



Surface Binding Energy Prediction using Machine Learning Models

Under Guidance of
Prof. Ali Haider
Dr. Tuhin Khan

Shivanshu Verma
2015CH70186



Why Machine Learning ?

- According to Sabatier Principle, binding energy is a good descriptor for catalytic activity.
- Using DFT calculations to find adsorbate binding energies on the surface takes a lot of time and computation even on HPC.
- Machine learning can predict the results based on previous results training in few minutes. More the data, more accurate the results.

Extracted Data *(88, 587 reactions)*



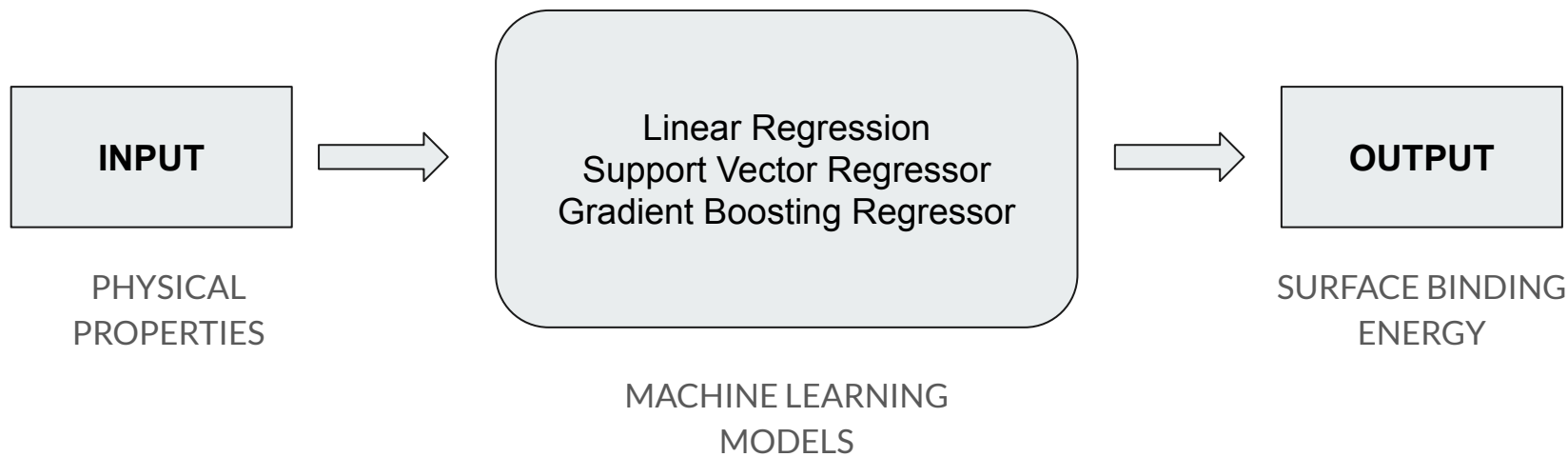
Equation	Chemical Composition	Facet	Reaction Energy
$0.5\text{N}_2(\text{g}) + * \rightarrow \text{N}^*$	Sc3V9	111	-2.810391181
$0.5\text{H}_2(\text{g}) + * \rightarrow \text{H}^*$	Pt9Ti3	111	-0.292809865
$\text{CH}_4(\text{g}) - 2.0\text{H}_2(\text{g}) + * \rightarrow \text{C}^*$	Ag6Pd6	211	4.541566089

Filtered Data

Equation	Chemical Composition	Facet	Reaction Energy
$\text{CH}_4(\text{g}) + 5* \rightarrow \text{C}^* + 4\text{H}^*$	Au3Cu	111	5.805189557
$\text{CH}_4(\text{g}) + 5* \rightarrow \text{C}^* + 4\text{H}^*$	Au3Pb	111	5.419068456
$\text{CH}_4(\text{g}) + 5* \rightarrow \text{C}^* + 4\text{H}^*$	Au3Pd	111	4.925831102

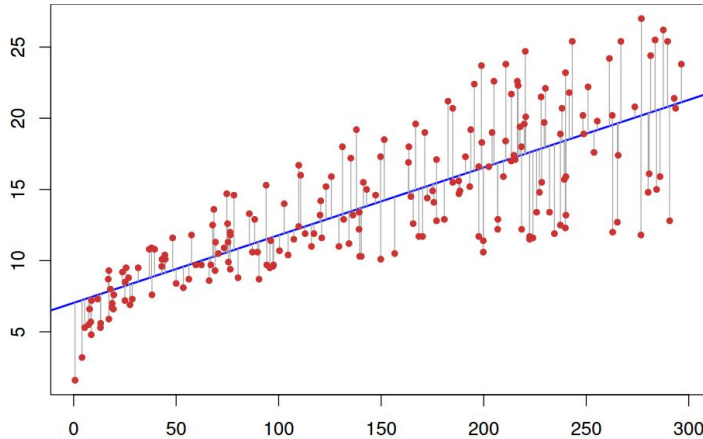
AIM

To predict carbon binding energy on 111 surfaces using catalyst atoms' properties as feature vector.

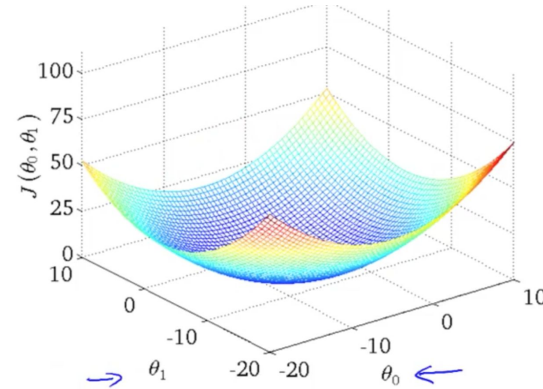


Feature Vector [AN, AM, group, period, radius, EN, MP, BP, H_FUS, density, IE, SE, same for second atom]

Linear Regression



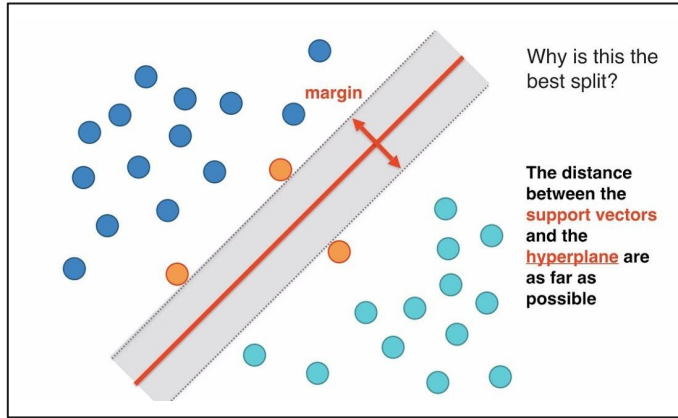
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$



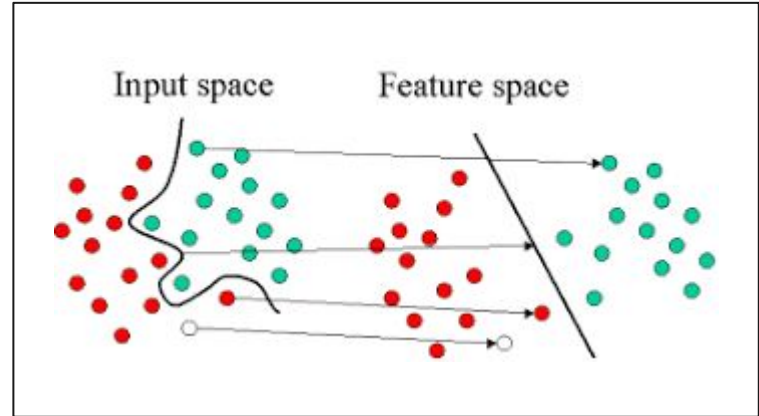
Gradient Descent
Algorithm

```
Percentage Train Error 21.53735417606063 % with 8 outlier points
Percentage Test Error 30.934410272375384 % with 3 outlier points
RMSE Test_error: 0.8875567696515375 eV
RMSE Train_error: 0.5520982254578807 eV
R square score for test data is 0.8429051804630732
R square score for train data is 0.9235168915460406
```

Support Vector Regression



Support Vectors



Kernels

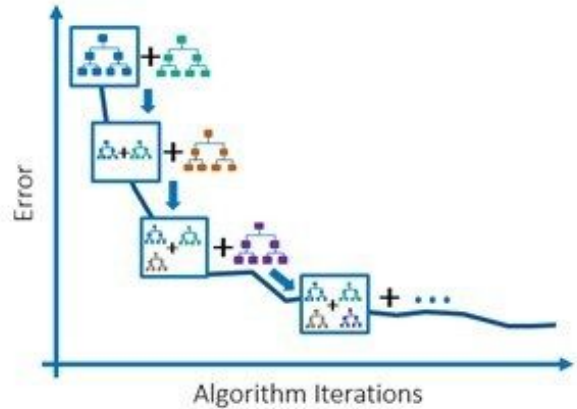
```
Percentage Train Error 19.369857561553 % with 10 of 56 outlier points
Percentage Test Error 23.270042167292488 % with 5 of 15 outlier points
RMSE Test_error: 1.0316915932174389 eV
RMSE Train_error: 0.8094243571324792 eV
R square score for test data is 0.8264718698190099
R square score for train data is 0.8468466755479362
```

Optimised with a set of

[Rbf, poly, linear, sigmoid]

kernels

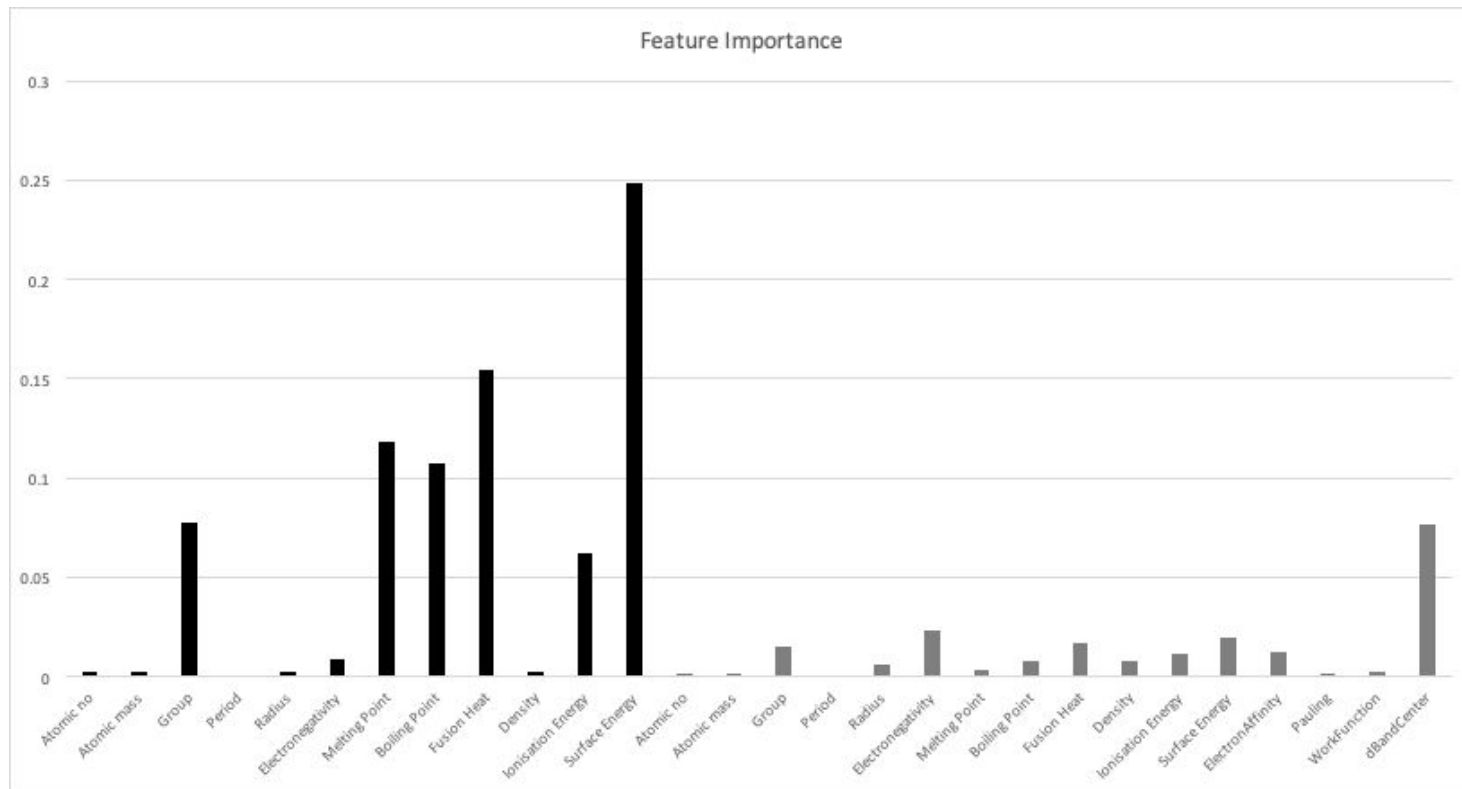
Gradient Boosting Regressor



Gradient boosting adds sub-models incrementally to minimize a loss function

```
Percentage Train Error 8.174628494774993 % with 0 of 56 outlier points
Percentage Test Error 5.53857972665733 % with 1 of 15 outlier points
RMSE Test_error: 0.6508331058742077 eV
RMSE Train_error: 0.04295257966966715 eV
R square score for test data is 0.9533205958009017
R square score for train data is 0.999580761529814
```

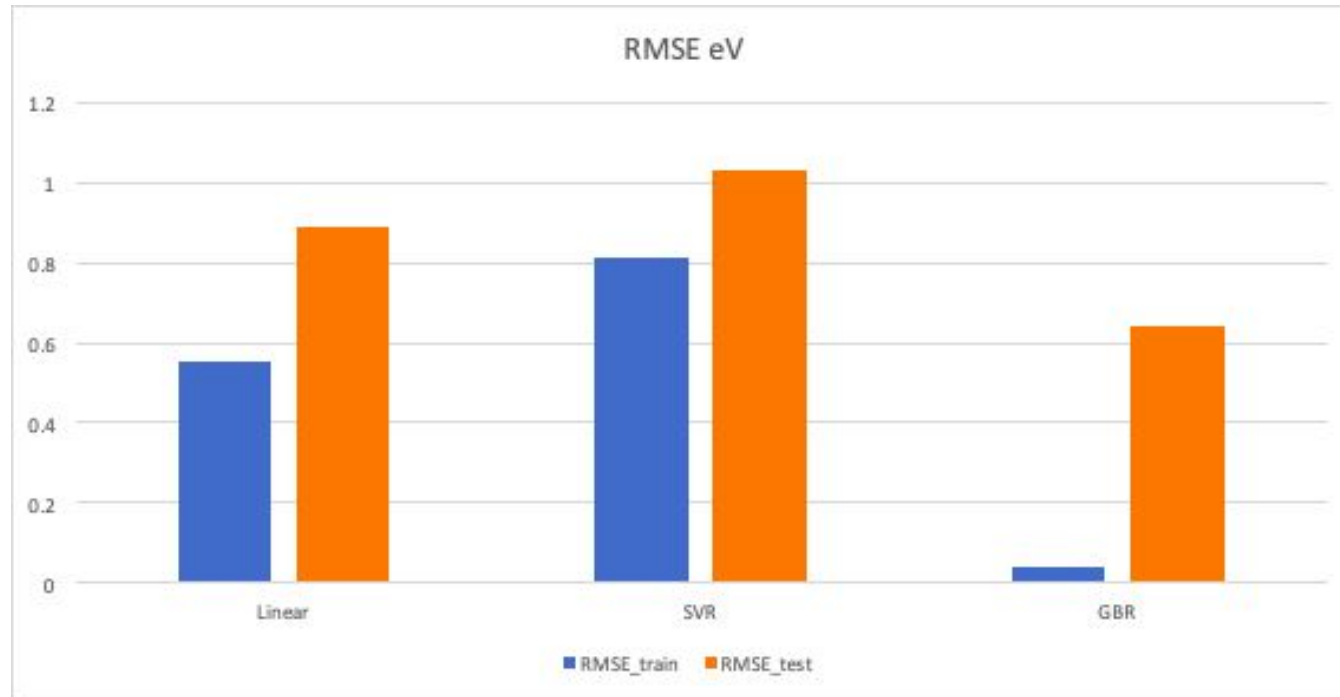
Feature Importance



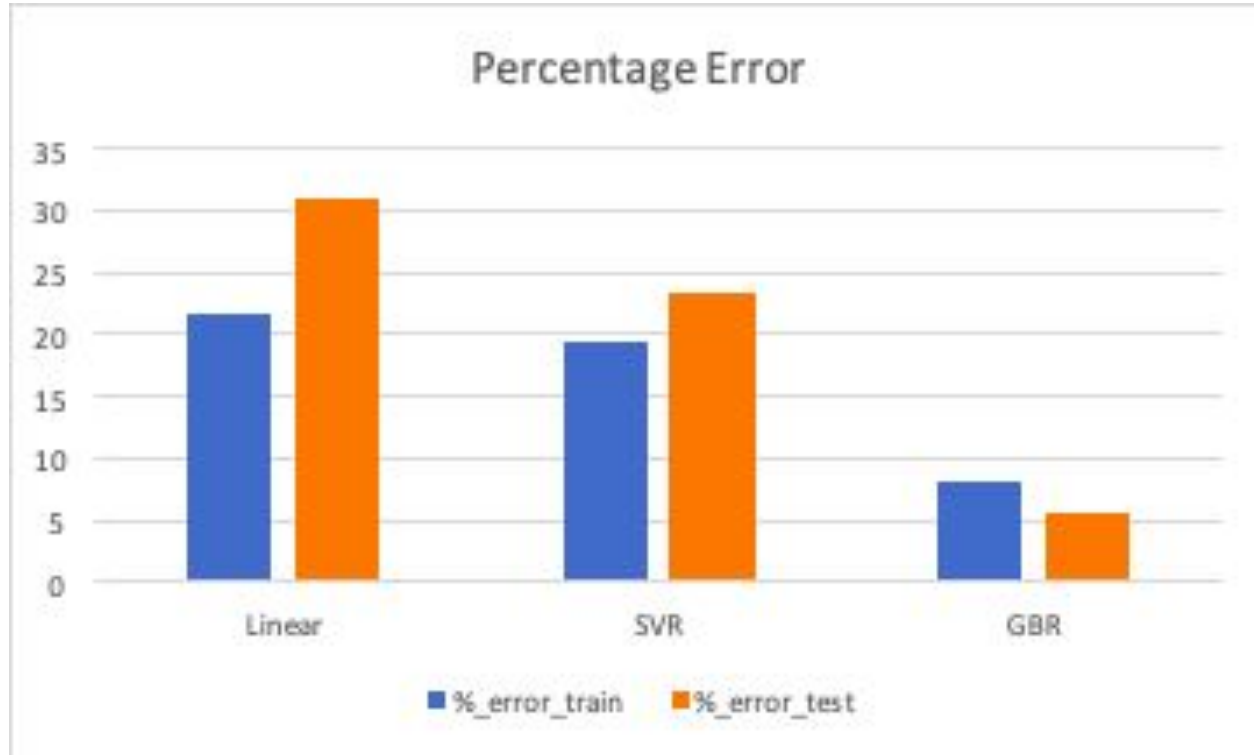
Comparison b/w Models



Comparison b/w Models



Comparison b/w Models





THANK YOU