

In [4]:

```
# Партика 1
import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import KFold
from sklearn.metrics import r2_score

# Загрузка данных
data = pd.read_csv('abalone.csv')

# Преобразование признака Sex
data['Sex'] = data['Sex'].map(lambda x: 1 if x == 'M' else (-1 if x == 'F' else 0))

# Разделение на признаки и целевую переменную
X = data.iloc[:, :-1]
y = data.iloc[:, -1]

# Настройка кросс-валидации
kf = KFold(n_splits=5, shuffle=True, random_state=1)

# Список для хранения результатов
results = []

# Перебор количества деревьев от 1 до 50
for n_trees in range(1, 51):
    model = RandomForestRegressor(n_estimators=n_trees, random_state=1)
    r2_scores = []

    # Кросс-валидация
    for train_index, test_index in kf.split(X):
        X_train, X_test = X.iloc[train_index], X.iloc[test_index]
        y_train, y_test = y.iloc[train_index], y.iloc[test_index]

        model.fit(X_train, y_train)
        y_pred = model.predict(X_test)
        r2_scores.append(r2_score(y_test, y_pred))

    mean_r2 = np.mean(r2_scores)
    results.append((n_trees, mean_r2))

    # Проверка условия для досрочного прерывания
    if mean_r2 > 0.52:
        print(f"Минимальное количество деревьев: {n_trees}")
        print(f"Качество (R2): {mean_r2:.4f}")
        break
    else:
        print("Не удалось достичь R2 > 0.52 в диапазоне 1-50 деревьев")

# Анализ изменения качества
if len(results) > 1:
    print("\nАнализ изменения качества:")
    for i in range(1, len(results)):
        prev_score = results[i-1][1]
        curr_score = results[i][1]
        if curr_score < prev_score:
            print(f"При {results[i][0]} деревьях качество ухудшилось с {prev_score} до {curr_score}")
    print("В целом качество стабилизируется после достижения определенного количества деревьев")
```

Минимальное количество деревьев: 21

Качество (R^2): 0.5205

Анализ изменения качества:

При 11 деревьях качество ухудшилось с 0.4954 до 0.4944

При 20 деревьях качество ухудшилось с 0.5198 до 0.5195

В целом качество стабилизируется после достижения определенного количества деревьев

In []: