

Практикум

Лоскутова Софья, 615 группа, химический факультет

<https://colab.research.google.com/drive/1ZLUDdBbFpuA2KioPPDGm87QvEMq1vVPJ?usp=sharing>

Сначала в качестве дескриптора использовались только атомные веса, данные были отнормированы методом MinMaxScaler(). Результаты метрики R^2 на разных моделях:

	R^2 train	R^2 test
Ridge(alpha=0.01)	0.034	0.005
KNeighborsRegressor()	0.688	0.521
RandomForestRegressor()	0.945	0.720

Результаты предсказаний не очень хорошие, лучше всего себя показала модель RandomForestRegressor()

Предсказания с помощью подхода Bag of Bonds улучшились только для линейной регрессии (Ridge). Чтобы сократить размерность признаков использовалось VarianceThreshold.

	R^2 train	R^2 test
Ridge(alpha=0.01)	0.843	0.571
KNeighborsRegressor()	0.684	0.549
RandomForestRegressor()	0.928	0.669

Подход ElementFraction дал достаточно хорошие результаты. Данные также были нормализованы

	R^2 train	R^2 test
Ridge(alpha=0.01)	0.902	0.835
KNeighborsRegressor()	0.863	0.750
RandomForestRegressor()	0.976	0.850

Используем нейросеть с функцией активации relu и стохастическим градиентным спуском.

Подход для признаков — ElementFraction

```
model = tf.keras.models.Sequential([
```

```
tf.keras.layers.Dense(dim_input, activation='relu'),
tf.keras.layers.Dense(512, activation='relu'),
tf.keras.layers.Dense(124, activation='relu'),
tf.keras.layers.Dense(1)
])
```

Количество эпох — 150

Метрики:

	R^2	MAE	RMSE
train	0.9570	10.8690	15.0440
test	0.8900	14.9770	23.4220

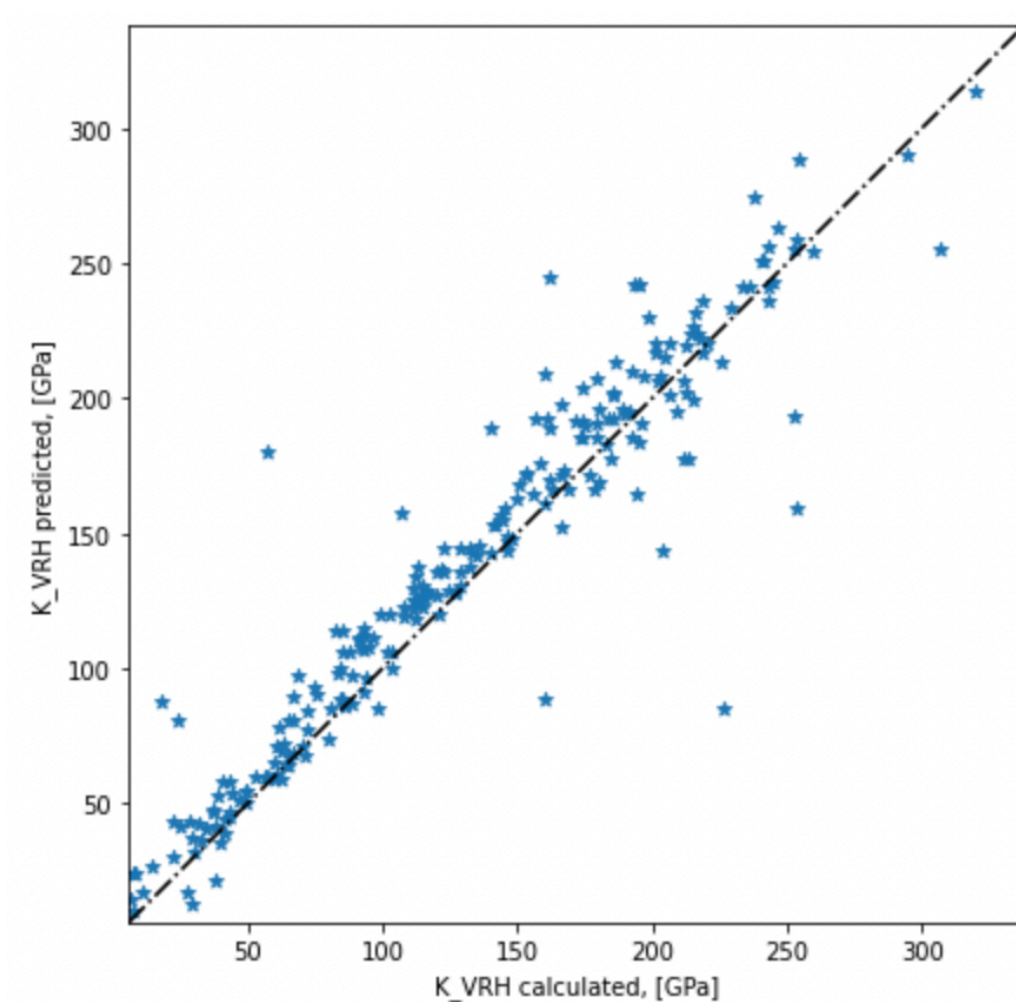


Рис. 1 График предсказанных и расчетных значений