

PJ：電力消費予測モデル化およびAPI化

開始月	2024年10月	チーム人数	1人
期間	1.5カ月		

開発環境（OS）	Windows11
言語	Python3.12.4
Framework/Library	Pandas, Numpy, Statsmodels, Plotly, Pmdarima, Scikit-learn, Prophet, LightGBM, Flask
ツール・その他	Jupyter Lab, Visual Studio Code
GitHub	<a href="#">GitHub URL</a>

【概要】

- 目的: 東京エリアの2023年の電力消費量を予測モデルを作成すること。また、最良モデルでAPIを構築しWeb上で可視化すること。
- 利用データ:
  - 東京エリアの1時間ごとの電力消費量データ（2016年～2023年）。出典：[東京電力パワーグリッド でんき予報](#)
  - 1時間ごとの平均気温データ（同上）。出典：[国土交通省 気象庁 過去の気象データ・ダウンロード](#)

【データの性質と前処理内容】

- train データ、test データの規模:
  - 学習データ：2016年4月～2022年12月（59,184行）。
  - テストデータ：2023年1月～12月（8,760行）。
- データの特徴量:
  - 目的変数：電力消費量（10MW単位）
  - 説明変数：気温（摂氏）、休日（土日祝日および祝日の振替休日を「1」、平日を「0」とするフラグ処理を実施）
- 与えられたデータの課題と改善点:
  - 複数の周期性（1日単位、1週間単位）が観察される非線形的なデータ特性。
  - 季節性や休日/平日の消費傾向を反映する特徴量の生成。

【実装内容・工夫した点】

- 前処理
  - ホームページよりcsvファイルを取得し、データクレンジングを実施。
  - EDA（探索的データ分析）を通じて、データの周期性や特徴を可視化。
  - 気温データを用い、平均気温からの絶対温度差を新たな特徴量として生成（閾値：18.1℃）。
  - 国民の祝日および振替休日も考慮したカレンダーデータを作成。
- モデル構築
  - モデル選定の基準と工夫:
    - 古典的時系列分析モデルが使えるかどうかをSTF分解やACF、PACFを分析することで考慮。
  - 古典的時系列分析モデル:
    - 性能を確かめるためAMモデルで基本的な予測を実施。
  - Prophetモデル:
    - 季節性を自動調整する能力を活用し、基本的な予測を実施。
    - 特徴量エンジニアリングを行い、グラフで予測傾向を比較。
  - 決定木系モデル（ランダムフォレスト、ヒストグラディエントブースティング木、LightGBM）:
    - 特徴量重要度を解析し、モデル改善の手がかりを得た。
    - 学習率、木の本数などのハイパーパラメータを最適化。
- モデルの学習と評価
  - RMSE（Root Mean Squared Error）とR<sup>2</sup>（決定係数）を評価指標に採用。
  - 学習データとテストデータの両方でモデルの予測精度を確認。また、グラフ化により、予測値と実測値の相関を評価。
- API化とフロントエンドの実装
  - Flaskフレームワークを用いたAPI開発:
    - 日付、時間、気温を入力として受け取り、予測結果を返す。
    - 予測結果として、指定時刻の予測値、1日の消費量グラフを生成し、結果をJSON形式で返却。
    - 入力データに気温がない場合でも、過去の同日同時間帯データで補充可能な仕組みを導入。

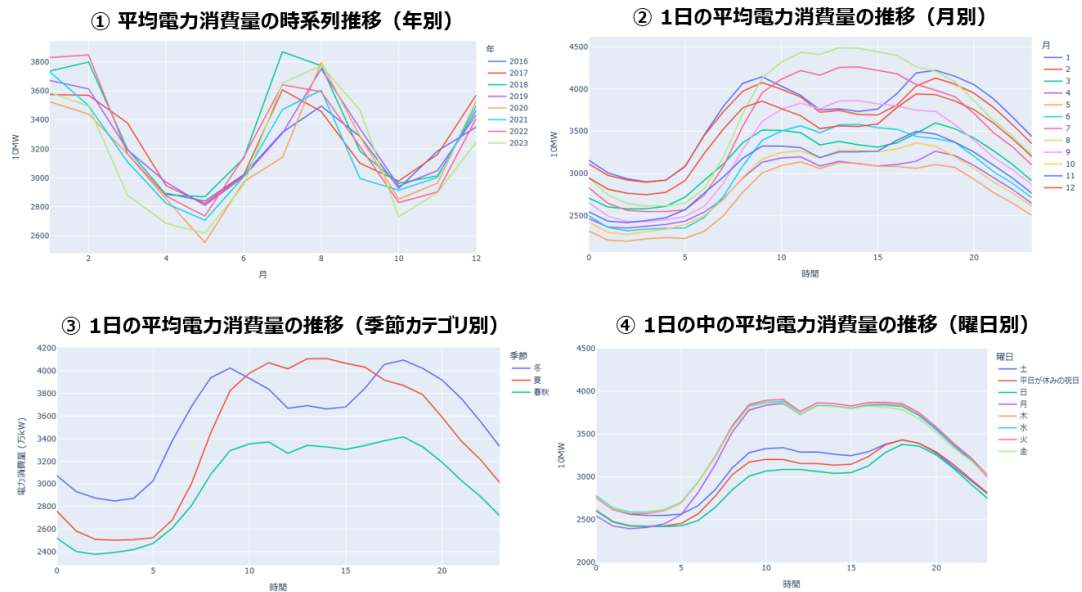
【結論】

- 今回は、時系列分析モデル、Prophetモデル、決定木系モデルを比較検証し、最も性能が良かったのはProphetモデルに休日フラグと気温の絶対値差分を追加し、さらに残差にLightGBMを適用するアンサンブル手法（テストデータにおける二乗平均平方根誤差: 215.5838、決定係数: 0.9083）。
- この結果の背景として、Prophetモデルがトレンドや季節要因を自動的に捕捉する能力を発揮し、さらにトレンドや季節要因を除いた非線形の残差をLightGBMが効果的に捉えたことが精度向上につながったと考えられる。
- 一方で、今回用いた説明変数（気温と休日）だけでは、電力使用量の全てを予測するには不十分であり、湿度や降水量などの気象要因や、新たな社会的要因を説明変数として加えることで、さらなる改善が期待できる。
- また、現状の手法ではさらなる非線形性の捉え方や長期依存性を十分に扱えていない可能性があるため、改善案として以下を試したい。
  - RNN（再帰型ニューラルネットワーク）やLSTM（長短期記憶）等の深層学習を活用したモデルを用いることで、複雑な要因間の相互作用をより正確に捉える。

## 【PJ：電力消費予測モデル化およびAPI化の要旨】

### 1. 可視化

- データの周期性と重要特徴量を分析するため、以下の時系列グラフを用いて可視化を行った。

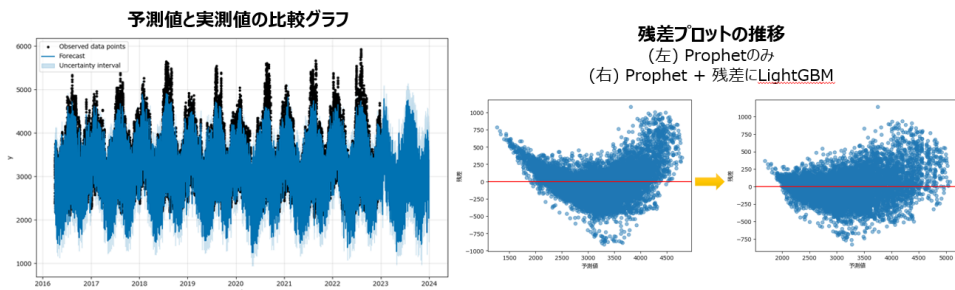


- 結果の要点

- 夏・冬に電力消費量が増加し、秋・春に減少することが確認された（①）。
- 一日の中では日中に電力消費量が増加し、夕方から夜にかけて減少する傾向が見られる（②、③）。
- 平日は土日・祝日と比較して電力消費量が多い（④）。

### 2. モデル化

- 気温と電力消費量の関係を考慮し、新たに気温の絶対値差分や休日フラグを特徴量として追加。複数のモデルを試行した結果、**Prophetモデル**に**LightGBM**で補正するアンサンブル手法が最も高い性能を示した。



- 結果の要点

- 予測と実測のグラフから周期性を正確に捉えられていることが分かる。ただし、残差プロットにはパターンが残っており、さらなる改善の余地がある。

### 3. API化

- 最良モデルをAPI化し、予測結果を取得できるシステムを構築した。
- システム構成図およびAPIの動作イメージを以下に示す。

#### ローカル環境（Python 3.12, Jupyter Lab）で動作



#### 入力画面

入力フォーム

目的:  
YYYY/MM/日

地域:  
00/00

気温（任意 / 未入力の場合は過去同日の平均気温）  
例: 15

予測ボタン

#### 出力画面

