

## Листок 05. Валидация моделей

**Упражнение 1.** Набор данных `sleep75` разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
sleep	totwrk, age, south, male

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distance'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distance'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

**Упражнение 2.** Набор данных `sleep75` разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
sleep	totwrk, age, south, male, smsa, yngkid, marr

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distance'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distance'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

**Упражнение 3.** Набор данных `wage2` разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
wage	age, IQ, south, married, urban

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distance'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distance'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

**Упражнение 4.** Набор данных **wage2** разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
log(wage)	age, IQ, south, married, urban

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distance'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distance'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

**Упражнение 5.** Набор данных **wage1** разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
wage	exper, female, married, smsa

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distance'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distance'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

**Упражнение 6.** Набор данных **wage1** разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
log(wage)	exper, female, married, smsa

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distance'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distance'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

**Упражнение 7.** Набор данных `Labour` разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
output	capital, labour, wage

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distance'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distance'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

**Упражнение 8.** Набор данных `Labour` разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
$\log(\text{output})$	$\log(\text{capital})$ , $\log(\text{labour})$ , $\log(\text{wage})$

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distance'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distance'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?