

Процесс решения задач машинного обучения

Визуализация матрицы корреляций

```
In cm = df.corr()  
sns.heatmap(cm, annot = True, square=True)
```

Отрисовка попарных графиков

```
In sns.scatterplot(df['Признак 1'], df['Признак 2'])
```

Замена категорий числовыми лейблами (Label Encoding)

```
In from sklearn.preprocessing import LabelEncoder  
  
# создаём переменную класса LabelEncoder – кодировщик  
encoder = LabelEncoder()  
# используем кодировщик, чтобы "перевести" строковые названия в числа  
df['column'] = encoder.fit_transform(df['column'])
```

Представление категориального поля в множество бинарных полей (One-hot Encoding)

```
In df = pd.get_dummies(df)
```

Задание у модели критерия оптимизации

```
In # Возможные значения критерия можно посмотреть в документации  
model = RandomForestRegressor(criterion='mae')
```

Получение коэффициентов линейной регрессии

```
In # Получение коэффициентов линейной регрессии  
feature_weights = model.coef_  
# Получение нулевого коэффициента  
weight_0 = model.intercept_
```

Получение важности признаков для дерева решений, случайного леса и градиентного бустинга

```
In importances = model.feature_importances_
```