	$^*N_2$ (side-on)	-0.4207922	-0.4465378	Training
Co	@N_1C_3	$^*N_2$ (end-on)	-0.5738711	-0.5718874	Training
Co	@N_1C_3	$^*N_2$ (side-on)	-0.5764249	-0.4312485	Test
V	@N_1C_3	$^*N_2$ (end-on)	-0.954832	-0.8012133	Test
V	@N_1C_3	$^*N_2$ (side-on)	-0.7339115	-0.7071313	Training
Ti	@N_1C_3	$^*N_2$ (end-on)	-0.7458274	-0.7781183	Training
Ti	@N_1C_3	$^*N_2$ (side-on)	-0.7445972	-0.7169227	Training
Sc	@N_1C_3	$^*N_2$ (end-on)	-0.4274947	-0.4495829	Training
Sc	@N_1C_3	$^*N_2$ (side-on)	-0.419891	-0.5142759	Training
Ni	@N_1C_3	$^*N_2$ (end-on)	-0.1939898	-0.2239279	Training
Ni	@N_1C_3	$^*N_2$ (side-on)	-0.1141969	-0.1088087	Training
Mn	@N_1C_3	$^*N_2$ (end-on)	-0.4661102	-0.3962384	Training
Mn	@N_1C_3	$^*N_2$ (side-on)	-0.4680776	-0.2555994	Training
Tc	@N_1C_3	$^*N_2$ (end-on)	-1.4385322	-1.077493	Training
Tc	@N_1C_3	$^*N_2$ (side-on)	-0.7833494	-0.9606695	Training
Ru	@N_1C_3	$^*N_2$ (end-on)	-0.8134872	-0.9504026	Training
Ru	@N_1C_3	$^*N_2$ (side-on)	-0.5275212	-0.6845019	Training
Rh	@N_1C_3	$^*N_2$ (end-on)	-0.6651124	-0.5271646	Training
Rh	@N_1C_3	$^*N_2$ (side-on)	-0.2846429	-0.3927525	Training
Mo	@N_1C_3	$^*N_2$ (end-on)	-1.1006048	-0.988664	Training
Mo	@N_1C_3	$^*N_2$ (side-on)	-0.7596997	-0.7051345	Training
Au	@N_1C_3	$^*N_2$ (end-on)	-0.1125256	-0.0118095	Training
Au	@N_1C_3	$^*N_2$ (side-on)	-0.0957819	-0.1580449	Training
Ag	@N_1C_3	$^*N_2$ (end-on)	-0.1250384	-0.0603872	Test

M	Structure	Adsorption species	DFT(eV)	LGBM (eV)	Type of data
Ag	@N_1C_3	$^*N_2$ (side-on)	-0.1054288	-0.1989271	Training
Ir	@N_1C_3	$^*N_2$ (end-on)	-0.7601747	-0.4532457	Training
Ir	@N_1C_3	$^*N_2$ (side-on)	-0.2676219	-0.2076268	Test
Pt	@N_1C_3	$^*N_2$ (end-on)	-0.0834193	-0.1344807	Training
Pt	@N_1C_3	$^*N_2$ (side-on)	-0.0813602	0.05787638	Training
Pd	@N_1C_3	$^*N_2$ (end-on)	-0.0925732	-0.1253467	Training
Pd	@N_1C_3	$^*N_2$ (side-on)	-0.0778092	0.12249389	Test
Fe	@N_2C_2	$^*N_2$ (end-on)	-0.5971324	-0.4870032	Training
Fe	@N_2C_2	$^*N_2$ (side-on)	-0.5991219	-0.3717392	Training
Cu	@N_2C_2	$^*N_2$ (end-on)	0.03931458	-0.1242686	Training
Cu	@N_2C_2	$^*N_2$ (side-on)	0.02887592	0.0187114	Training
Cr	@N_2C_2	$^*N_2$ (end-on)	-0.7219705	-0.7331897	Training
Cr	@N_2C_2	$^*N_2$ (side-on)	-0.3513142	-0.3530946	Training
Co	@N_2C_2	$^*N_2$ (end-on)	-0.7407023	-0.6299749	Training
Co	@N_2C_2	$^*N_2$ (side-on)	-0.7453868	-0.4569294	Training
V	@N_2C_2	$^*N_2$ (end-on)	-0.9246307	-0.8531223	Training
V	@N_2C_2	$^*N_2$ (side-on)	-0.6636331	-0.7686602	Training
Ti	@N_2C_2	$^*N_2$ (end-on)	-0.7233012	-0.8650679	Training
Ti	@N_2C_2	$^*N_2$ (side-on)	-0.5707619	-0.8107131	Training
Sc	@N_2C_2	$^*N_2$ (end-on)	-0.4498179	-0.5223829	Training
Sc	@N_2C_2	$^*N_2$ (side-on)	-0.4496847	-0.398396	Training
Ni	@N_2C_2	$^*N_2$ (end-on)	-0.1136445	-0.2934363	Training
Ni	@N_2C_2	$^*N_2$ (side-on)	-0.1255103	-0.1203908	Training
Mn	@N_2C_2	$^*N_2$ (end-on)	-0.4274666	-0.3972724	Training
Mn	@N_2C_2	$^*N_2$ (side-on)	-0.4275923	-0.2242269	Test
Tc	@N_2C_2	$^*N_2$ (end-on)	-1.1343432	-1.0876506	Training
Tc	@N_2C_2	$^*N_2$ (side-on)	-1.0777318	-0.9374049	Training
Ru	@N_2C_2	$^*N_2$ (end-on)	-1.0366456	-0.894765	Training
Ru	@N_2C_2	$^*N_2$ (side-on)	0.10511505	-0.5823287	Training
Rh	@N_2C_2	$^*N_2$ (end-on)	-0.4962524	-0.4779433	Training
Rh	@N_2C_2	$^*N_2$ (side-on)	-0.0509304	-0.3561215	Training
Mo	@N_2C_2	$^*N_2$ (end-on)	-1.0432793	-1.0387498	Training
Mo	@N_2C_2	$^*N_2$ (side-on)	-0.9455138	-0.8676061	Test
Au	@N_2C_2	$^*N_2$ (end-on)	-0.1082125	-0.1574426	Training
Au	@N_2C_2	$^*N_2$ (side-on)	-0.1079261	-0.0808303	Training
Ag	@N_2C_2	$^*N_2$ (end-on)	-0.108221	-0.1074314	Training
Ag	@N_2C_2	$^*N_2$ (side-on)	-0.1049765	-0.1863112	Training

M	Structure	Adsorption species	DFT(eV)	LGBM (eV)	Type of data
Ir	$@N_2C_2$	$^*N_2$ (end-on)	-0.6656983	-0.662607	Training
Ir	$@N_2C_2$	$^*N_2$ (side-on)	-0.0116255	0.04045759	Training
Pt	$@N_2C_2$	$^*N_2$ (end-on)	-0.1055131	-0.1762872	Test
Pt	$@N_2C_2$	$^*N_2$ (side-on)	-0.105759	-0.1347416	Training
Pd	$@N_2C_2$	$^*N_2$ (end-on)	-0.0959427	-0.1357052	Test
Pd	$@N_2C_2$	$^*N_2$ (side-on)	-0.1043901	-0.0102486	Test
Fe	$@N_3C_1$	$^*N_2$ (end-on)	-0.6028458	-0.5606431	Training
Fe	$@N_3C_1$	$^*N_2$ (side-on)	-0.0716999	-0.3733501	Test
Cu	$@N_3C_1$	$^*N_2$ (end-on)	-0.113096	-0.3071667	Training
Cu	$@N_3C_1$	$^*N_2$ (side-on)	-0.1060563	-0.0463157	Training
Cr	$@N_3C_1$	$^*N_2$ (end-on)	-0.7678833	-0.7631348	Training
Cr	$@N_3C_1$	$^*N_2$ (side-on)	-0.7742601	-0.7059997	Training
Co	$@N_3C_1$	$^*N_2$ (end-on)	-0.637376	-0.6413775	Training
Co	$@N_3C_1$	$^*N_2$ (side-on)	0.13240921	-0.470203	Training
V	$@N_3C_1$	$^*N_2$ (end-on)	-1.032985	-0.9503306	Training
V	$@N_3C_1$	$^*N_2$ (side-on)	-0.8065837	-0.8951246	Training
Ti	$@N_3C_1$	$^*N_2$ (end-on)	-3.0699242	-2.0401055	Training
Sc	$@N_3C_1$	$^*N_2$ (end-on)	-0.5169557	-0.528667	Test
Sc	$@N_3C_1$	$^*N_2$ (side-on)	-0.5294164	-0.44942	Test
Ni	$@N_3C_1$	$^*N_2$ (end-on)	-0.1159272	-0.3896619	Training
Ni	$@N_3C_1$	$^*N_2$ (side-on)	-0.1143464	-0.1553075	Training
Mn	$@N_3C_1$	$^*N_2$ (end-on)	-0.3078969	-0.34258	Training
Mn	$@N_3C_1$	$^*N_2$ (side-on)	0.01714834	-0.2057419	Training
Tc	$@N_3C_1$	$^*N_2$ (end-on)	-1.4414109	-1.2993717	Training
Tc	$@N_3C_1$	$^*N_2$ (side-on)	-1.4377011	-1.1738671	Test
Ru	$@N_3C_1$	$^*N_2$ (end-on)	-1.3355724	-1.01878	Test
Ru	$@N_3C_1$	$^*N_2$ (side-on)	-0.670788	-0.8709927	Test
Rh	$@N_3C_1$	$^*N_2$ (end-on)	-0.4794018	-0.3535791	Training
Rh	$@N_3C_1$	$^*N_2$ (side-on)	-0.4788738	-0.3283482	Training
Mo	$@N_3C_1$	$^*N_2$ (end-on)	-1.2262183	-1.2600568	Training
Mo	$@N_3C_1$	$^*N_2$ (side-on)	-1.2505401	-1.1144949	Training
Au	$@N_3C_1$	$^*N_2$ (end-on)	-0.1083577	-0.1249483	Training
Au	$@N_3C_1$	$^*N_2$ (side-on)	-0.1112852	-0.048336	Training
Ag	$@N_3C_1$	$^*N_2$ (end-on)	-0.1162146	-0.0484312	Test
Ag	$@N_3C_1$	$^*N_2$ (side-on)	-0.1194169	-0.1194373	Training
Ir	$@N_3C_1$	$^*N_2$ (end-on)	-0.6057149	-0.4496596	Training
Ir	$@N_3C_1$	$^*N_2$ (side-on)	-0.2568056	-0.3760662	Training

M	Structure	Adsorption species	DFT(eV)	LGBM (eV)	Type of data
Pt	@N_3C_1	$^*N_2$ (end-on)	-0.1042239	-0.2156503	Training
Pt	@N_3C_1	$^*N_2$ (side-on)	-0.1208125	-0.082497	Training
Pd	@N_3C_1	$^*N_2$ (end-on)	-0.1036452	-0.1532953	Training
Pd	@N_3C_1	$^*N_2$ (side-on)	-0.1177953	-0.227655	Training
Fe	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.4624389	-0.6956954	Training
Fe	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.0503088	-0.321627	Training
Cu	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.1102661	-0.104053	Training
Cu	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.1103359	-0.0553299	Training
Cr	@N_2C_2(Cis)	$^*N_2$ (end-on)	-1.0022138	-0.7331897	Training
Cr	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.181355	-0.3902483	Training
Co	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.6042606	-0.6299749	Training
Co	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.6089451	-0.4569294	Training
V	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.9350226	-0.8531223	Training
V	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.6601324	-0.6908259	Test
Ti	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.7669517	-1.1367981	Training
Ti	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.5829068	-0.9412999	Training
Sc	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.4607695	-0.5054588	Training
Sc	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.4700263	-0.5595435	Training
Ni	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.1161718	-0.3901476	Training
Ni	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.109085	-0.1203908	Training
Mn	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.3587533	-0.3581062	Training
Mn	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.1207425	-0.1790924	Training
Tc	@N_2C_2(Cis)	$^*N_2$ (end-on)	-1.4943514	-1.0876506	Training
Tc	@N_2C_2(Cis)	$^*N_2$ (side-on)	-1.0162423	-0.9579065	Training
Ru	@N_2C_2(Cis)	$^*N_2$ (end-on)	-1.3569061	-1.0500254	Test
Ru	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.8592559	-0.8237982	Training
Rh	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.7233946	-0.2217112	Training
Rh	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.06878	-0.3561215	Training
Mo	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.9555222	-0.9412209	Training
Mo	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.6532356	-1.001459	Training
Au	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.1108014	-0.1893284	Training
Au	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.1103142	-0.0274063	Training
Ag	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.1151192	-0.5032897	Training
Ag	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.1183769	0.07211319	Training
Ir	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.6950209	-0.3747444	Test
Ir	@N_2C_2(Cis)	$^*N_2$ (side-on)	0.20406662	0.04045759	Training
Pt	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.1085205	-0.2838257	Training

M	Structure	Adsorption species	DFT(eV)	LGBM (eV)	Type of data
Pt	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.1097484	-0.0583966	Training
Pd	@N_2C_2(Cis)	$^*N_2$ (end-on)	-0.1079235	-0.2348276	Training
Pd	@N_2C_2(Cis)	$^*N_2$ (side-on)	-0.1230199	-0.1091652	Test
Fe	@C_4	$^*N_2$ (end-on)	-0.5420216	-0.5518086	Training
Fe	@C_4	$^*N_2$ (side-on)	-0.1798949	-0.4044904	Training
Cu	@C_4	$^*N_2$ (end-on)	-0.1238216	-0.2555145	Test
Cu	@C_4	$^*N_2$ (side-on)	-0.1158216	-0.2158261	Test
Cr	@C_4	$^*N_2$ (end-on)	-0.6615316	-0.5093424	Training
Cr	@C_4	$^*N_2$ (side-on)	-0.6681562	-0.5730243	Training
Co	@C_4	$^*N_2$ (end-on)	-0.6613516	-0.6015663	Training
Co	@C_4	$^*N_2$ (side-on)	-0.2838516	-0.4312485	Training
V	@C_4	$^*N_2$ (end-on)	-0.7679116	-0.8137833	Training
V	@C_4	$^*N_2$ (side-on)	-0.7610568	-0.645681	Training
Ti	@C_4	$^*N_2$ (end-on)	-0.7292216	-0.7239427	Training
Ti	@C_4	$^*N_2$ (side-on)	-0.7300032	-0.7250758	Training
Sc	@C_4	$^*N_2$ (end-on)	-0.4152216	-0.5850614	Training
Sc	@C_4	$^*N_2$ (side-on)	-0.4279518	-0.4567474	Training
Ni	@C_4	$^*N_2$ (end-on)	-0.2598216	-0.2494477	Training
Ni	@C_4	$^*N_2$ (side-on)	-0.1228216	-0.2246868	Training
Mn	@C_4	$^*N_2$ (end-on)	-0.5402976	-0.3184041	Training
Mn	@C_4	$^*N_2$ (side-on)	-0.2061483	-0.1777652	Test
Tc	@C_4	$^*N_2$ (end-on)	-0.7386829	-0.9038027	Training
Tc	@C_4	$^*N_2$ (side-on)	-0.7938646	-0.7694562	Training
Ru	@C_4	$^*N_2$ (end-on)	-0.6652144	-0.5298654	Training
Ru	@C_4	$^*N_2$ (side-on)	-0.6620166	-0.6991085	Training
Rh	@C_4	$^*N_2$ (end-on)	-0.4594145	-0.1304886	Training
Rh	@C_4	$^*N_2$ (side-on)	-0.155208	-0.179998	Training
Mo	@C_4	$^*N_2$ (end-on)	-0.8640704	-0.8997314	Training
Mo	@C_4	$^*N_2$ (side-on)	-0.8226017	-0.691969	Training
Au	@C_4	$^*N_2$ (end-on)	-0.1214082	-0.1655398	Training
Au	@C_4	$^*N_2$ (side-on)	-0.0778098	-0.3238006	Test
Ag	@C_4	$^*N_2$ (end-on)	-0.071599	-0.5221536	Training
Ag	@C_4	$^*N_2$ (side-on)	-0.0845156	-0.2826619	Training
Ir	@C_4	$^*N_2$ (end-on)	-0.6417499	-0.3478212	Training
Ir	@C_4	$^*N_2$ (side-on)	-0.093067	-0.2344029	Training
Pt	@C_4	$^*N_2$ (end-on)	-0.0831932	-0.2300224	Training
Pt	@C_4	$^*N_2$ (side-on)	-0.0681633	-0.0518215	Training

M	Structure	Adsorption species	DFT(eV)	LGBM (eV)	Type of data
Pd	@C_4	$^*N_2$ (end-on)	-0.127983	0.03864099	Training
Pd	@C_4	$^*N_2$ (side-on)	-0.0729265	0.31877015	Training
Fe	@N_4	$^*N_2$ (end-on)	-0.1939416	-0.4306905	Test
Fe	@N_4	$^*N_2$ (side-on)	-0.1899916	-0.4329245	Training
Cu	@N_4	$^*N_2$ (end-on)	-0.1357416	-0.3349077	Training
Cu	@N_4	$^*N_2$ (side-on)	-0.1007916	-0.1354911	Training
Cr	@N_4	$^*N_2$ (end-on)	-0.1649216	-0.7727024	Training
Cr	@N_4	$^*N_2$ (side-on)	0.14427838	-0.3271927	Training
Co	@N_4	$^*N_2$ (end-on)	-0.1881216	-0.7535442	Training
Co	@N_4	$^*N_2$ (side-on)	-0.0872816	-0.4269229	Test
V	@N_4	$^*N_2$ (end-on)	-1.1080916	-0.9854593	Training
V	@N_4	$^*N_2$(side-on)	-0.7300616	-0.7692654	Training
Ti	@N_4	$^*N_2$ (end-on)	-1.0426216	-1.1573641	Test
Ti	@N_4	$^*N_2$ (side-on)	-1.1020016	-1.0659442	Training
Sc	@N_4	$^*N_2$ (end-on)	-0.7910616	-0.7103831	Training
Sc	@N_4	$^*N_2$ (side-on)	-0.6373416	-0.5994546	Training
Ni	@N_4	$^*N_2$ (end-on)	-0.1302866	-0.3663901	Training
Ni	@N_4	$^*N_2$ (side-on)	-0.0994476	-0.1566774	Training
Mn	@N_4	$^*N_2$ (end-on)	-0.1773216	-0.4335602	Training
Mn	@N_4	$^*N_2$ (side-on)	-0.1406116	-0.2704275	Training
Tc	@N_4	$^*N_2$ (end-on)	-1.1083743	-1.1258556	Training
Tc	@N_4	$^*N_2$ (side-on)	-0.6838642	-1.0680354	Training
Ru	@N_4	$^*N_2$ (end-on)	-1.259592	-1.0026812	Training
Ru	@N_4	$^*N_2$ (side-on)	-0.4130307	-0.5855781	Training
Rh	@N_4	$^*N_2$ (end-on)	-0.1063012	-0.2452938	Training
Rh	@N_4	$^*N_2$ (side-on)	-0.0290651	-0.045172	Training
Mo	@N_4	$^*N_2$ (end-on)	-1.2321619	-1.080281	Training
Mo	@N_4	$^*N_2$ (side-on)	-1.2849525	-1.0531667	Training
Au	@N_4	$^*N_2$ (end-on)	-0.0876914	-0.2276173	Training
Au	@N_4	$^*N_2$ (side-on)	-0.084375	-0.0706088	Training
Ag	@N_4	$^*N_2$ (end-on)	-0.5248738	-0.5519001	Training
Ag	@N_4	$^*N_2$ (side-on)	-0.5294164	-0.5769579	Training
Ir	@N_4	$^*N_2$ (end-on)	0.24552688	-0.1002983	Training
Ir	@N_4	$^*N_2$ (side-on)	-0.0778002	-0.072508	Training
Pt	@N_4	$^*N_2$ (end-on)	-0.0936812	-0.1832603	Training
Pt	@N_4	$^*N_2$ (side-on)	-0.0844344	-0.0895786	Training
Pd	@N_4	$^*N_2$ (side-on)	-0.0403662	0.1876369	Training

Reference

1. C. Francois, *Deep learning with Python*, Manning Publications Company, 2017.