**Отчет по проекту: Прогнозирование потребления электроэнергии в Индии**

**2. Выбор источников данных**

* **Источник**: Исторические данные о потреблении электроэнергии в штатах Индии (2019–2020).
* **Поиск**: Датасет был найнед на Kaggle: [ссылка](https://www.kaggle.com/datasets/twinkle0705/state-wise-power-consumption-in-india/code).
* **Анализ**: Данные представлены в виде временного ряда за период 17 месяцев с 2 января 2019 года по 23 мая 2020 года. Строки индексированы датами, столбцы соответствуют штатам. Каждое значение отражает потреблённую электроэнергию для указанного штата на указанную дату.

**3. Предобработка данных**

* **Предобработка**:
  + Кодирование штатов (LabelEncoder, сохранен в state\_encoder.pkl).
  + Масштабирование числовых признаков (StandardScaler, feature\_scaler.pkl).
  + Извлечение временных признаков: день года, месяц, день недели, сезон.
  + Проверка целостности данных, удаление пропусков.

**4. Разведочный анализ данных**

* **Закономерности**:
  + Существуют сезонные колебания (высокое потребление летом).
  + Существуют различия между штатами (индустриальные штаты потребляют больше).
  + Дневное потребление выше ночного.

**5. Выбор признаков и инженерия**

* **Признаки**:
  + State\_Encoded: Закодированное название штата.
  + DayOfYear: 1–365/366.
  + Month: 1–12.
  + DayOfWeek: 0–6.
  + Season: 0 (зима), 1 (весна), 2 (лето), 3 (осень).
* **Инженерия**:
  + Сезон определен на основе месяцев (зима: 12,1,2; весна: 3,4,5; лето: 6,7,8; осень: 9,10,11).

**6. Выбор и обучение моделей**

* **Модель**: XGBoost Regressor.
* **Обоснование**: Высокая точность, устойчивость к шуму, поддержка категориальных признаков.

**7. Анализ результатов**

* **Метрики** (из model\_metadata.pkl):
  + Train RMSE: 22.03 MW | MAE: 13.53 MW
  + Test RMSE: 22.20 MW | MAE: 13.56 MW
* **Вывод**: XGBoost показал достаточную точность для прогнозирования.

**10. Разработка графического интерфейса**

* **Инструмент**: Streamlit.
* **Компоненты**:
  + Боковая панель: выбор даты, штата, кнопка "Predict".
  + Основная область: карта, графики, таблицы.
* **Файл**: app.py.