Прогнозирование характеристик ШАЛ

Работу выполняла команда sudo rm-rf

Задача: Комплексный анализ ШАЛ

В ходе работы на основе 100000 событий ШАЛ из бинарного файла необходимо было обучить модель, способную предсказывать:

- Направление ливня (Зенитный угол θ, Азимут φ)
- Положение оси (Координаты X, Y)
- Мощность линиия (lgN_e, где N_e - это число электронов)
- Возраст ливня (этап развития, где 0 - это начало ливня, 1 – это максимальное число частиц, 2 – погашение ливня)

Предобработка данных: подготовка к анализу

Перед обучением модели, было необходимо выполнить несколько шагов:

- Очистка переменных от пропущенных значений
- Обработка кортежей и добавление их в датасет
- Циклическое преобразование азимута
- Заполнение пропусков

```python
# 1. Загрузка данных
df = pd.read_csv('result.csv')

# 2. Очистка данных
df = df.dropna(subset=['tetta', 'phi', 'x', 'y', 'power', 'age'])

# 3. Обработка кортежей
def expand_tuple_column(df, col_name):
    if col_name in df.columns:
        # Извлекаем кортежи
        tuples = df[col_name].apply(lambda x: eval(x) if isinstance(x, str) else x)

        # Создаем временный DataFrame с развернутыми значениями
        expanded = pd.DataFrame(tuples.tolist())
        expanded.columns = [f'{col_name}_{i}' for i in range(expanded.shape[1])]

        # Удаляем исходный столбец и добавляем развернутые
        df = df.drop(col_name, axis=1)
        df = pd.concat([df, expanded], axis=1)
    return df

# Применяем к столбцам с кортежами
df = expand_tuple_column(df, 'energy')
df = expand_tuple_column(df, 'threshold_time')

# 4. Циклическое преобразование азимута
df['phi_sin'] = np.sin(np.radians(df['phi']))
df['phi_cos'] = np.cos(np.radians(df['phi']))

# 5. Подготовка признаков
# Выбираем только числовые столбцы
numeric_cols = df.select_dtypes(include=[np.number]).columns.tolist()
X = df[numeric_cols].drop(['tetta', 'phi', 'x', 'y', 'power', 'age', 'phi_sin', 'phi_cos'], axis=1, errors='ignore')

# Заполнение пропущенных значений
imputer = SimpleImputer(strategy='mean')
X_imputed = imputer.fit_transform(X)
X = pd.DataFrame(X_imputed, columns=X.columns)
```

Выбор моделей для прогнозирования

- MultiOutputRegressor

Был выбран для эффективного предсказания нескольких целевых переменных (азимут и координаты), ввиду способности моделировать каждый выход независимо

- RandomForestRegressor

Применялся для направления (θ, φ), положения оси (X, Y) и мощности (lgN_e)

- GradientBoostingRegressor

Выбран для прогнозирования возраста ливня ввиду способности к последовательному улучшению

Оценка качества моделей: Метрики MSE и R^2

Как можно заметить, метрика R^2 почти по всем параметрам точнее чем MSE

```
============================================================
Обучение модели для направления прихода...
Время обучения: 20.4 сек
Направление прихода:
MSE зенитного угла (tetta): 69.3003, R²: 0.5059
MSE азимута (phi): 18610.7066, R²: -0.8031

============================================================
Обучение модели для положения оси...
Время обучения: 13.0 сек
Положение оси ливня:
MSE координаты X: 138.9975, R²: 0.7498
MSE координаты Y: 168.2518, R²: 0.8855

============================================================
Обучение модели для мощности ливня...
Время обучения: 6.8 сек
Мощность ливня:
MSE: 0.0189, R²: 0.9531

============================================================
Обучение модели для возраста ливня...
Время обучения: 15.2 сек
Возраст ливня:
MSE: 0.0034, R²: 0.1912
```

Визуализация результатов: положение и направление

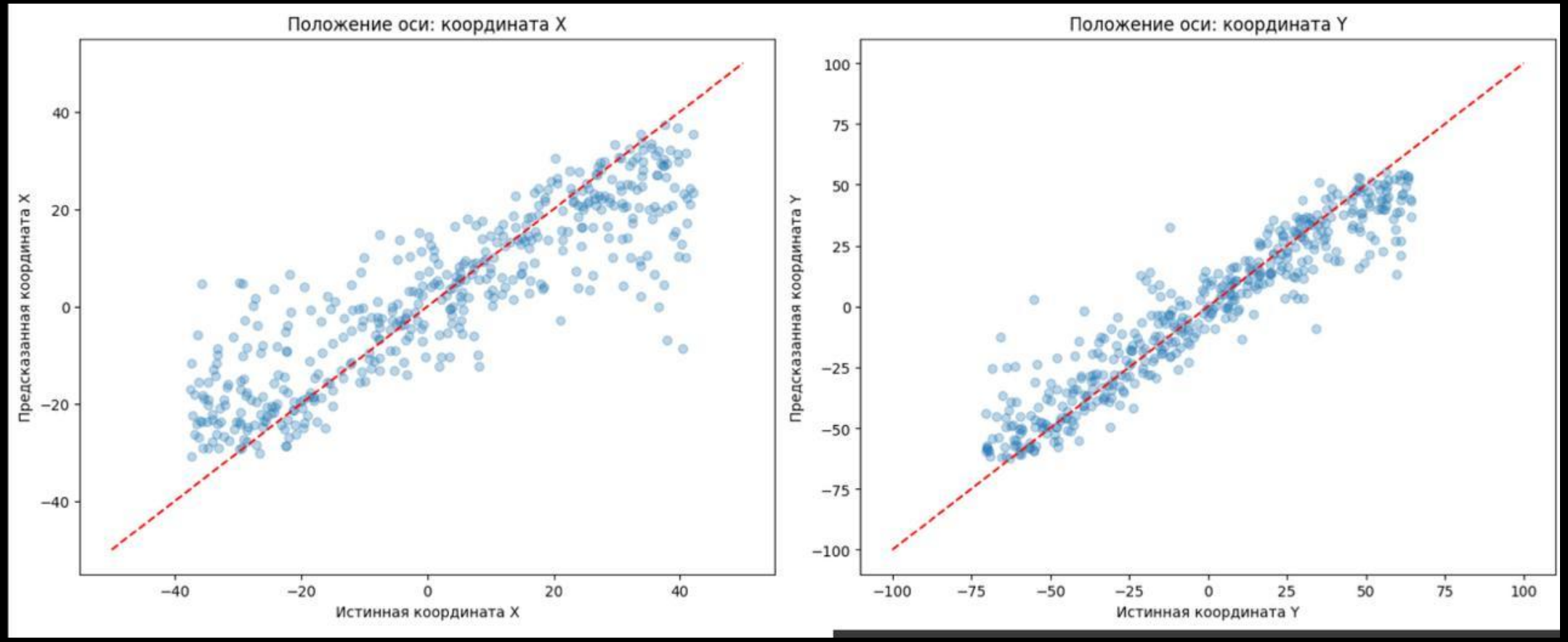


График предсказанных значений против истинных показывают, что модель неплохо справляется с прогнозированием координат

Визуализация результатов: положение и направление

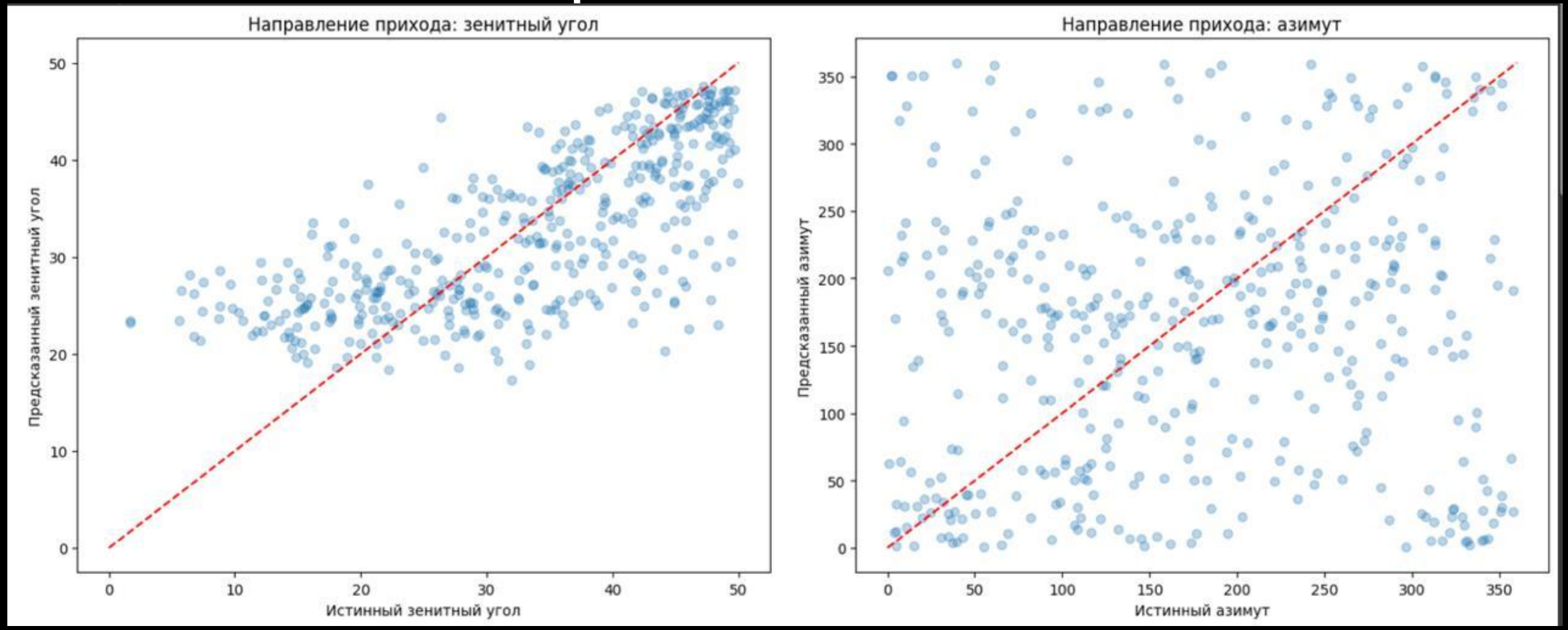
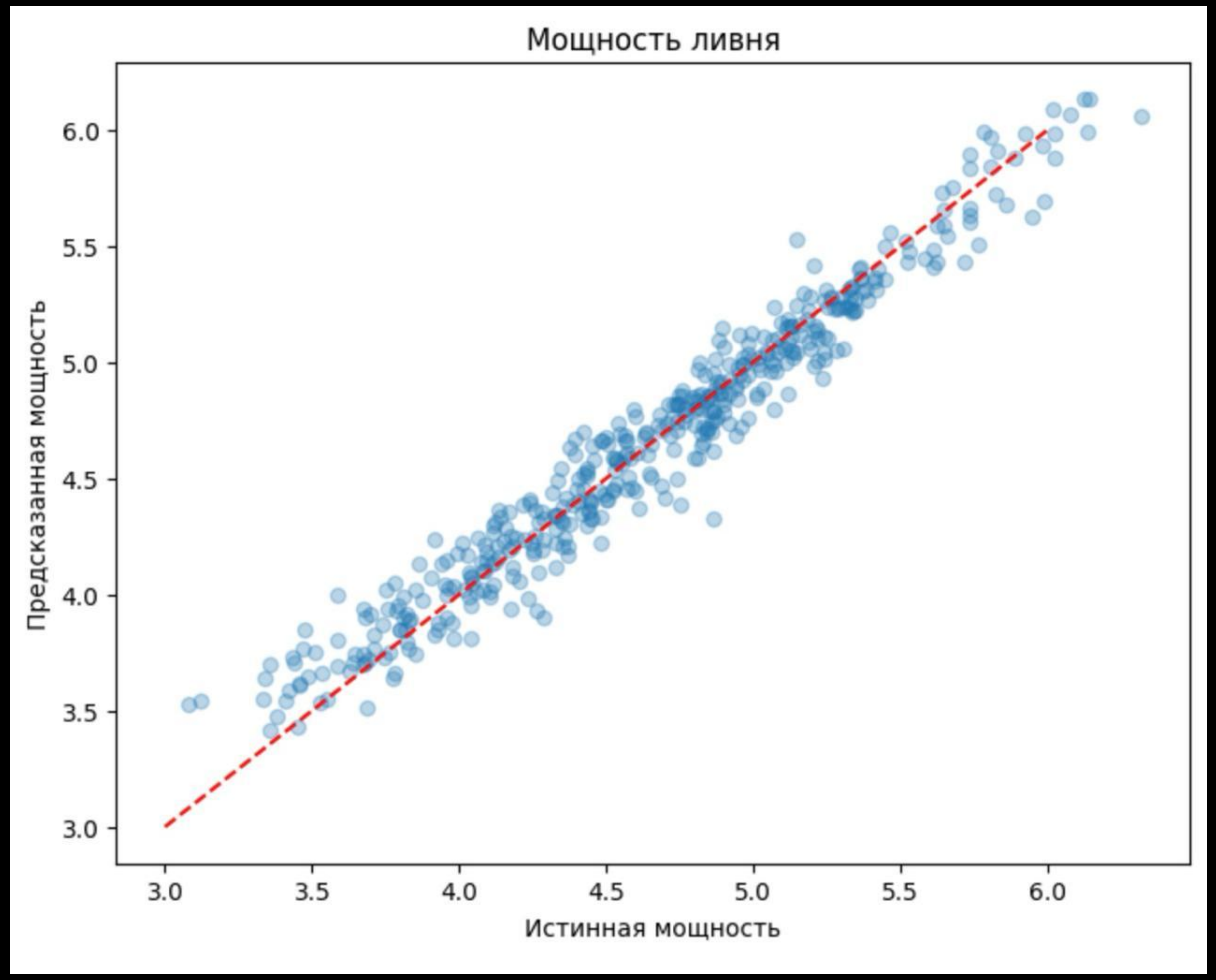


График предсказанных значений против истинных показывают, что модель неплохо справляется с прогнозированием зенитного угла, однако с прогнозированием азимута возникают проблемы

Визуализация результатов: мощность и возраст



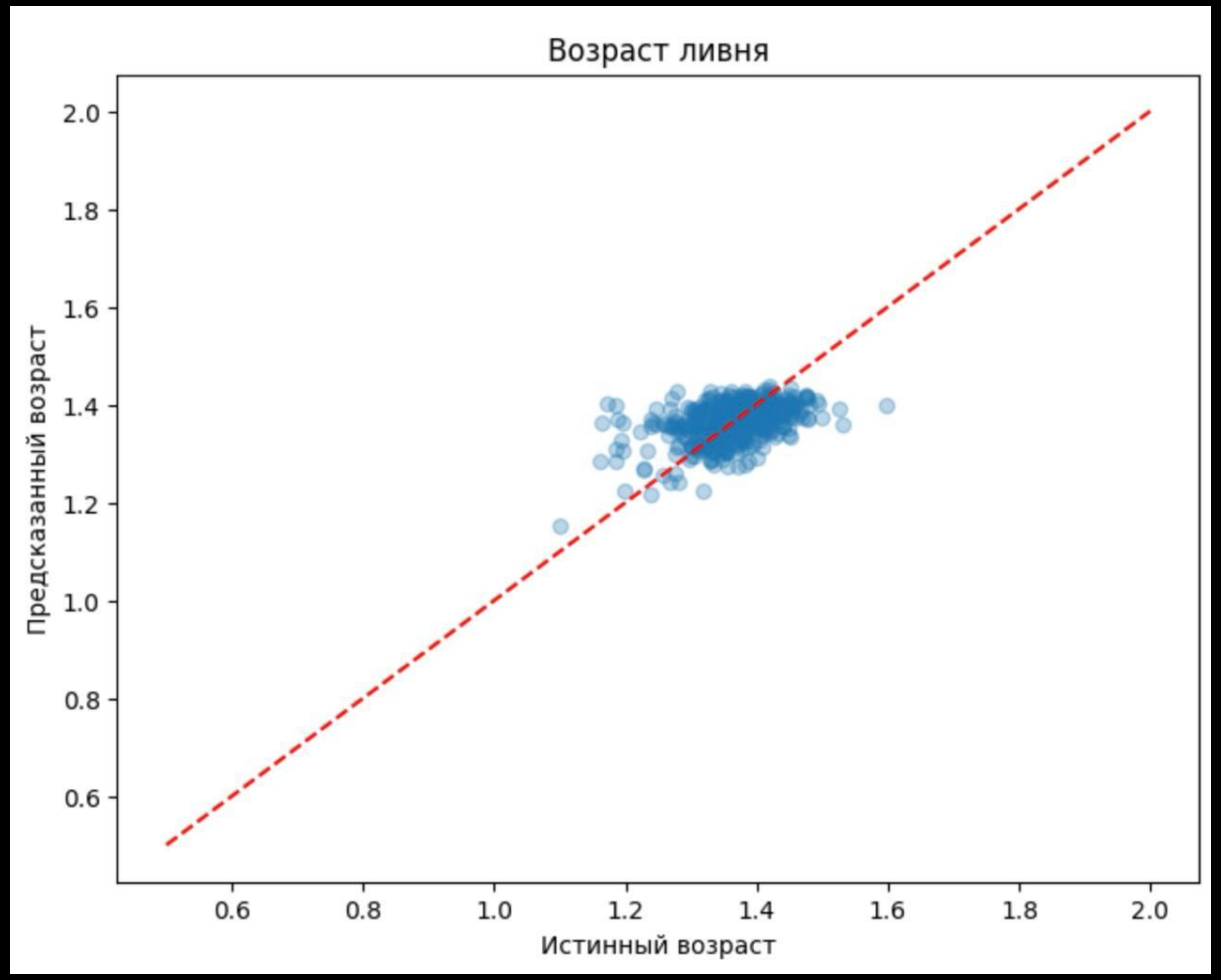


График предсказанных значений против истинных показывают, что модель неплохо справляется с прогнозированием возраста и мощности ливня