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

Local Descriptors-Based Machine Learning Model Refined by Cluster Analysis for Accurately Predicting Adsorption Energies on Bimetallic Alloys

A. F. Usuga[a], C. S. Praveen[b,c] and A. Comas-Vives[a,d]

[a] Department of Chemistry, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Catalonia, Spain

[b] International School of Photonics, Cochin University of Science and Technology, University Road, South Kalamassery, Kalamassery, Ernakulam, Kerala, 682022, India

[c] Inter University Centre for Nano Materials and Devices, Cochin University of Science and Technology, University Road, South Kalamassery, Kalamassery, Ernakulam, Kerala, 682022, India

[d] Institute of Materials Chemistry, TU Wien, 1060 Vienna, Austria

E-mail: aleix.comas@tuwien.ac.at; aleix.comas@uab.cat

Table S1. Summary of the performance of the evaluated ML-based models. Mean accuracy and standard deviation for 10 splits on the K-fold methodology.

ML-based model	Parameters
XGBoost	• n_estimators=900 • max_depth=5 • reg_lambda=0 • reg_alpha=1.2 • learning_rate=0.12 • objective='reg:squarederror' • min_child_weight=6 • gamma=0
CatBoost	• n_estimators=1300 • learning_rate=0.12 • depth=6 • loss_function='LogCosh' • silent=True • min_data_in_leaf=0 • L2_leaf_reg=1.4
LightGBM	• n_estimators=500 • learning_rate=0.2 • max_depth=5 • boosting_type='gbdt' • num_leaves=(2**5) • verbosity=-1 • min_data_in_leaf=2 • lambda_L2=0.8
Kernel Ridge	• alpha=0.01 • kernel='laplacian' • gamma=None
Linear Regression	-

Table S2. Summary of the proposed descriptors

Descriptor	Meaning
Geometric features for the surface	
'atoms_surf'	Numerical stratification to encode the adsorption site top (1), bridge(2), and hollow (3).
'CN'	Coordination number from the adsorption site.
'gCN'	General coordination number from the adsorption site.
'CN_max'	Maximum coordination number from the adsorption site.
Electronic features for the surface	
'FermiEnergy_surf'	Surface's Fermi Energy.
'Total_charge_surf'	Average of the total charge between the labeled atoms on the surface.
's_charge_surf'	Average of the s-partial orbital charge between the labeled atoms on the surface
'p_charge_surf'	Average of the p-partial orbital charge between the labeled atoms on the surface.
'd_charge_surf'	Average of the d-partial orbital charge between the labeled atoms on the surface.
'd_center'	Surface's d-band center.
'WorkFunction'	Surface's work function.
'stm_surf'	Electronic density at the adsorption site.
Atomic/elemental features for the surface	
'AtomicMass_surf'	Average of the atomic mass between the labeled atoms on the surface.
'Electronegativity_surf'	Average of the electronegativity between the labeled atoms on the surface.
'FirstIonization_surf'	Average of the first ionization between the labeled atoms on the surface.
'AtomicRadius_surf'	Average of the atomic radius between the labeled atoms on the surface.
'MeltingPoint_surf'	Average of the melting point between the labeled atoms on the surface.
'BoilingPoint_surf'	Average of the boiling point between the labeled atoms on the surface.
'NumberofShells_surf'	Average of the number of shells between the labeled atoms on the surface.
Geometric features for the adsorbate	
'H_ads'	Number of hydrogen bonds in the adsorbate.
Electronic features for the adsorbate	
'HOMO_ads'	HOMO energy of the adsorbate
'Total_charge_ads'	Total charge of the main bonded atom from the adsorbate with the surface.
's_charge_ads'	s-partial orbital charge of the main bonded atom from the adsorbate with the surface.
'p_charge_ads'	p-partial orbital charge of the main bonded atom from the adsorbate with the surface.
'stm_ads'	Electronic density at 1.5 Å of the main bonded atom from the adsorbate with the surface.
Atomic/elemental features for the adsorbate	
'AtomicMass_ads'	The atomic mass of the main bonded atom from the adsorbate with the surface.
'Electronegativity_ads'	The electronegativity of the main bonded atom from the adsorbate with the surface.
'Group_ads'	Group of the main bonded atom from the adsorbate with the surface.
'FirstIonization_ads'	First ionization of the main bonded atom from the adsorbate with the surface.
'AtomicRadius_ads'	Atomic radius of the main bonded atom from the adsorbate with the surface.
'MeltingPoint_ads'	The melting point of the main bonded atom from the adsorbate with the surface.
'BoilingPoint_ads'	The boiling point of the main bonded atom from the adsorbate with the surface.
'SpecificHeat_ads'	Specific heat of the main bonded atom from the adsorbate with the surface.

'NumberofShells_ads'	Number of shells of the main bonded atom from the adsorbate with the surface.

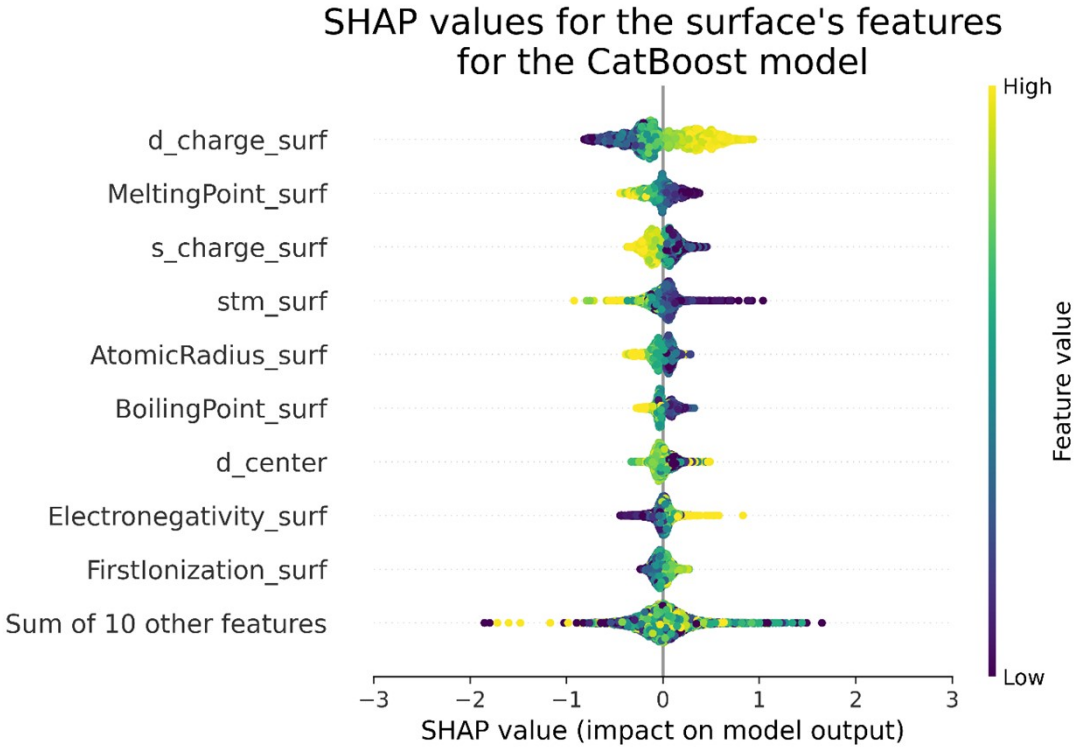

Figure S1. Summary of the entire distribution of SHAP values for the surface-related descriptors with the CatBoost model.

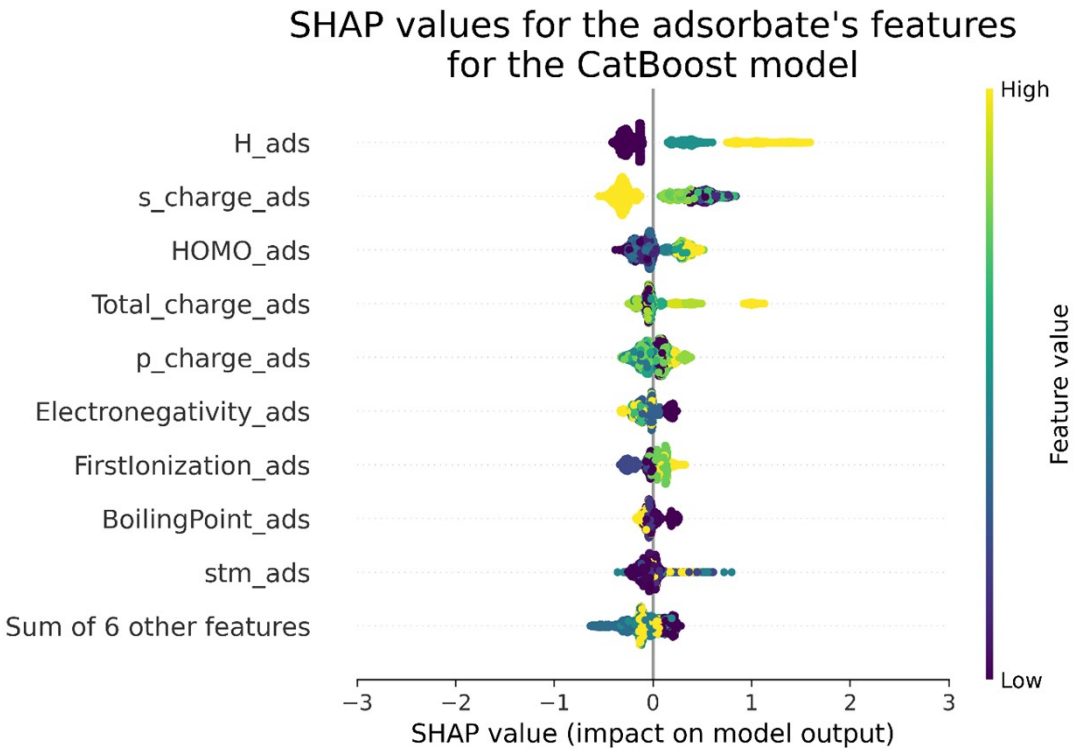

Figure S2. Summary of the entire distribution of SHAP values for the adsorbate-related descriptors with the CatBoost model.

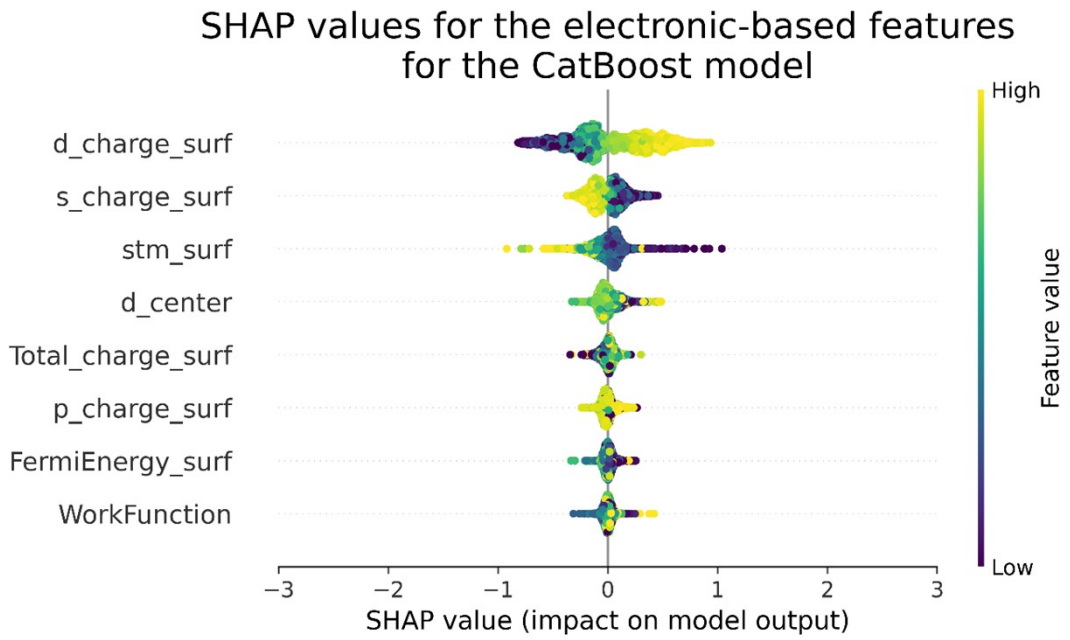

Figure S3. Summary of the entire distribution of SHAP values for the electronical properties at the surface-related descriptors with the CatBoost model.

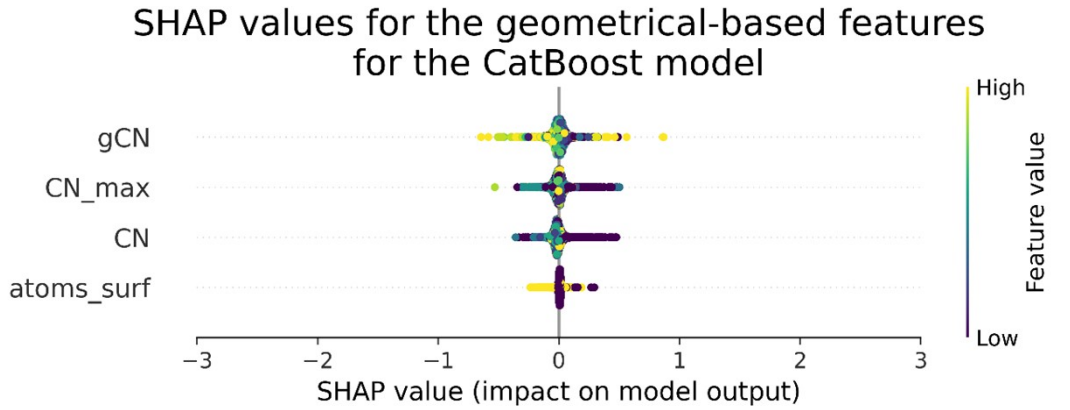

Figure S4. Summary of the entire distribution of SHAP values for the geometrical information at the surface-related descriptors with the CatBoost model.

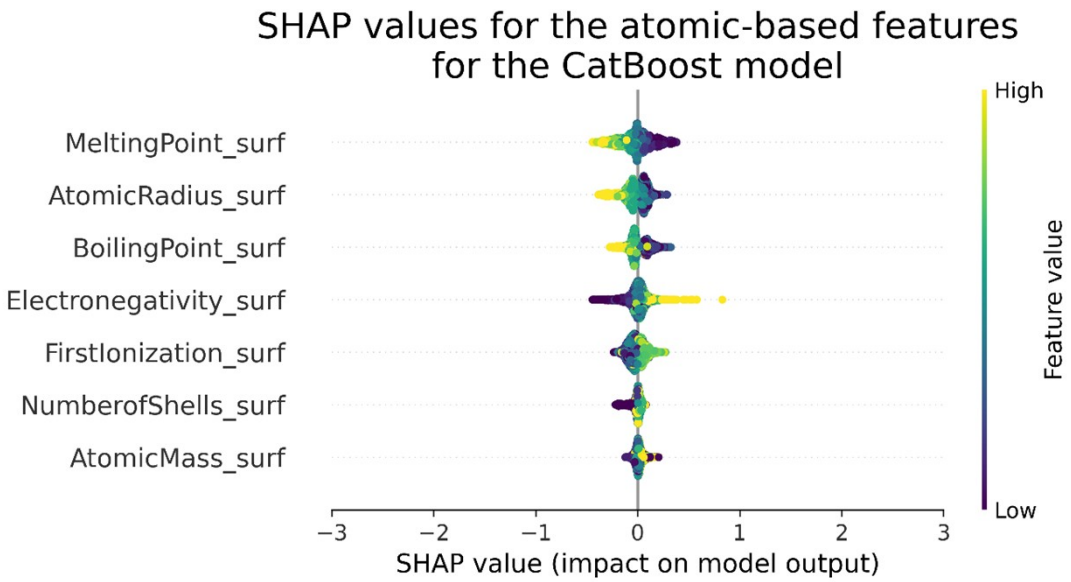

Figure S5. Summary of the entire distribution of SHAP values for the atomical information at the surface-related descriptors with the CatBoost model.

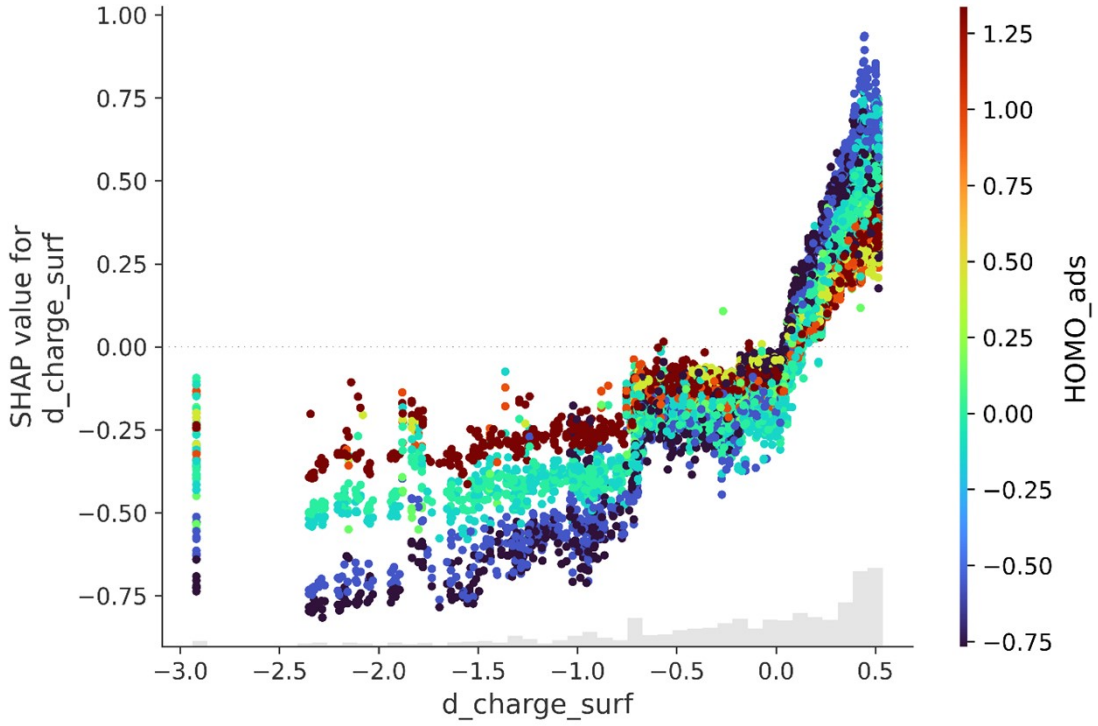

Figure S6. Dependence scatter plot for the d_charge_surf (the d-partial orbital charge for the labeled atoms at the surface) feature with the HOMO_ads feature (HOMO values for each main bonded atom in the adsorbate to the surface) based in the SHAP values for the CatBoost model.

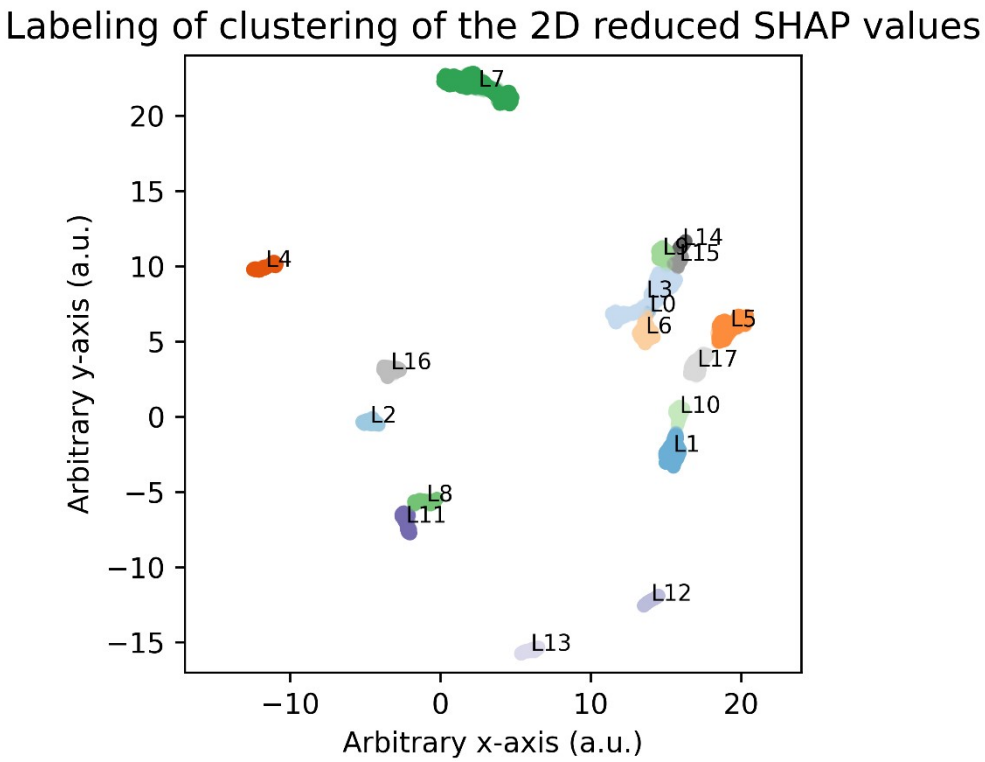

Figure S7. Labeling of the clustering for the 2D reduced dimension visualization of SHAP values using the CatBoost model.

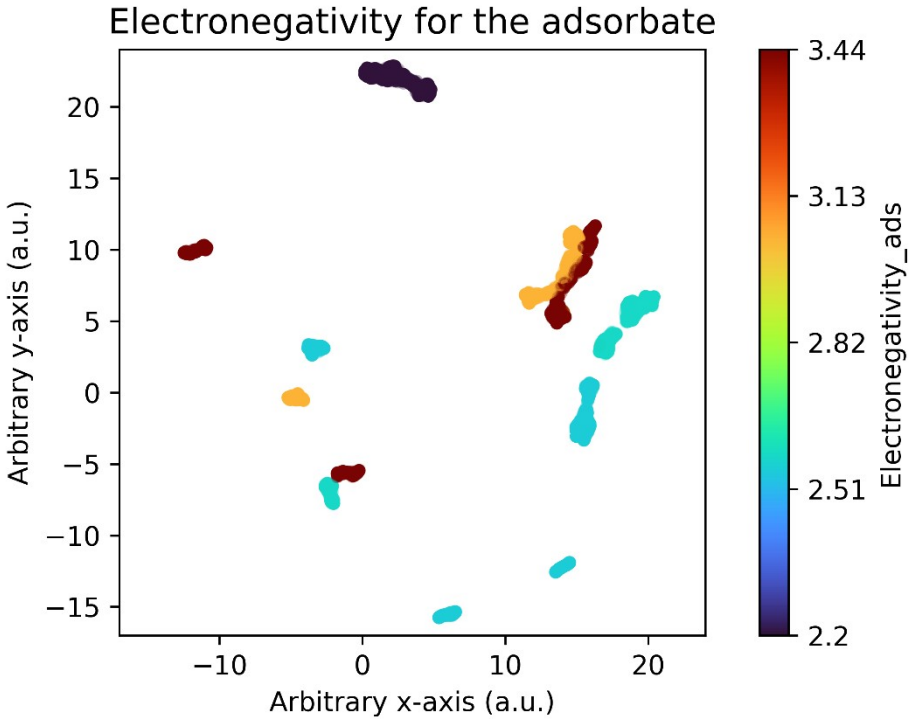

Figure S8. 2D reduced dimension visualization of SHAP values using the CatBoost model for the electronegativity of the main atom in the adsorbate bonded at the surface.

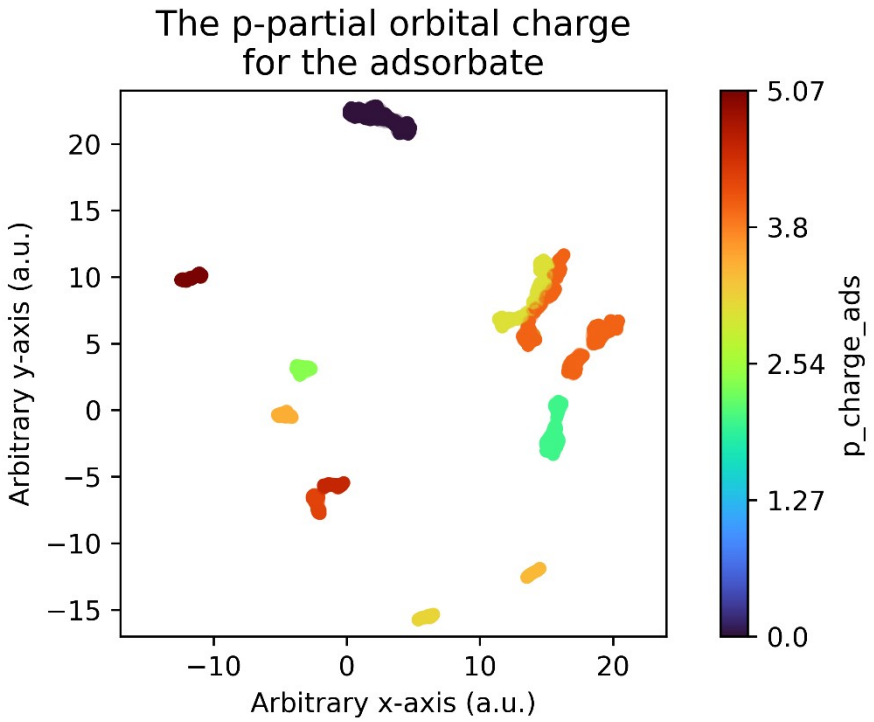

Figure S9. 2D reduced dimension visualization of SHAP values using the CatBoost model for the p-partial orbital charge of the main atom in the adsorbate bonded at the surface.

Table S3. Summary of the performance of the evaluated ML-based models. Mean accuracy and standard deviation for 10 splits on the K-fold methodology.

Adsorbate	H_ads (no. bonds between hydrogen and main elements)	HOMO_ads [eV] (HOMO values for adsorbates)
H	0	-3,94
CH_2	2	-4,55
CH_3	3	-4,61
H_2O	2	-5,44
NH	1	-5,92
CH	1	-6,16
S	0	-6,16
SH	1	-6,31
C	0	-6,38
N	0	-7,10
OH	1	-7,41
O	0	-7,43