**지속가능한 전력 생산**

**전력 수요, 태양열 발전량, 풍력 발전량 예측**

**Team. 빵빵이들**

**팀원**

이정현 201802798

이의진 201902743 (PM)

정경서 201903156

최서여 202103464

**프로젝트 목표**

전기 수요량, 풍력 발전량, 태양광 발전량 예측을 통한 화석 에너지 발전량 제안

전기 수요량을 예측하고 풍력, 태양열 발전량을 예측하여 생산해야 할 화석 에너지를 통한 전기생산량을 도출

**연구 배경**

1. 우리 삶에 필수적인 요소, 전기

전력은 4차 산업혁명의 주춧돌이면서 문명사회의 근간을 이루는 필수 요소이다.

전력 수요를 잘못 예측하여 정부가 발전설비를 충분히 짓지 않게 되면 전국적으로 대정전을 겪을 수 있다. 이는 실제 2011년 ‘9·15 순환 대정전’에서 신호 고장으로 도로를 아수라장으로 만들고, 기업들의 업무 마비 등과 같은 피해로 이어졌다. 그렇기 때문에, 첨단사회일수록 정전이 가져오는 피해도 커진다고 할 수 있고, 이 점이 바로 전력 수요를 제대로 예측하고 필요한 만큼만 발전소를 가동할 수 있도록 준비해야 하는 이유이다.

1. 지구 온난화의 주범, 화석 연료

최근 몇 십년 동안 온난화로 인한 기후 변화가 우리의 일상에 지속적인 영향을 미치고 있다. 이로 인해 화석 연료 사용량을 줄여야 할 필요성이 대두되고 있다. 신재생 에너지, 특히 태양광 및 풍력 에너지는 이러한 화석 연료 대체에 큰 기여를 할 수 있다.

1. 전력 수요 예측의 핵심 요인, 기후

전력 수요는 평균기온 등 기상상황과 밀접한 관련이 있다. 통상 평균기온이 오르면 에어컨 등 전력 다소비 기기 사용이 늘어 전력 수요가 급증하고, 평균 기온이 내리면 전력 수요가 줄어든다.

**프로젝트 범위**

대한민국의 3개년 데이터

**데이터**

1. **기상 데이터**

**기상청 API HUB**, 종관기상관측 데이터를 바탕으로, 전국, 3년(20,21,22년도), 5분단위의 기온, 강수, 기압, 습도, 풍향, 풍속, 일사, 일조, 적설, 구름, 시정, 지면 · 초상온도 등에 대한 기상 데이터를 제공한다.

1. **전력 데이터**

**한국전력거래소,** 정보공시 데이터를 바탕으로, 전국, 3년, 5분단위의 공급능력(MW), 현재수요(MW), 최대예측수요(MW) ,공급예비력(MW), 공급예비율(퍼센트), 운영예비력(MW), 운영예비율(퍼센트) 등의 전력 사용량에 관련된 데이터를 제공한다.

1. **태양열 및 풍력 발전량 데이터**

**한국전력거래소,** 정보공시 데이터를 바탕으로, 전국, 3년, 1일 단위의 태양열 및 풍력 발전량 데이터를 수집하였다. 이는 앞으로 예측할 예측 발전 생산량(화력 등)을 결정하는데 사용할 예정이다. ( ex)예측 발전 생산량 = 예측 전력 수요량 – 예측 태양열 및 풍력 발전량 )

**중간 진행 상황**

단일 모델 분석

1. 단일 모델 분석을 위한 데이터 전처리

풍력, 태양 에너지를 통한 전기 생산량의 예측을 위한 데이터셋 생성

|  |  |
| --- | --- |
| 구분 | 구성 |
| 기상요소 | 습도, 풍속, 강우량, 풍향 |
| 시간 데이터 | 년, 월, 일 |
| 위치 데이터 | 도시명 |
| 예측 목표값 | 태양광 발전량, 풍력 발전량 |

풍력, 태양 에너지를 통한 전기 생산량의 예측을 위한 데이터셋 생성

|  |  |
| --- | --- |
| 요소명 | 전처리 방법 |
| 습도, 풍속, 강우량, 풍향 | 표준화 |
| 도시명 | One-Hot Encoding을 통해 벡터화 |

모델의 학습을 위해서 train test split을 8:2 로 진행하여 8에 해당하는 데이터로 기계 학습을 진행.

1. 모델을 통한 예측

사용 모델 Regressor,

1. Decision Tree

Decision Tree는 tree 구조를 활용하여 entropy가 최소화되는 방향으로 데이터를 분류하거나 원하는 어떤 결과값을 예측하는 분석 방법

1. Multi Linear Regression

여러 개의 독립 변수와 하나의 종속 변수의 선형 관계를 모델링 하는 분석 방법

1. Random Forest

기계 학습에서의 랜덤 포레스트는 분류, 회귀 분석 등에 사용되는 앙상블 학습 방법의 일종으로, 훈련 과정에서 구성한 다수의 결정 트리로부터 부류(분류) 또는 평균 예측치(회귀 분석)를 출력함으로써 동작한다.

1. Multi-Layer Perceptron

다층 퍼셉트론(Multilayer perceptron, MLP)은 퍼셉트론을 여러층 쌓은 순방향의 인공 신경망이다. 입력층(input layer)과 은닉층(hidden layer)과 출력층(output layer)으로 구성된다. 각 층에서는 활성함수를 통해 입력을 처리하여 데이터를 분류한다.

1. 모델을 통한 예측 평가

태양 에너지 예측

|  |  |  |
| --- | --- | --- |
| 모델명 | RMSE | Parameter |
| Decision Tree | 77.42 | max\_depth=8 |
| Multi Linear Regression | 108.63 |  |
| Random Forest | 54.86 | max\_depth=8 |
| Multi-Layer Perceptron | 49.69 | max\_iter=500 |

풍력 에너지 예측

|  |  |  |
| --- | --- | --- |
| 모델명 | RMSE | Parameter |
| Decision Tree | 47.57 | max\_depth=8 |
| Multi Linear Regression | 52.24 |  |
| Random Forest | 41.48 | max\_depth=8 |
| Multi-Layer Perceptron | 41.08 | max\_iter=500 |

1. 추후 계획

풍력 발전과 태양에너지 발전에 0 값이 많이 분포하고 있다. 그렇기 때문에 0인 데이터와 아닌 데이터를 랜덤 포레스트를 통해 분류하고 랜덤 포레스트에서 0이 아닌 값으로 분류한 결과에 대해서 현재 예측에 가장 좋은 지표를 보이고 있는 MLP를 통해 예측을 진행하는 두단계 모델을 구성할 계획임

**전력 데이터 시계열 분석**

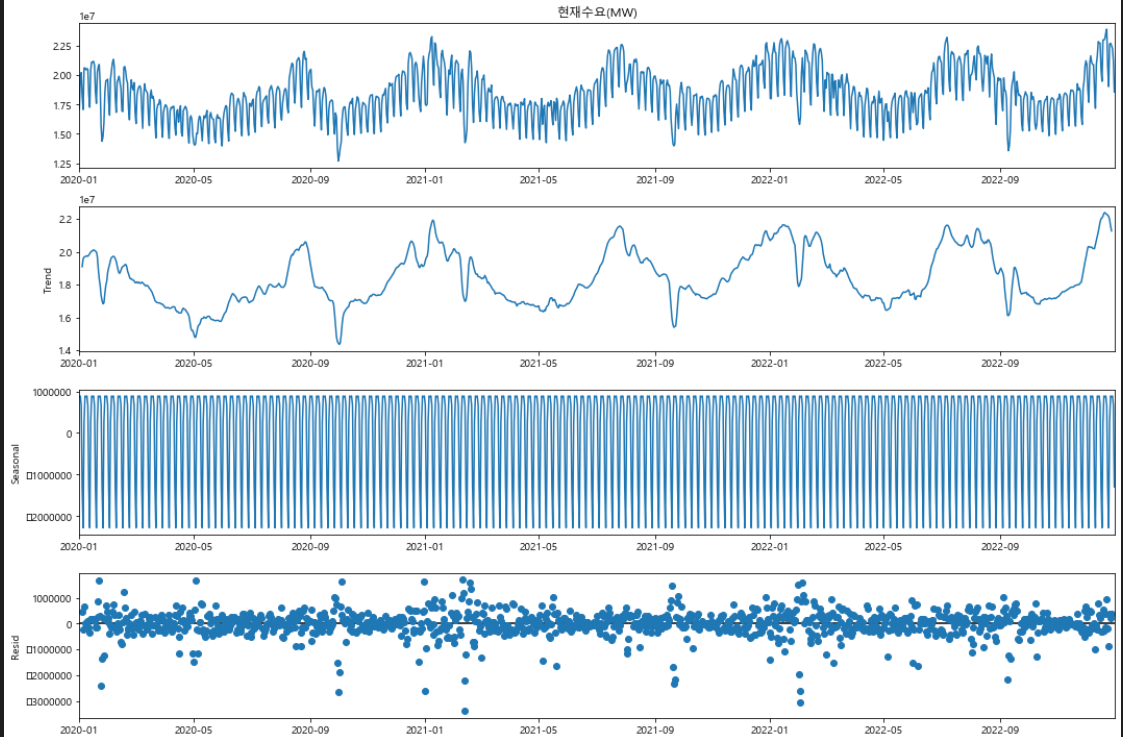
시계열 분석을 위한 데이터 전처리

- 시계열 데이터

시간에 순차적으로 관측한 값들의 집합으로, 예측 모델에서 시간을 변수로 사용하는 특징이 있다. 본 프로젝트에서는 전력 수요 데이터를 시계열 데이터로 판단하여 시계열 분석을 하기로 결정했다.

- 시계열 분해

시계열 분해는 시계열 데이터를 구성하는 다양한 구성 요소를 분리하여 각 구성 요소의 특성을 더 잘 이해하고 모델링하기 위한 기술이다.



- Observed (관측값): 원본 데이터인 '현재수요(MW)' 열

- Trend (추세): 데이터의 장기적인 변동(증가, 감소)을 보여주는 부분

- Seasonal (계절성): 주기적으로 반복되는 패턴을 보여주는 부분

- Residual (잔차): 추세와 계절성을 제외한 나머지 부분으로, 모델이 설명하지 못하는 남은 변동이 표시된다.

- 정상성 검증

정상 시계열은 시간의 흐름에 따라 그 통계적 특성이 변하지 않는 시계열을 의미한다. 정상 시계열의 경우, 시간이 변해도 평균과 분산(표준편차), 공분산은 일정하다. 많은 통계적 시계열 모델들(AR, MA, ARMA, ARIMA 등)은 시계열 데이터의 정상성을 가정하고 있다.

추세나 계절성이 있는 시계열은 정상 시계열이 아니다. 정상성을 나타내지 않는 데이터는 복잡한 패턴을 모델링하여 분석하기 어렵기 때문에, 정상성을 갖도록 차분, 로그 변환 등 전처리 후 분석을 시행한다.

차분은 연이은 관측값들의 차이를 계산하는 것으로, 시계열의 수준에서 나타나는 변화를 제거하여 시계열의 평균 변화를 일정하게 만드는데 도움이 된다. 결과적으로 추세나 계절성이 제거 혹은 감소된다.

**1. Kwiatkowski-Phillips-Schmidt-Shin(KPSS) 검정**

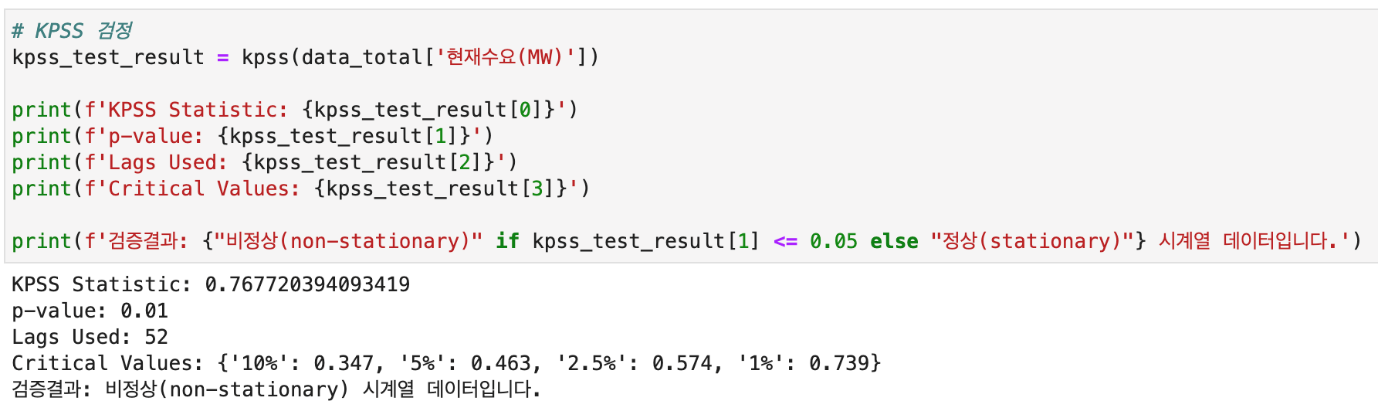
95% 신뢰도를 바탕으로 0.05 유의수준을 두면

- H0: 해당 시계열은 정상 시계열이다.

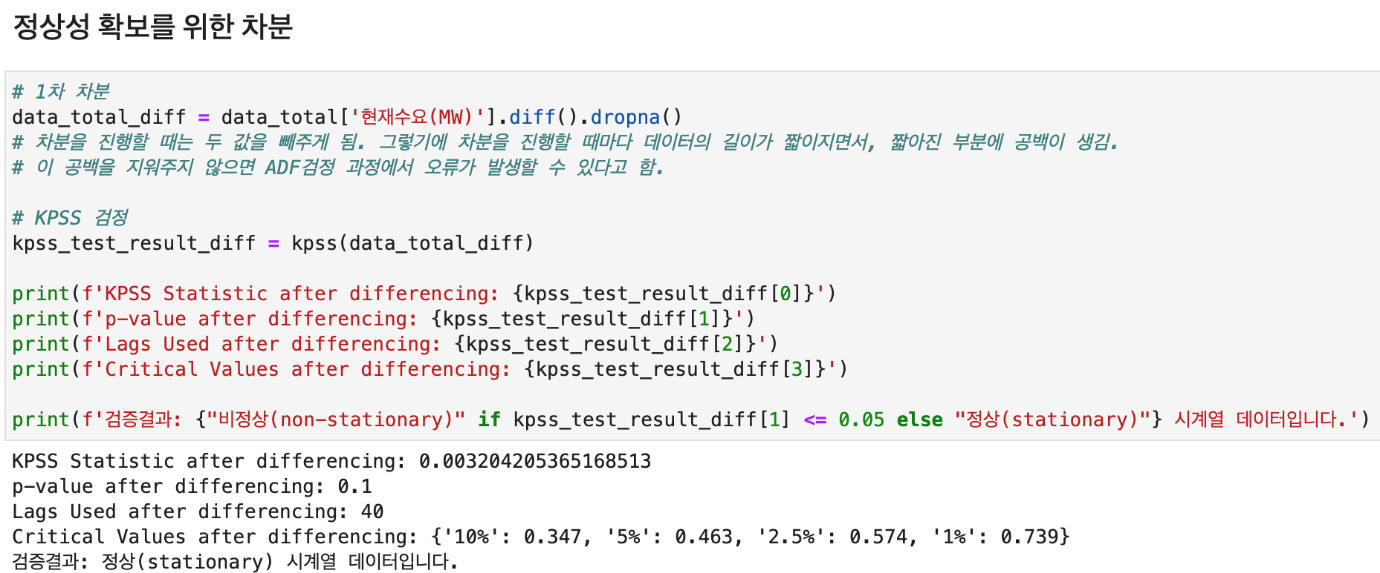
- H1: 해당 시계열은 비정상 시계열이다.

p-value <= 0.05 이면 H0 기각, H1 채택으로 비정상 시계열 데이터이다.

p-value > 0.05 이면 H0 채택, H1 기각으로 정상 시계열 데이터이다.



KPSS 검정 결과 p-value 0.01<=0.05로, 비정상 시계열 데이터이다. 비정상 시계열 데이터를 정상 시계열 데이터로 바꾸어 주기 위해 차분을 진행한다.



1차 차분한 데이터는 p-value 0.1>0.05로, 정상 시계열 데이터이다.

**2. Augmented Dickey-Fuller(ADF) 검정**

ADF 검정을 통해 시계열 데이터의 정상성을 확인하는 작업을 다시 수행한다.

95% 신뢰도를 바탕으로 0.05 유의수준을 두면

- H0: 해당 시계열은 비정상 시계열이다. (시계열에 단위근이 존재한다.)

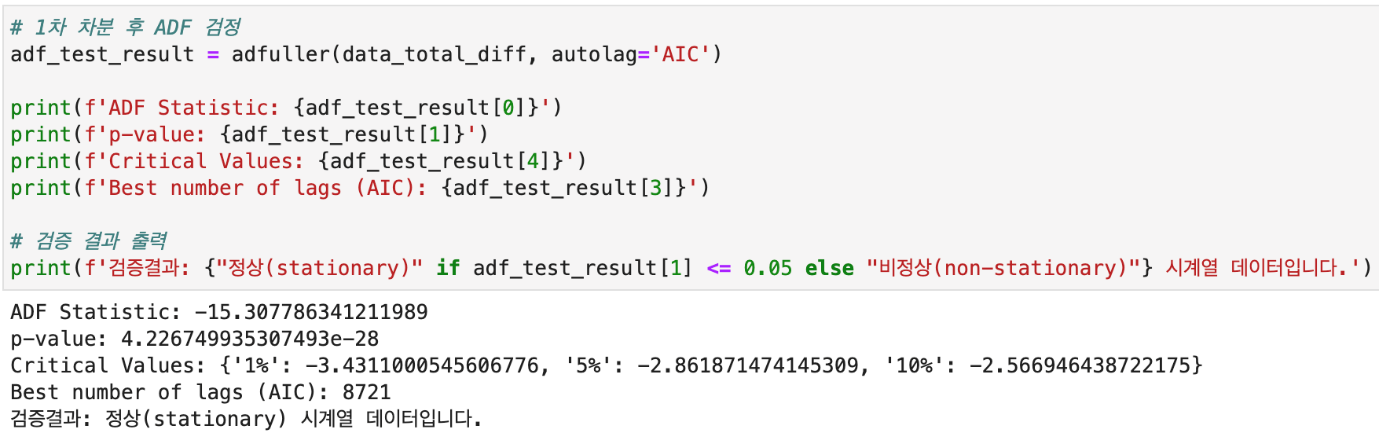
- H1: 해당 시계열은 정상 시계열이다.

p-value <= 0.05 이면 H0 기각, H1 채택으로 정상 시계열 데이터이다.

p-value > 0.05 이면 H0 채택, H1 기각으로 비정상 시계열 데이터이다.

KPSS 검정과 귀무가설이 반대이다.

1차 차분한 데이터에 대해 ADF 검정을 진행해보면,



p-value <= 0.05로, 정상 시계열 데이터이다.

**참고문헌**

송경빈, 문찬호 and 권보성. (2022). 한국 전력시스템의 240시간 전력수요예측에 대한 딥러닝 모델과 학습기법. 전기학회논문지, 71(4), 585-591.

김지은, 천관호. (개최날짜). LSTM을 활용한 단기 전력수요 예측기법. 대한전기학회 학술대회 논문집, 개최지.

권보성, 전재성, 공병철. (2023). 단기 전력수요예측이 한국의 전력시장 가격에 미치는 영향 분석. 대한전기학회 학술대회 논문집, 개최지.

차현종, 강아름. (2023). 기상정보를 활용한 머신러닝 기반의 전력수요 예측 모델. 한국콘텐츠학회논문지, 23(2), 117-124, 10.5392/JKCA.2023.23.02.117

김형욱, 「날씨 따라 바뀌는 전력수요 예측…한전, 기상청 빅데이터 공유 확대」, 『이데일리』 2022.07.12, https://m.edaily.co.kr/news/Read?newsId=03289846632393864&mediaCodeNo=257&utm\_source=https://www.google.com/(2023-11-03 접속).

부형권, 「심야전력 수요예측 잘못 한전 125억 경영손실」. 『동아일보』 2009-09-21, https://www.donga.com/news//article/all/20010204/7643869/1 (2023-11-03 접속).

성수영, 「전력수요 예측은 왜 번번이 틀리나」, 『생글생글』 2018.08.20, https://sgsg.hankyung.com/article/2018081702641(2023-11-03 접속).

편집팀, 「[에릭인사이트] 4차 산업혁명에는 ‘전력’이 핵심이다」, 『전기신문』 2021.06.03, https://www.electimes.com/news/articleView.html?idxno=218285 (2023-11-06 접속).