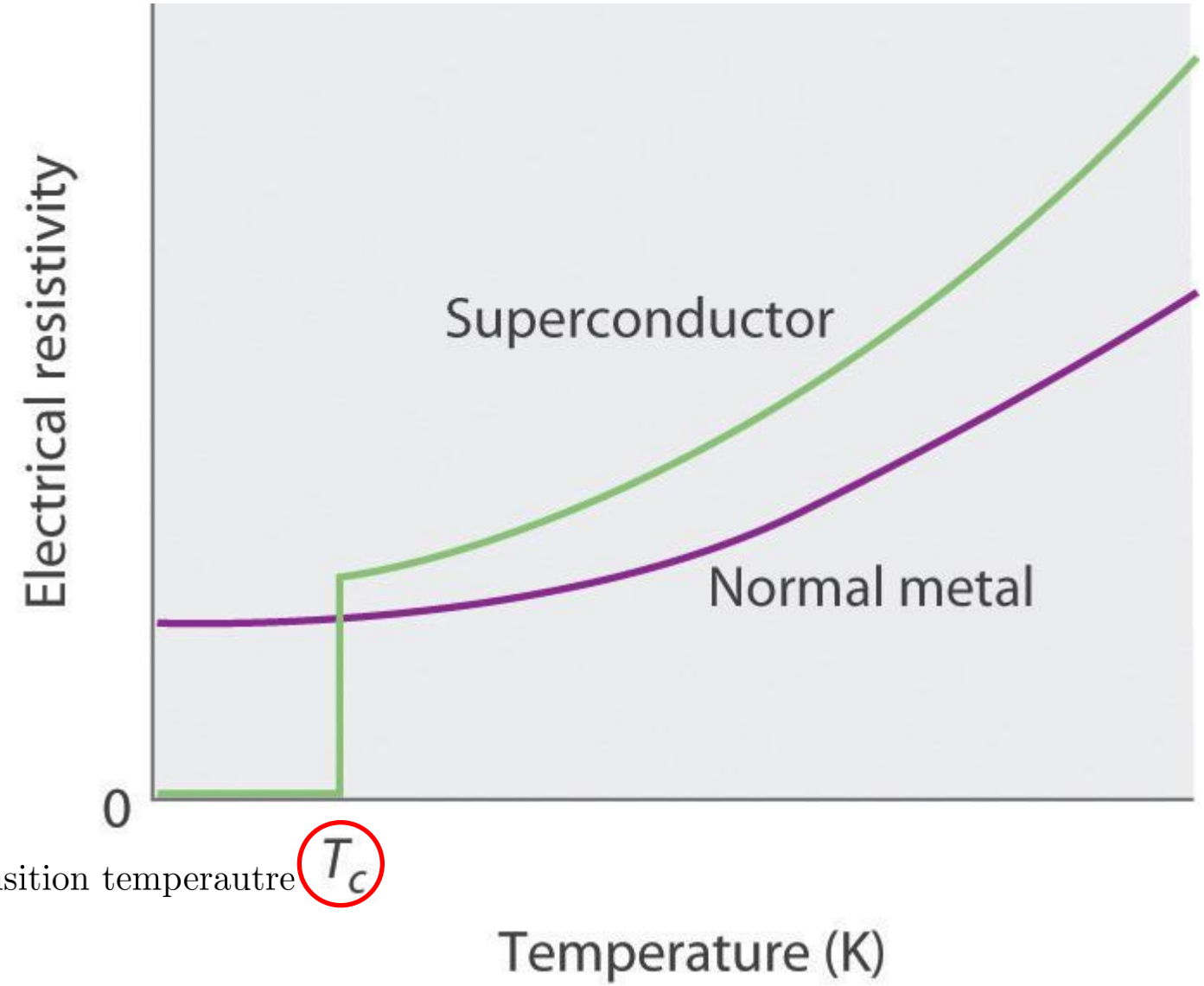
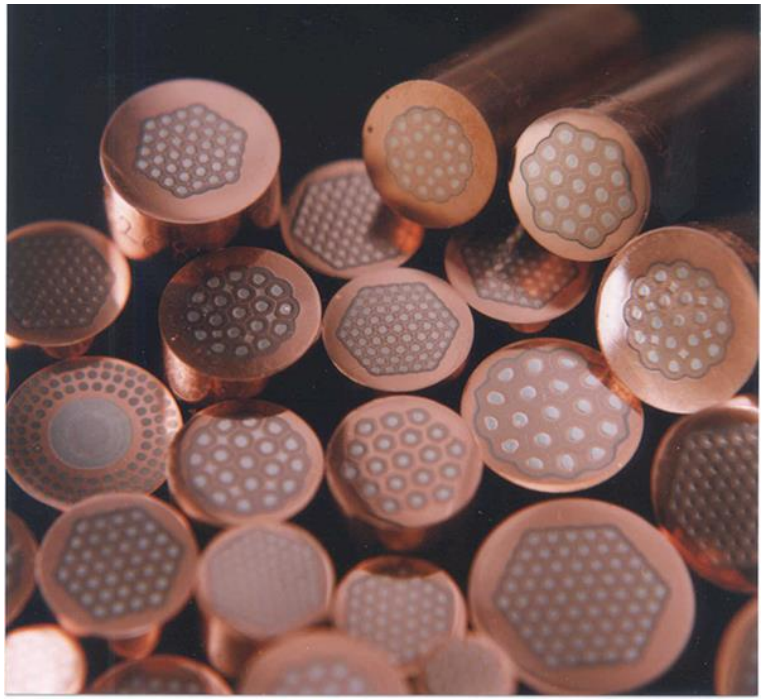
Superconducting Transition Temperature Prediction from Chemical Formula and Elemental Properties

What is a superconductor?



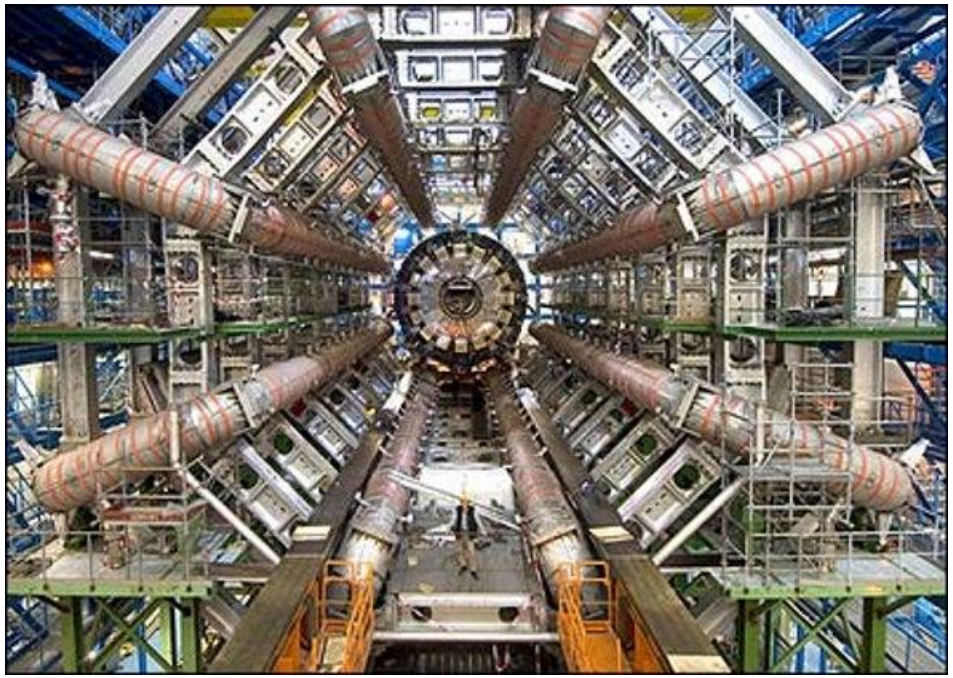
Superconducting transition temperautre

What are applications of superconductors?



Superconducting wire

MRI

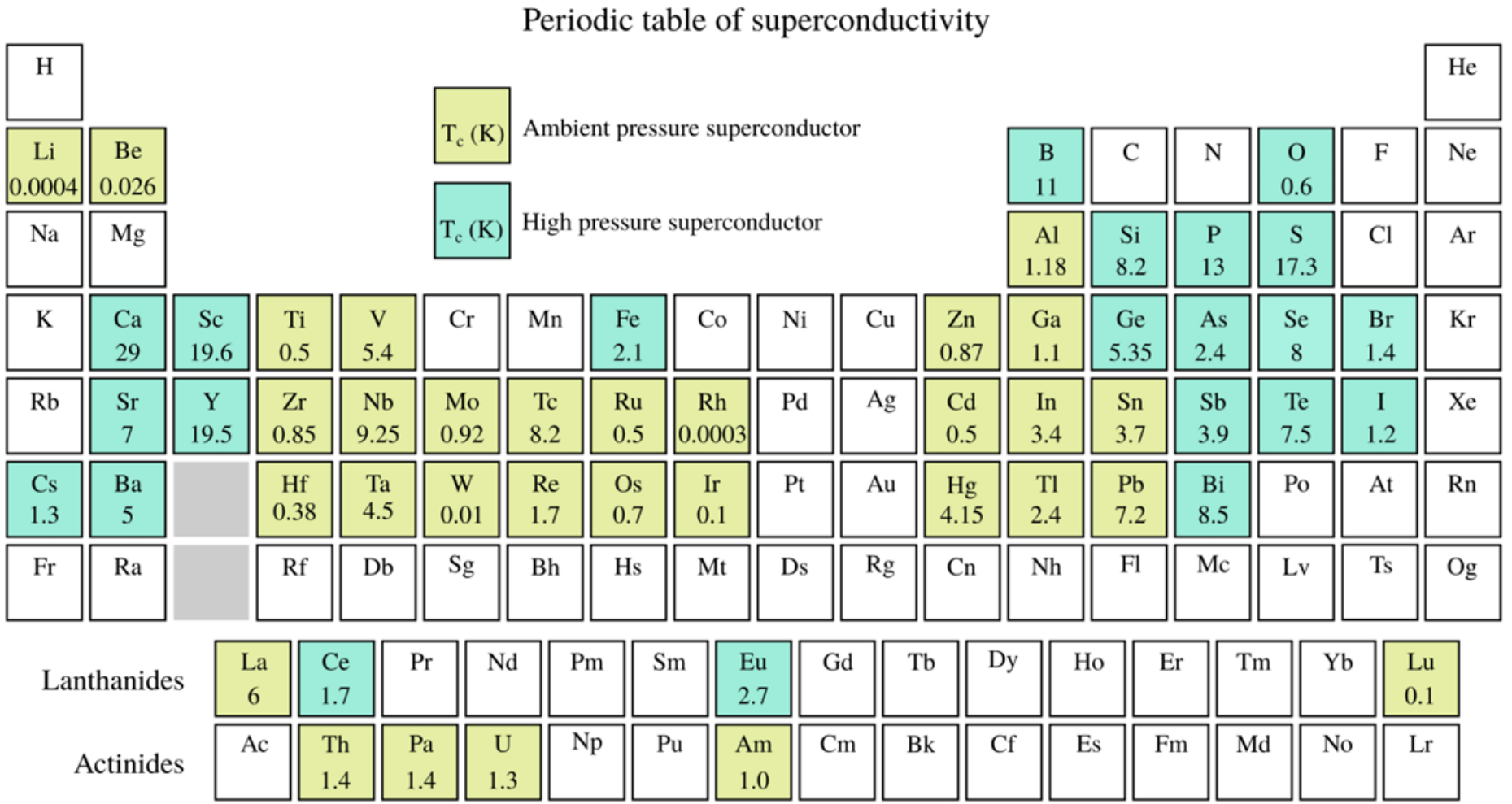


Particle accelerator

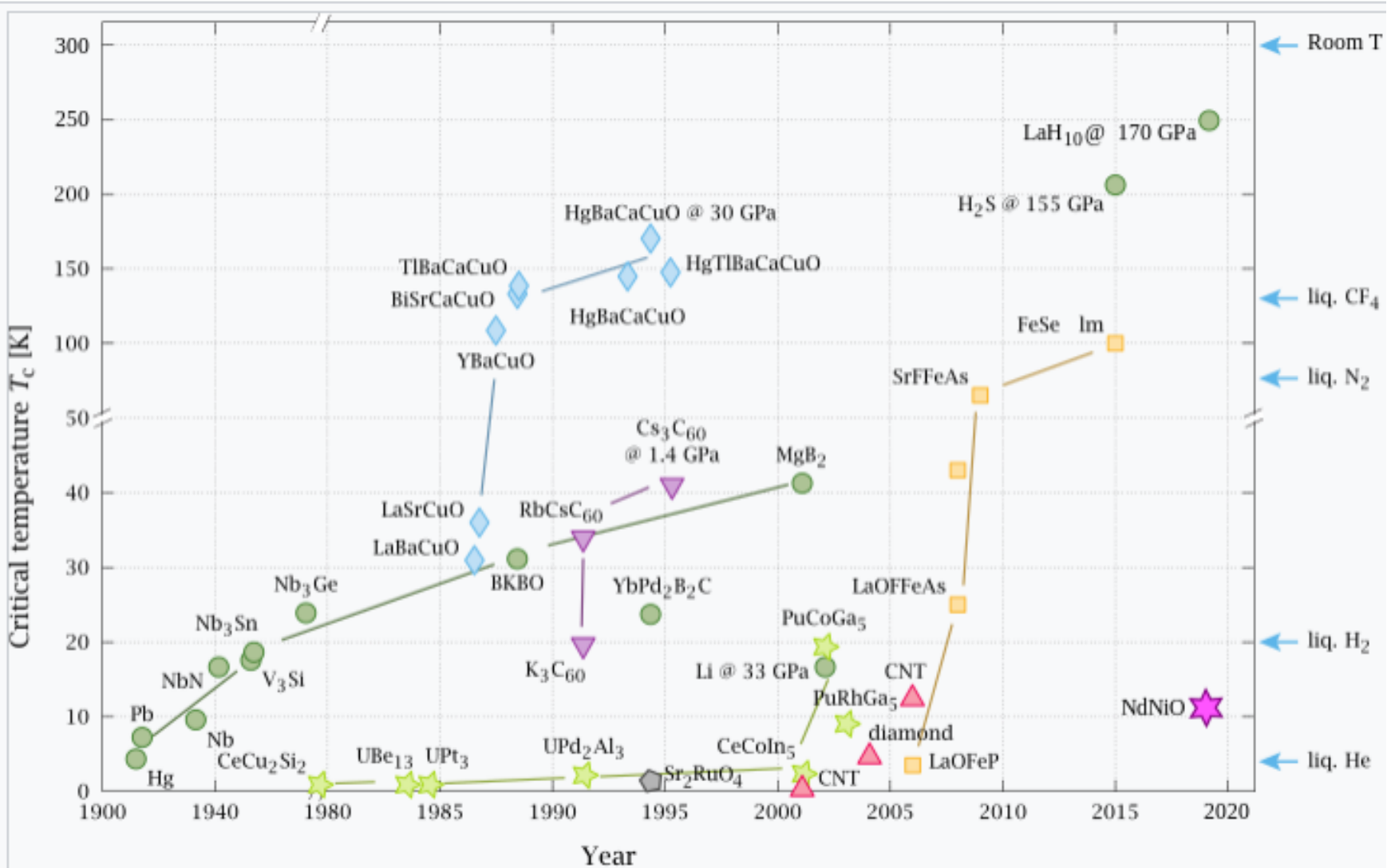
Maglev train

Why don’t we see superconductors everywhere?

- Hard to find materials with high T_c



- Generally, require very high pressure



Most importantly, No general theory that can predict T_c

Data driven approach

Chemical formula	Critical temperature
YBaCuO	93
MgB_2	42
Ni_3Ge	23
⋮	⋮

Element	Properties		
	Atomic mass	Atomic radius	⋯
O	16	66	
Cu	64	132	
Fe	56	132	
⋮			

Elemental properties

Property	Unit	Description
Atomic Mass	Da	Rest mass of an atom
First Ionization Energy	kJ/mol	Energy required to remove a valence electron
Atomic Radius	pm	Calculated atomic radius
Density	kg/m^3	Density at room temperature and ambient pressure
Electron Affinity	kJ/mol	Energy released on formation of anions
Fusion Heat	kJ/mol	Energy required to change from solid to liquid
Thermal Conductivity	W/m·K	Ability to conduct heat
Valence	No units	Typical number of chemical bonds formed by the element

Features

Feature	Chemical formula AB_2	
Mean (μ)	$\frac{p_A+p_B}{2}$	; p_A = property of element A
Weighted mean (ν)	$\frac{1}{3}p_A + \frac{2}{3}p_B$	
Geometric mean	$(p_A p_B)^{\frac{1}{2}}$	
Weighted geometric mean	$p_A^{\frac{1}{3}} p_B^{\frac{2}{3}}$	
Entropy	$-v_A \ln v_A - v_B \ln v_B$	; $v_A = \frac{p_A}{p_A+p_B}$
Weighted entropy	$-z_A \ln z_A - z_B \ln z_B$	; $z_A = \frac{\frac{1}{3}v_A}{\frac{1}{3}v_A+\frac{2}{3}v_B}$
Range	$\lvert p_A - p_B \rvert$	
Weighted range	$\lvert \frac{1}{3}p_A - \frac{2}{3}p_B \rvert$	
Standard deviation	$\sqrt{\frac{(p_A-\mu)^2+(p_B-\mu)^2}{2}}$	
Weighted standard devieation	$\sqrt{\frac{1}{3}(p_A-\nu)^2 + \frac{2}{3}(p_B-\nu)^2}$	
Number of elements	2	

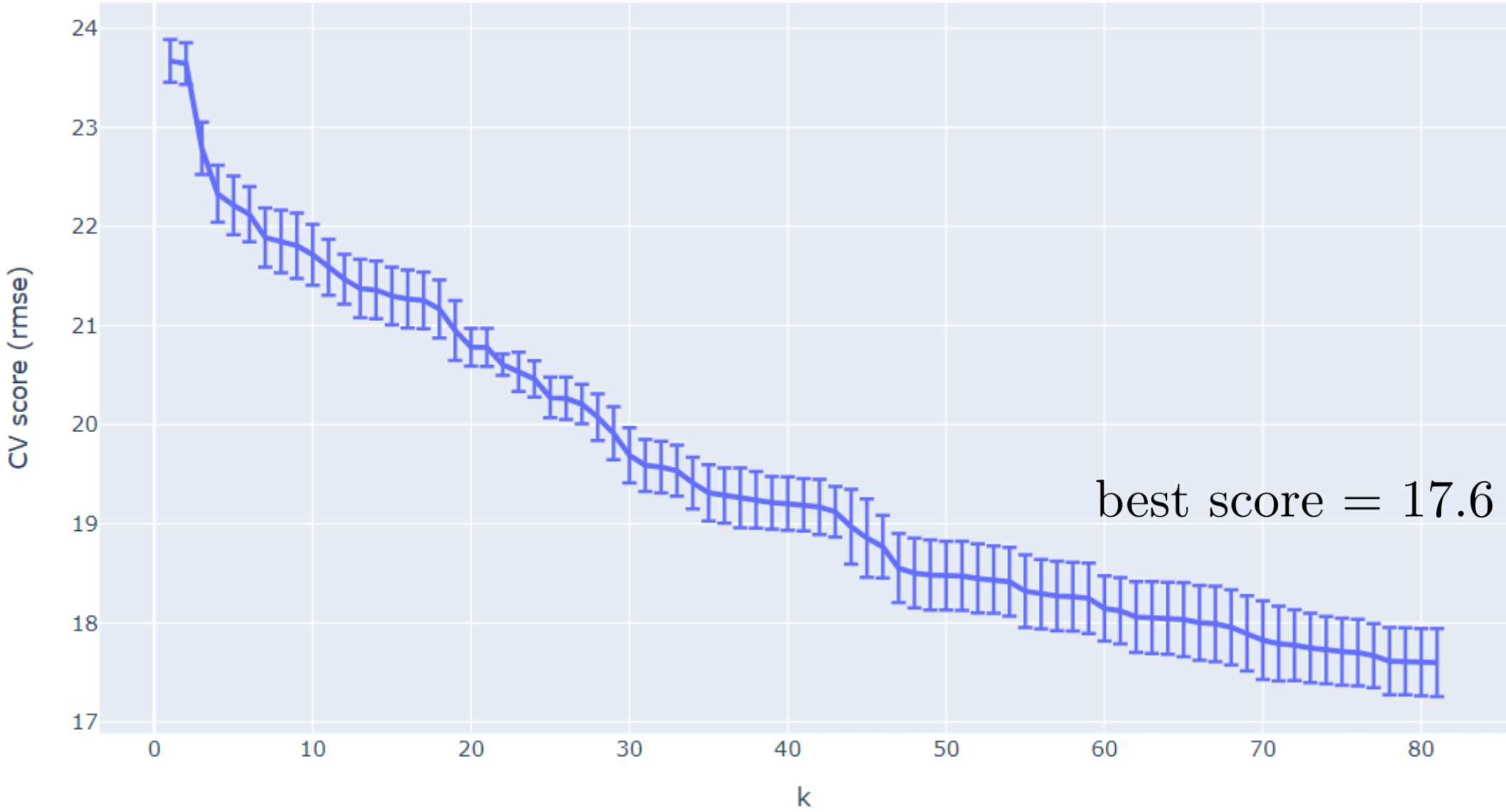
Input

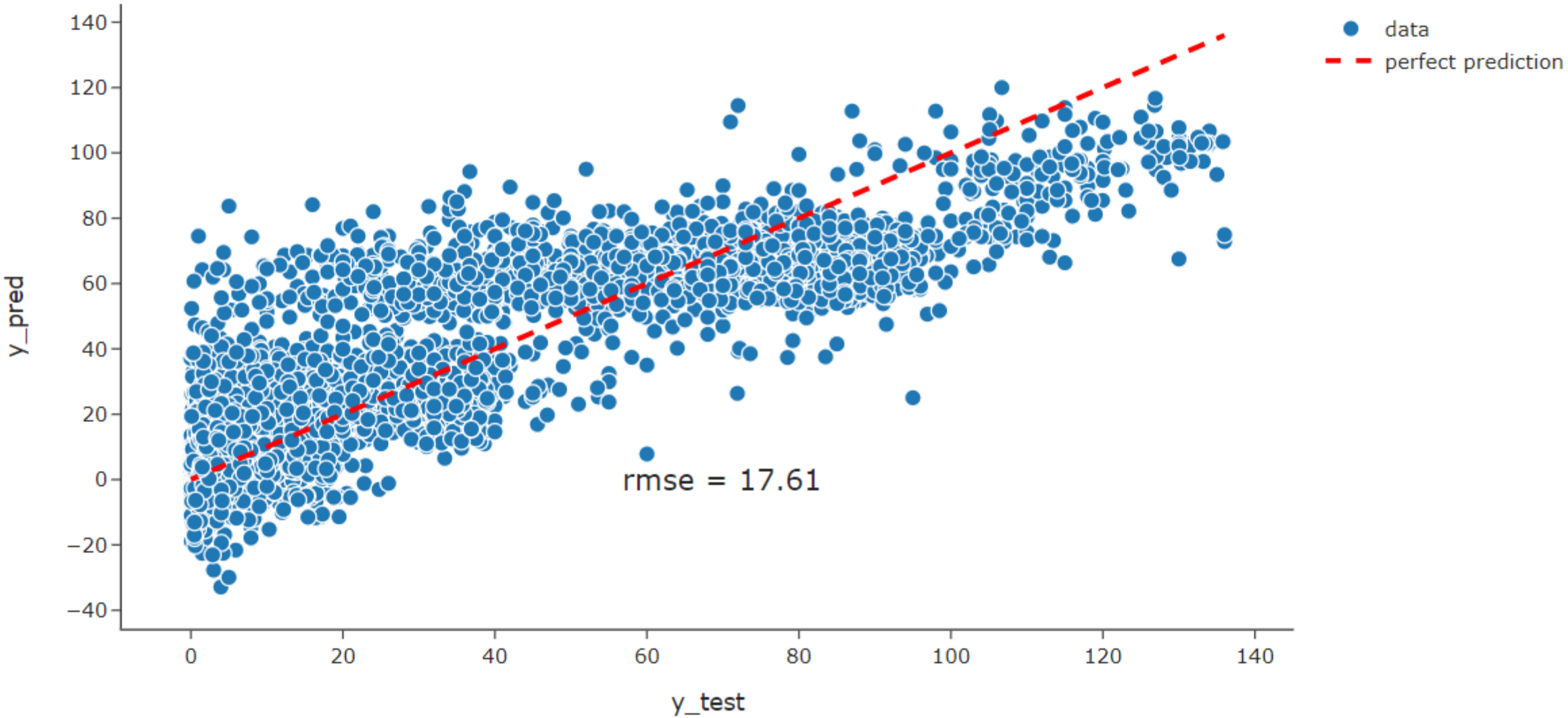
- Repeat the process for each property
- Drop chemical formula
- Number of features
  $= 10$ (properties) $\times$ 8 (aggregate functions) + 1 (# of elements)
  $= 81$
- Number of samples $= 21196$
- Data taken from previous work,
  K. Hamidieh, Computational Materials Science 154 (2018) 346–354

Methodology

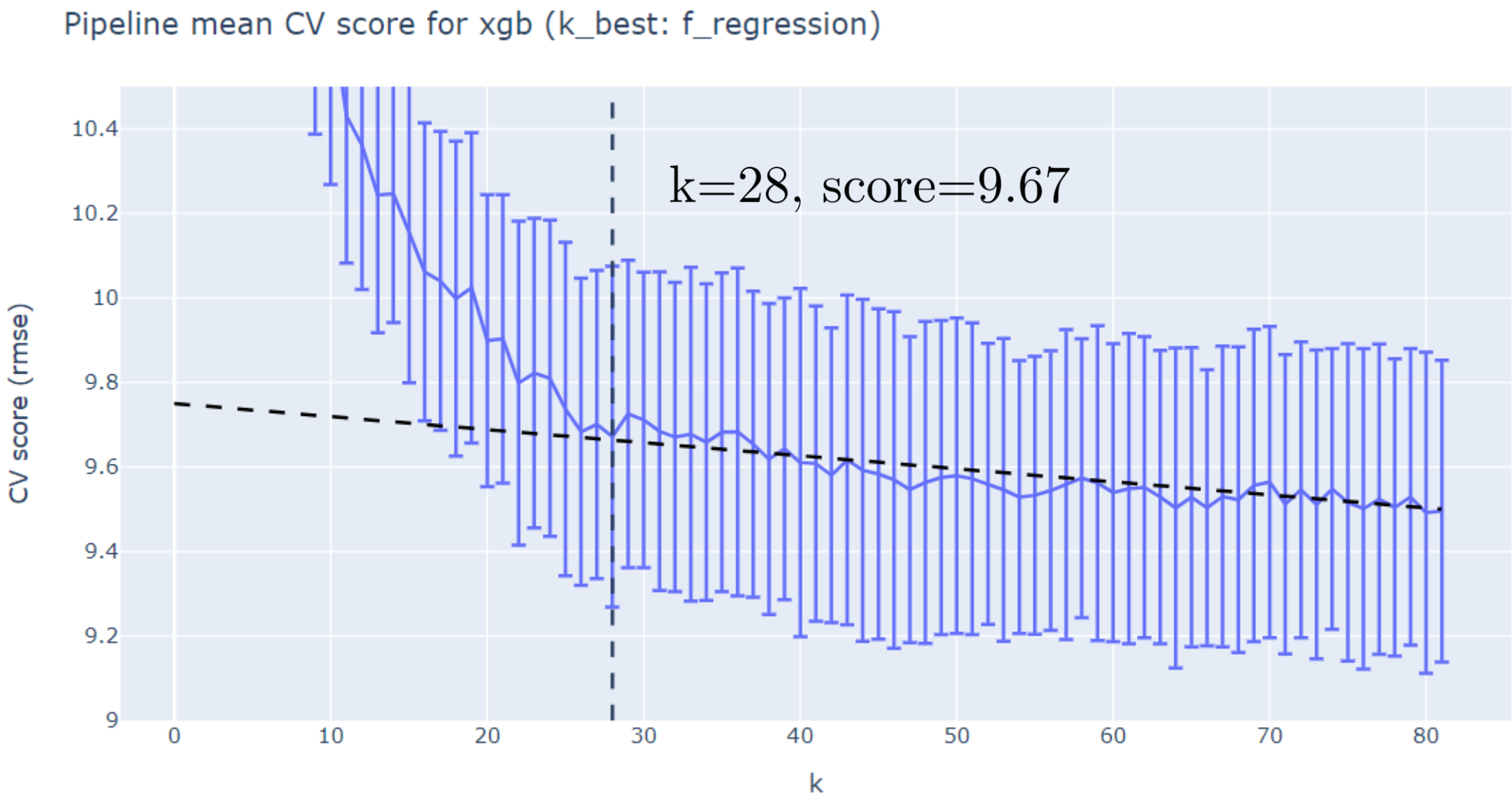
- Metric: Root Mean Squared Error (RMSE)
- Linear regression models
- XGBoost model
  - Bayes hyperoptimization
  - Feature selection
- Compare final result to previous study by K. Hamidieh
  – XGBoost: RMSE = 9.5

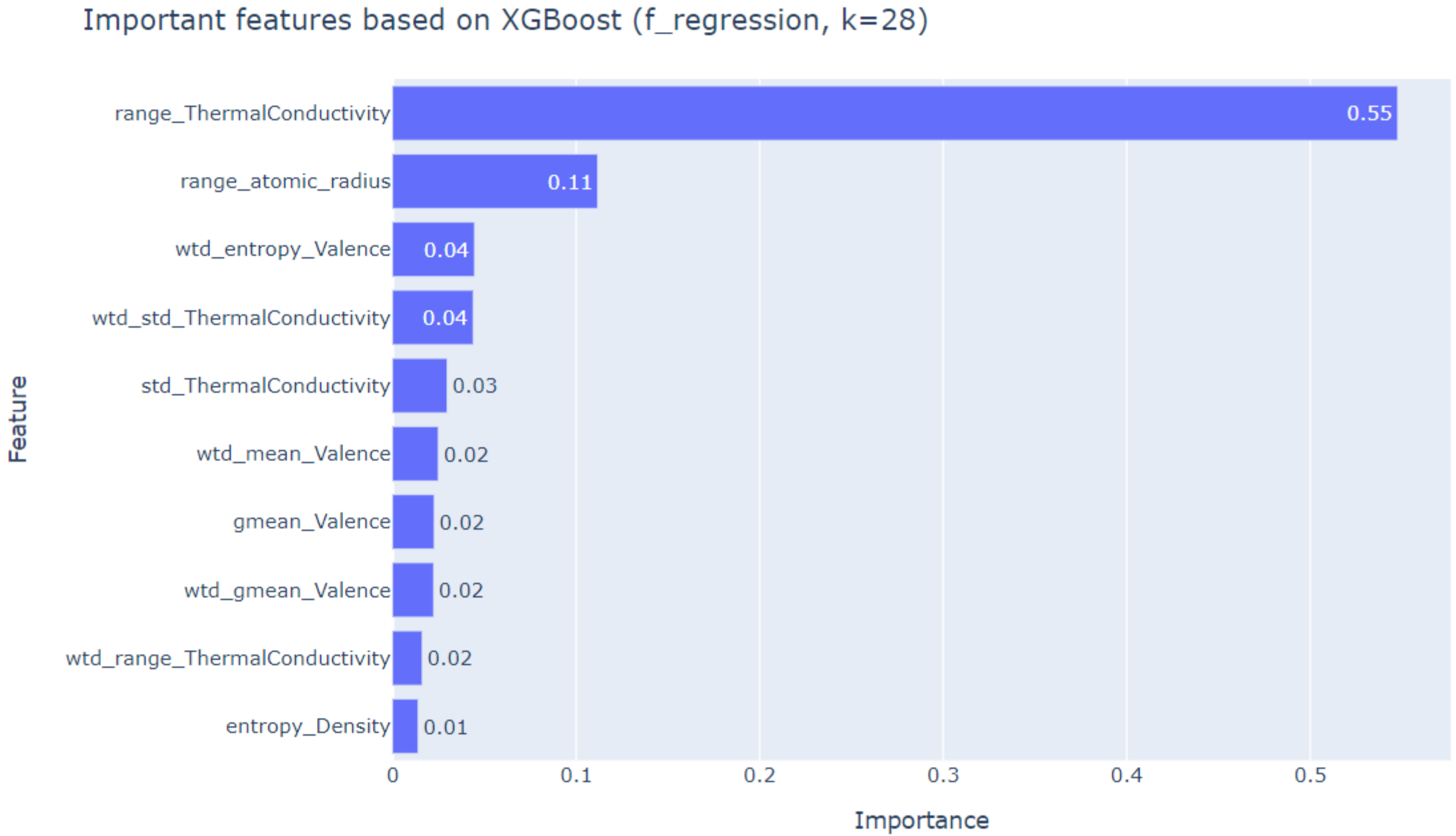
Linear regression (ols, lasso, ridge)



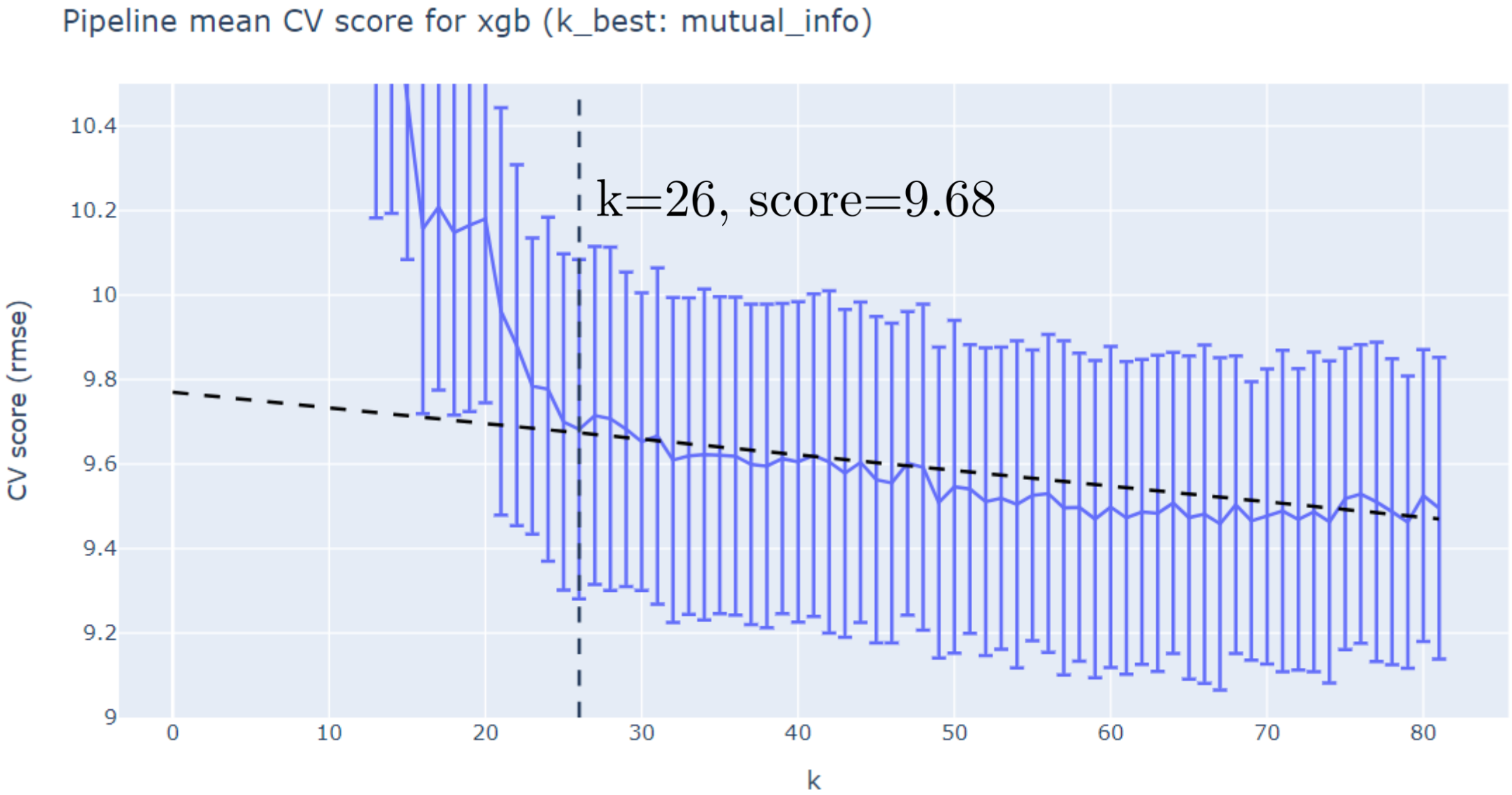

data
perfect prediction
rmse = 17.61
y_pred
y_test
140
120
100
80
60
40
20
0
−20
−40
0
20
40
60
80
100
120
140

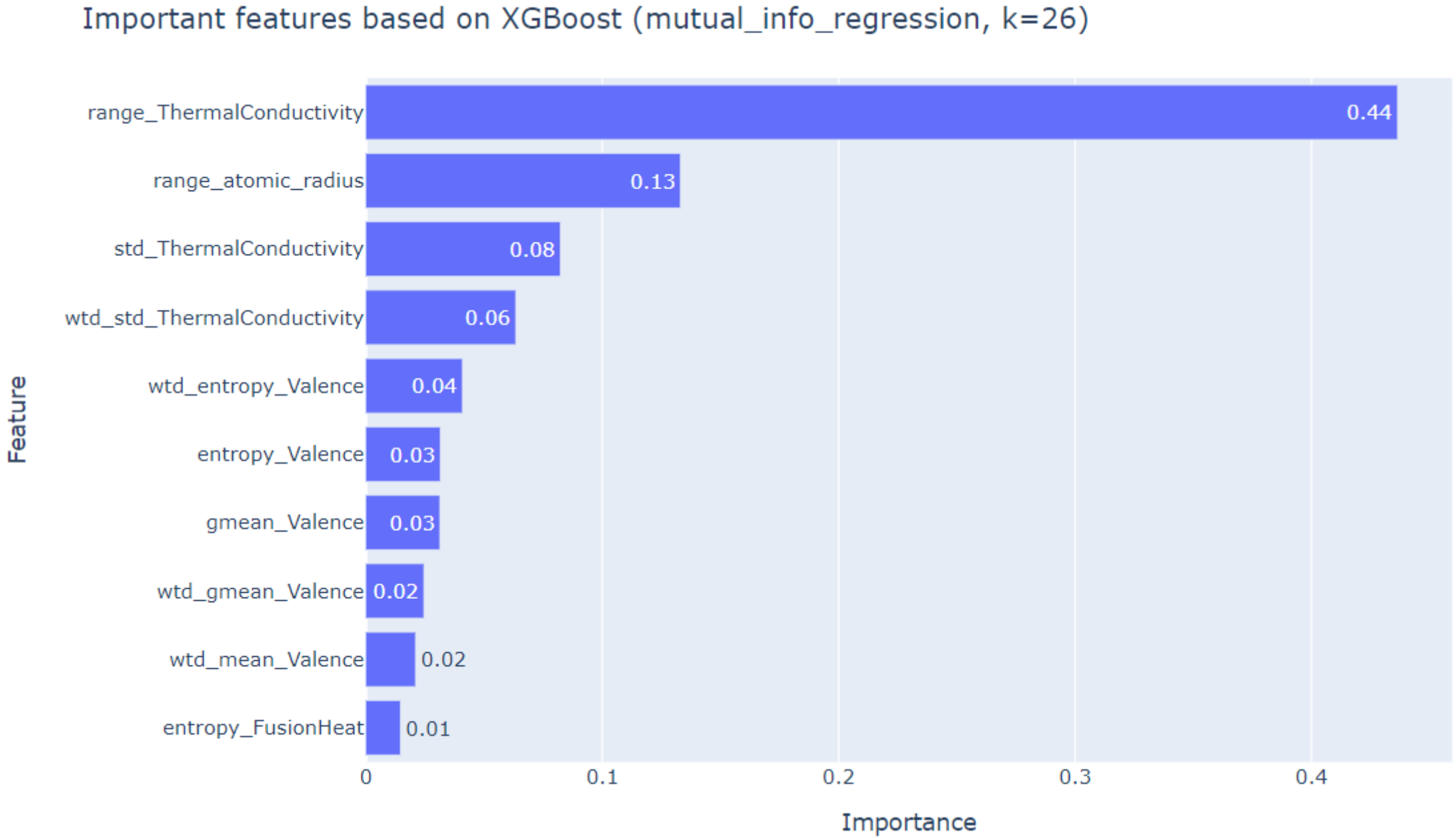
XGBoost


Pipeline mean CV score for xgb (k_best: f_regression)
k=28, score=9.67
CV score (rmse)
10.4
10.2
10
9.8
9.6
9.4
9.2
9
0
10
20
30
40
50
60
70
80
k

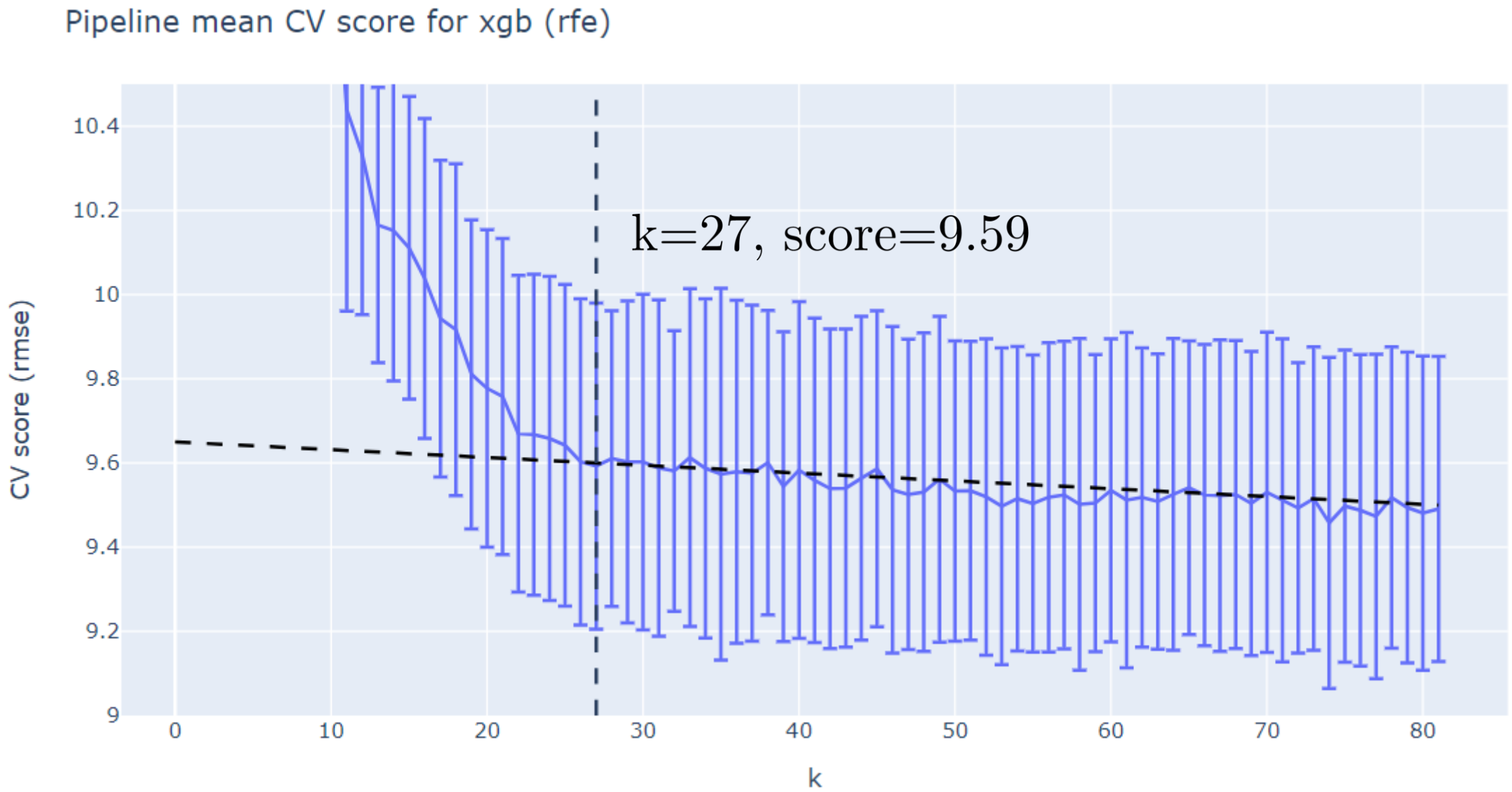

Important features based on XGBoost (f_regression, k=28)
range_ThermalConductivity
range_atomic_radius
wtd_entropy_Valence
wtd_std_ThermalConductivity
std_ThermalConductivity
wtd_mean_Valence
gmean_Valence
wtd_gmean_Valence
wtd_range_ThermalConductivity
entropy_Density
0.55
0.11
0.04
0.04
0.03
0.02
0.02
0.02
0.02
0.01
Feature
0
0.1
0.2
0.3
0.4
0.5
Importance

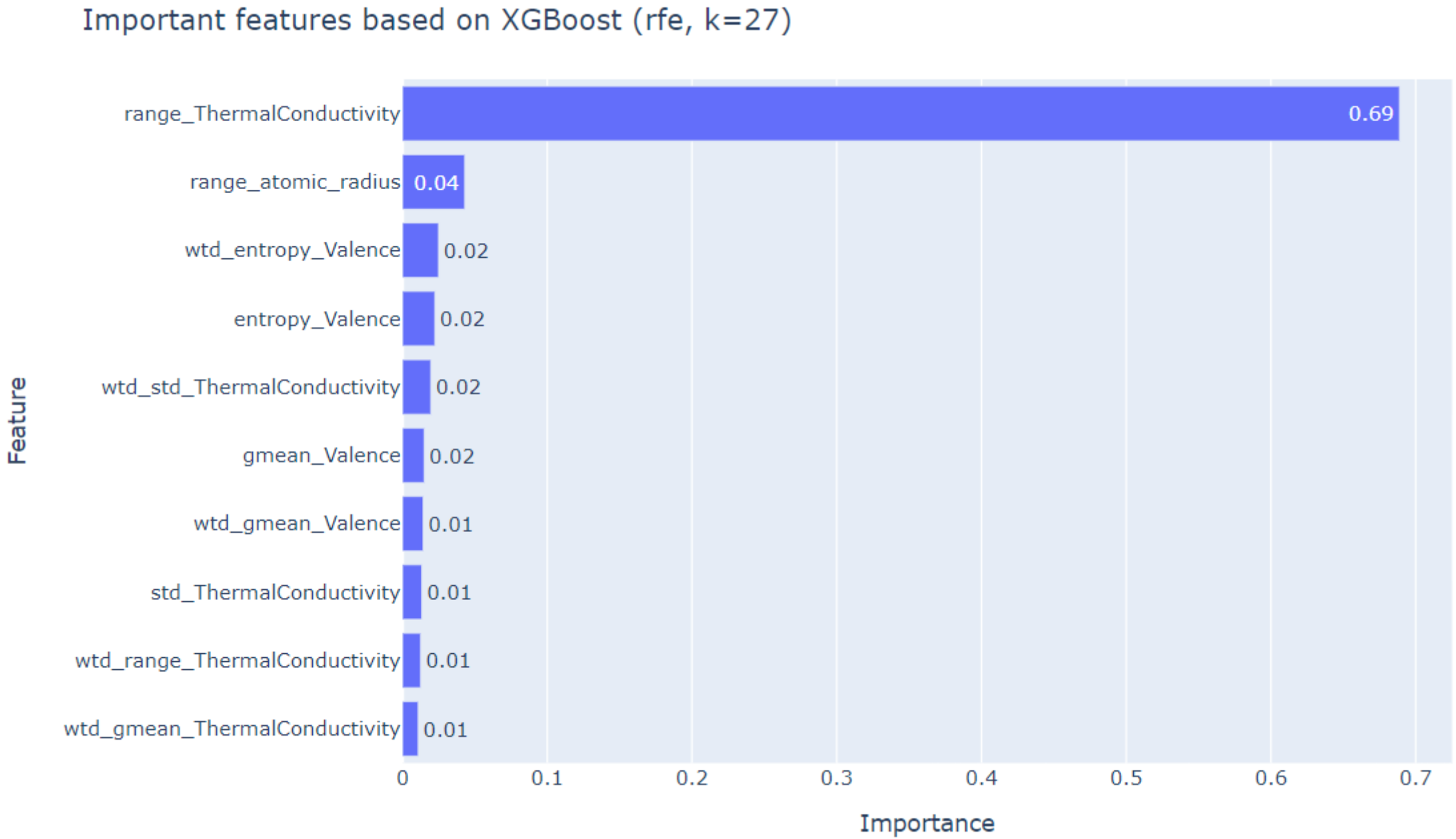
XGBoost

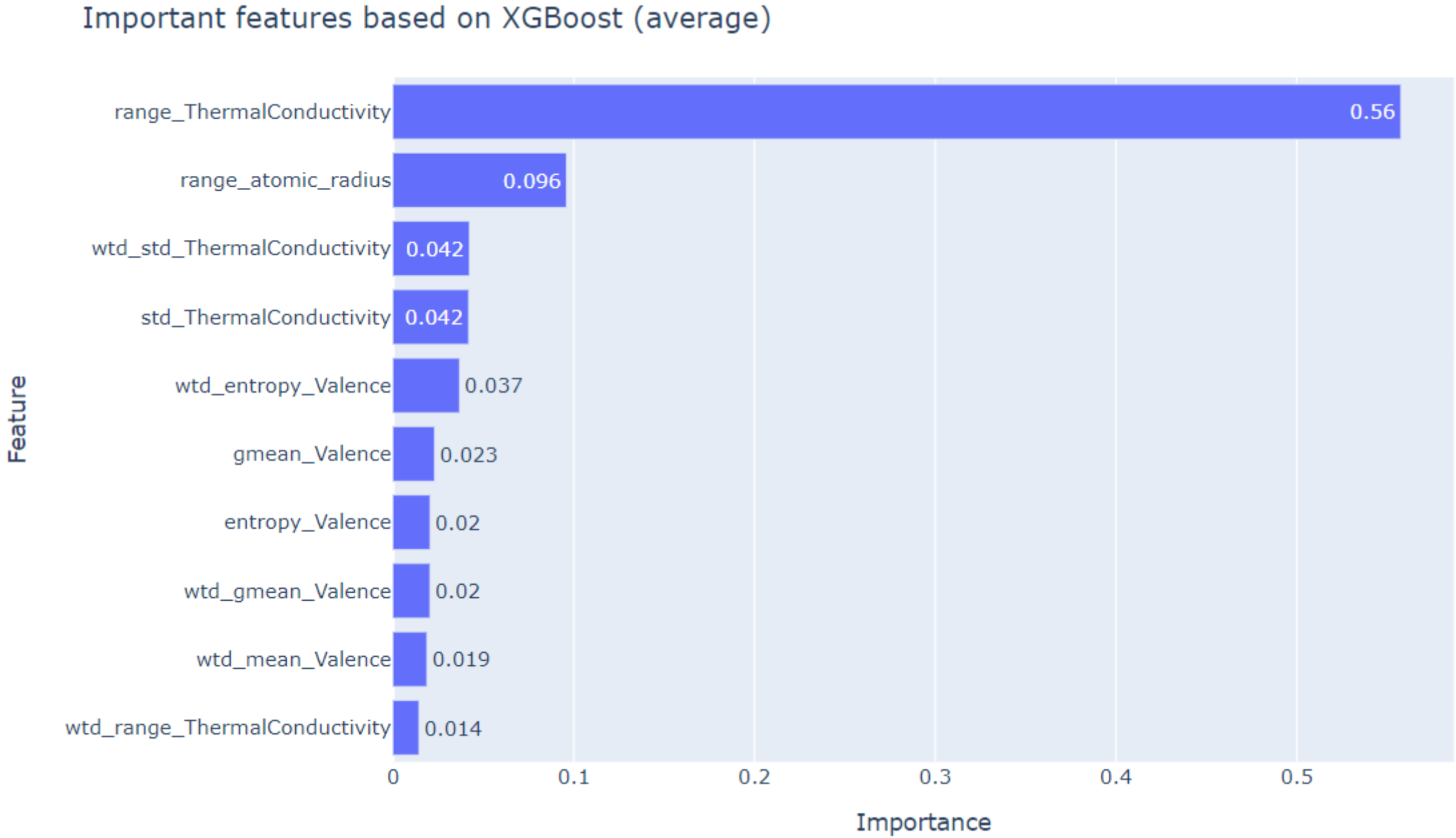

Pipeline mean CV score for xgb (k_best: mutual_info)
k=26, score=9.68
CV score (rmse)
10.4
10.2
10
9.8
9.6
9.4
9.2
9
0
10
20
30
40
50
60
70
80
k


Important features based on XGBoost (mutual_info_regression, k=26)
range_ThermalConductivity
range_atomic_radius
std_ThermalConductivity
wtd_std_ThermalConductivity
wtd_entropy_Valence
entropy_Valence
gmean_Valence
wtd_gmean_Valence
wtd_mean_Valence
entropy_FusionHeat
0.44
0.13
0.08
0.06
0.04
0.03
0.03
0.02
0.02
0.01
Feature
0
0.1
0.2
0.3
0.4
Importance

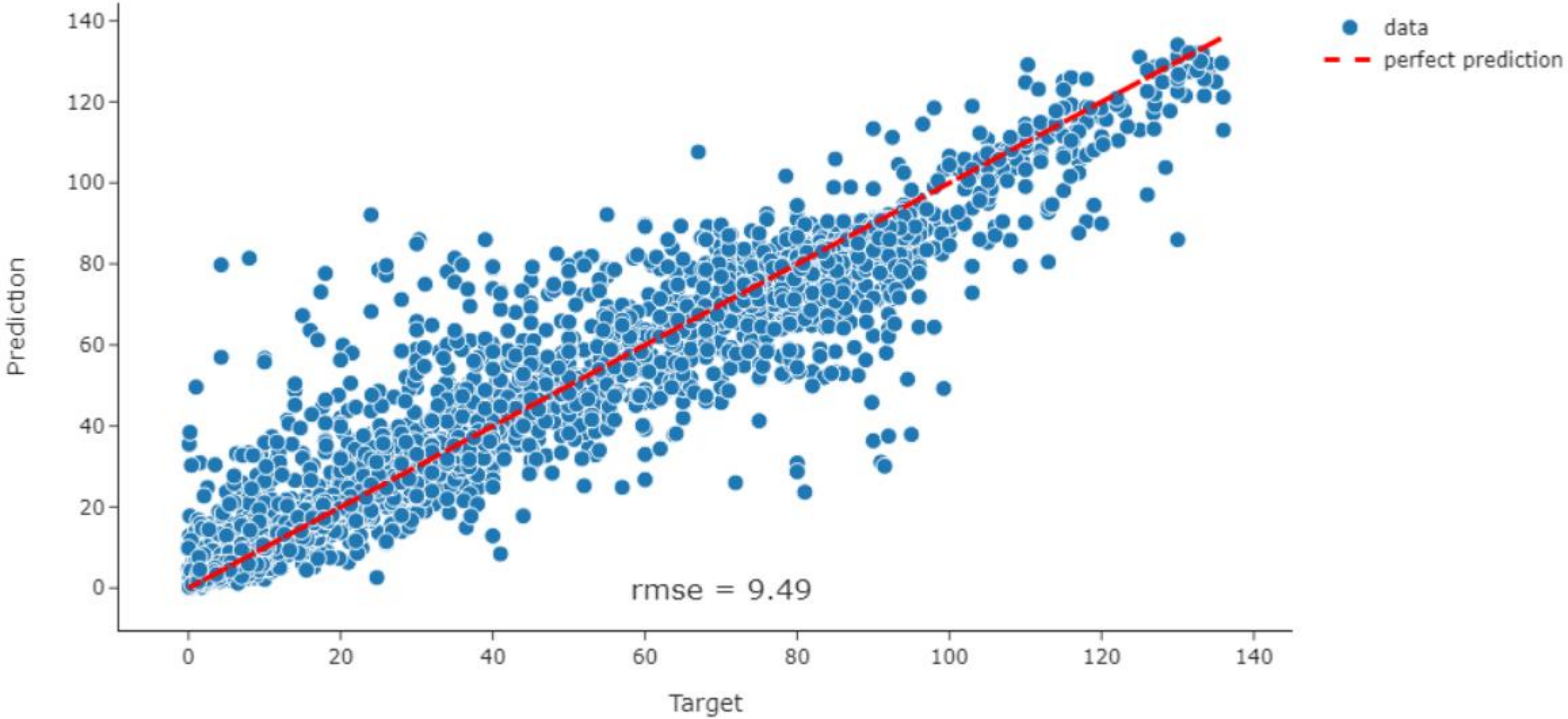
XGBoost


Pipeline mean CV score for xgb (rfe)
k=27, score=9.59
CV score (rmse)
10.4
10.2
10
9.8
9.6
9.4
9.2
9
0
10
20
30
40
50
60
70
80
k

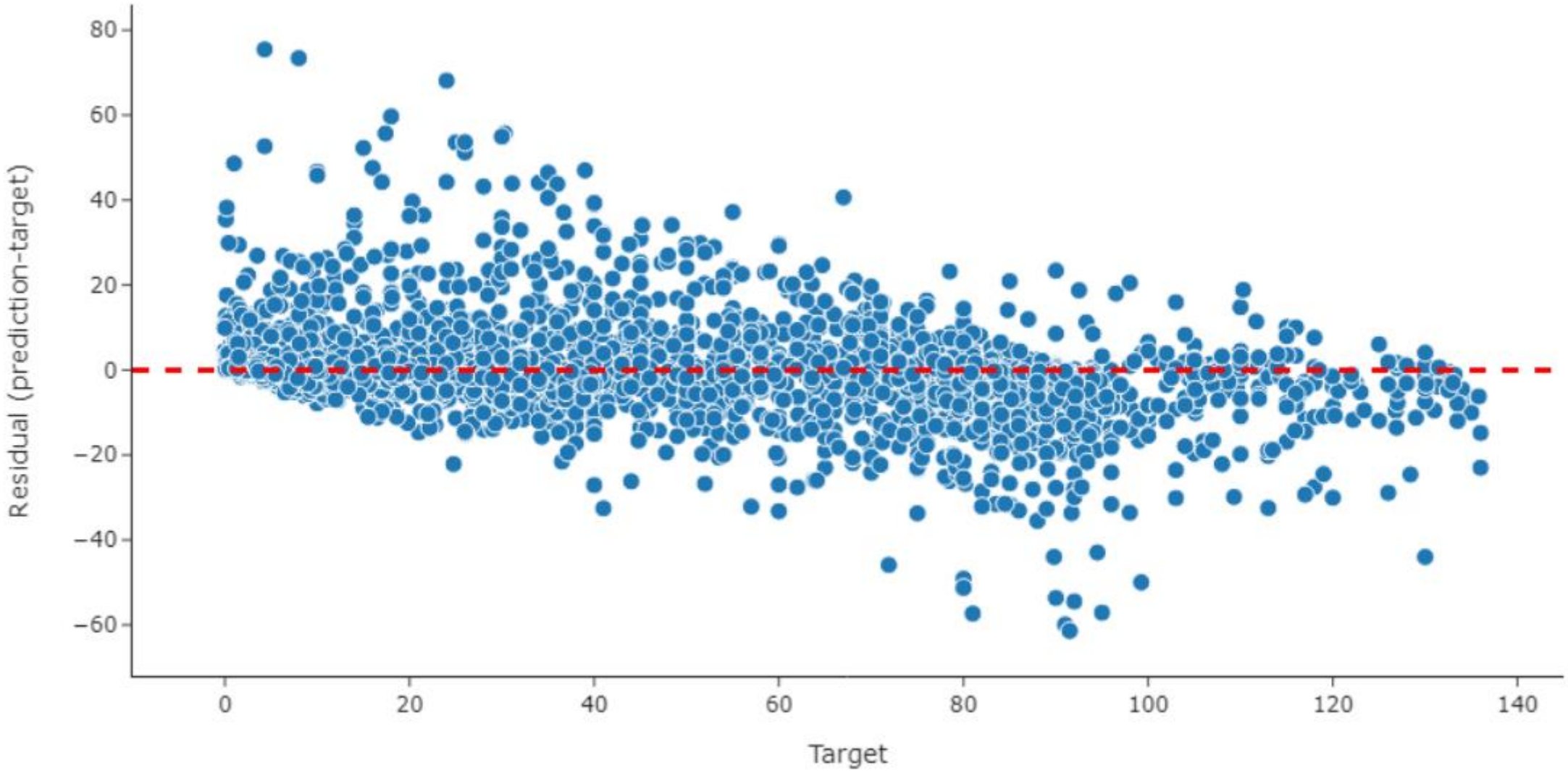

Important features based on XGBoost (rfe, k=27)
range_ThermalConductivity
range_atomic_radius
wtd_entropy_Valence
entropy_Valence
wtd_std_ThermalConductivity
gmean_Valence
wtd_gmean_Valence
std_ThermalConductivity
wtd_range_ThermalConductivity
wtd_gmean_ThermalConductivity
0.69
0.04
0.02
0.02
0.02
0.02
0.01
0.01
0.01
0.01
Feature
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Importance


Important features based on XGBoost (average)
range_ThermalConductivity
range_atomic_radius
wtd_std_ThermalConductivity
std_ThermalConductivity
wtd_entropy_Valence
gmean_Valence
entropy_Valence
wtd_gmean_Valence
wtd_mean_Valence
wtd_range_ThermalConductivity
0.56
0.096
0.042
0.042
0.037
0.023
0.02
0.02
0.019
0.014
Feature
0
0.1
0.2
0.3
0.4
0.5
Importance

Prediction based on XGB model (rfe)



Residual


Residual (prediction-target)
80
60
40
20
0
−20
−40
−60
0
20
40
60
80
100
120
140
Target

Potential improvements

- Use one-hot encoder on chemical formula
- Different model, such as neural network
- Classify data into low and high critical temperature before regression

Summary

- model: XGBoost
- feature selection: recursive feature elimination k=27 (from 81)
- best test score: 9.49
- most important features: range thermal conductivity, and range atomic radius
- most important properties: thermal conductivity, atomic radius, and valence