# 페이지 수 적어놓기

#1 ===========================

안녕하십니까. 빵빵이들 Term Project 최종 발표를 맡은 정경서입니다.

발표 시작하겠습니다.

저희 팀은 지속가능한 전력 생산을 주제로 프로젝트를 진행했습니다.

~~지속가능한 전력 생산을 위해서는 화력 발전을 대체할 수 있는 친환경 ….~~

~~대체 에너지 자원 정확히 예측 -> 화석에너지를 줄여야 한다….~~

#2 ===========================

발표는 다음과 같은 목차 순으로 진행하겠습니다.

#3 ===========================

**// 데이터 설명**

먼저 데이터 설명입니다.

저희 팀이 데이터 분석을 위해 수집한 데이터는 총 3가지입니다.

1번째, 풍력/태양광 발전량에 해당하는 친환경 에너지 발전 데이터와,  
2번째, AWS(즉, 기상 정보 데이터),  
3번째, 전력 수요 데이터입니다.

기간은 20, 21, 22년도 1월부터 12월까지로 총 3년치이며/ 범위는 전국/ 단위는 한시간입니다.

#4 ===========================

**// 분석 기법 선택**

저희는 데이터 예측에 있어 2가지 데이터 분석 기법을 이용하고자 했는데요.

1번째, 기계학습 분석 방법과 2번째, 시계열 분석 방법입니다.

#7 ===========================

**// 풍력&태양광 – 전처리**

먼저 ~~풍력&태양광 발전 데이터에 해당하는~~ 친환경 에너지 분석입니다.

# ===========================

**데이터 분석을 진행하기에 앞서 데이터 전처리 과정에 대해 말씀드리겠습니다.**

#8 ===========================

먼저 풍력/태양광 데이터 모두 위치 데이터 전처리 과정을 거쳤는데요.

**501개의** 지점에서 관측된 “시간별 기상 요소 데이터”와 & **501개의** “관측소 위치 데이터”//

이 2가지 데이터를 **풍력/태양광 발전 지역 10개로** 줄이는 작업을 거쳤습니다.

#9 ===========================

**그리고 특별히 정확도를 높이기 위해, 풍력/태양광 데이터에 추가적인 전처리 과정을 시행했습니다.**

먼저 풍력의 경우, 미가공 상태인 풍향을 16개로 카테고리화 하는 논문을 찾을 수 있었는데요.

이 논문을 바탕으로 저희는 One-Hot Encoding을 이용해서 풍향을 16개 카테고리화 하는 전처리 과정을 거쳤습니다.

다음으로 태양광 발전량 그래프에서 유의미한 변곡점을 기준으로 하여 08~20시로 시간을 제한하는 데이터 전처리 과정을 거쳤습니다.

또한 태양 고도가 태양광 발전량에 영향을 준다는 논문을 찾을 수 있었는데요

이 논문을 바탕으로/ 저희 팀이 수집한 데이터에 피어슨 상관관계 분석을 진행한 결과, 태양 고도와 태양광 발전량에 높은 상관관계가 존재함을 확인했습니다.

때문에, 저희는 ~~(날짜, 시간, 위도, 경도 데이터를 input하면 태양 고도 데이터를 얻을 수 있는)~~ **Python Package 중 (파이쏠라|pysolar)**를 이용해서 태양고도 데이터를 수집하고, 이를 새로운 컬럼으로 추가했습니다.

그리고 풍력/태양광 데이터 모두에 지역 데이터의 One-Hot Encoding을 추가로 진행했습니다.

===========================

**// 태양광&풍력 – 머신러닝 분석 & 시계열**

위와 같은 데이터 전처리 과정 후,

기계학습 분석(머신러닝 분석)을 진행했습니다.

기계학습의 경우 중간보고와 같이 Decision Tree, Random Forest, Multi-Layer Perceptron을 사용하였고 Decision Tree를 제외한 두가지 모델은 Grid Search를 통해서   
각각 적절한 파라미터를 탐색한 후 예측 과정을 진행하였습니다.  
더불어, 중간 발표 피드백을 반영하여 시계열 분석까지 진행해보았습니다.

===========================

**// 전력 – 시계열 분석**

다음으로 전력 수요 데이터 분석입니다.

===========================

전력 수요 데이터의 경우 저희는 원본 데이터의 일일 총 수요량을 바탕으로 데이터를 분석하고자 했습니다.

전기 수요량의 경우, 시간에 따라 데이터가 나타나 있기 때문에, 시계열분석을 진행하고자 했고요.

시계열분석에도 다양한 분석 방법이 존재하는데, 전력 수요 데이터는 다음과 같은 3가지 조건을 만족하기에 ARIMA 시계열 분석을 진행하고자 했습니다.

ARIMA 분석 가능 여부를 판단하기 위해, 보이는 바와 같이 시계열 분해를 진행했습니다. 그 결과, 추세가 없고, 계절성과 주기성은 있는 것으로 파악되어 ARIMA 모델의 선행조건을 만족하는 것을 확인했습니다.

~~(PPT 자료) (1. 단기 예측, 2. 계절적 변동 요인, 3. 데이터 수가 50개 이상)~~ ~~따라서 이에 적합한 ARIMA 모델을 이용해 시계열 분석을 진행했습니다.~~

#13 ===========================

또한, ARIMA 모델의 또다른 선행조건인 데이터 정상성을 파악해 보았습니다.

여기서 시계열 데이터의 정상성이란, 시간에 따른 평균/분산/공분산이 일정한 상태를 말합니다.

#14 ===========================

하지만 원본 데이터로 ADF 및 KPSS 검정 결과, 해당 데이터는 비정상 시계열이라는 결과를 얻었습니다.

저희는 정상성을 만족시키기 위한 방법으로 차분을 이용했습니다. 차분이란, 연 이은 관측값들의 차이를 계산하는 것으로 데이터의 추세와 계절성을 제거하는 방법인데요. 즉, 평균 변화를 일정하게 만드는 방법인거죠. VVVVV -> vvvvv

~~차분(differencing)은 시계열의 수준에서 나타나는 변화를 제거하여 시계열의 평균 변화를 일정하게 만드는데 도움이 될 수 있습니다.~~

이러한 1차 차분을 적용한 결과, 두 검정에서 정상성을 만족하게 됐으므로, 시계열분석을 진행하였습니다.

#15 ===========================

(@PPT: plot) 이로써 얻은 ARIMA 모델을 바탕으로

===========================

다음과 같이 미래 전력 수요량 예측을 진행할 수 있었습니다.

===========================

더불어, 전력 수요 예측에서도 중간 발표 피드백을 반영하여 기계 학습 분석까지 진행해보았습니다.

#16 ===========================

**// 최종 결과**

최종 결과입니다.

===========================

각 데이터에 대해 RMSE와 MAPE를 바탕으로 어떤 분석 방법이 더 적절한지 파악하고자 했습니다.

여기서 RMSE는 평균 제곱근 오차로 **(실제 값과 예측 값 간의 차이)**를 의미합니다.

그리고 MAPE는 실제 값과 예측 값 간의 **백분율 오차**를 의미합니다.

즉, 두 평가 지표 모두 값이 작을수록 예측 모델의 성능이 더 좋다고 할 수 있겠죠.

이러한 2가지 평가지표 값이 작게 나오는 부분을 확인했을 때,

태양광&풍력 발전 데이터의 경우 Random Forest와 MLP 분석 기법에서,

전력 수요 데이터의 경우 시계열 분석 기법에서 예측 모델의 성능이 좋음을 알 수 있었습니다.

#6 ===========================

더 나아가, 태양광과 풍력에 관해서는 시간에 따른 데이터 편차가 크므로 시간별로 추가 분석을 진행했는데요. 그 결과, Random Forest가 가장 좋은 지표를 보여 태양광&풍력 발전 데이터는 Random Forest 기법에서 예측 모델의 성능이 좋음을 알 수 있었습니다.

(공부 필요) 추가로, Random Forest와 MLP를 혼합하여 결과를 도출하는 **앙상블 기법을** 사용해보았지만, 기존 모델의 비하여 더 좋은 성능을 보여주지 못해서 **앙상블 기법은** 채택하지 않았습니다.

===========================

최종적으로 각각의 분석을 수행한 결과, 다음과 같이 각각의 지표에서 우수한 값을 보이는 모델을 선별할 수 있었습니다.

#17 ===========================

각 모델을 바탕으로 저희는 2023년 1월 1일에 필요한 화석 연료량을 예측해보았습니다.

총 전기 수요량은 ARIMA 시계열 분석을 통해 다음과 같은 값을 얻을 수 있었고,

풍력 및 태양광 발전량은 Random Forest 분석을 통해 다음과 같은 값을 얻을 수 있었습니다.

따라서 저희 팀은 2023년 1월 1일에 약 (1760만)만큼의 화석연료를 통한 에너지 생산량이 필요하다는 결론을 도출할 수 있었습니다.

#18 ===========================

**// 한계점 및 보완점**

마지막으로 한계점 및 보완점입니다.

1. 먼저 모~든 신재생 에너지 발전량을 고려한 것이 아니기 때문에 실제 화석연료 발전이 필요한 양을 예측했다고 보기에는 어렵다는 한계점이 있었음. 때문에, 다양한 에너지원을 바탕으로 전력을 생산하는 현실을 추가 고려할 필요가 있다고 생각.
2. 저희 분석에서 Grid Search를 통해 탐색한 하이퍼 매개변수들은 Random Forest의 깊이와 MLP의 max\_iter (max iteration) 밖에 없었음. 높은 정확성을 위해서는 다양한 하이퍼 매개변수에 대해서 탐색을 진행해야 하기 때문에, 이 부분이 한계점이라 생각
3. 전력 수요 시계열의 경우, 전력 수요가 높은 시간대를 따로 분리해서 분석하면/ 더 좋은 결과를 얻을 수 있을 것이라 기대되기 때문에 이를 보완점으로 생각할 수 있었음.
4. 풍력/태양광 발전 데이터 분석에서 앙상블 기법으로 Voting Regressor를 사용해 보았지만, 더 좋은 결과를 얻을 수 없었습니다. 이는 저희의 앙상블에 대한 이해가 부족하기 때문이라고 생각했고, 그렇기에 다양한 기법을 수행하지 못한 점을 한계점으로 생각할 수 있었음.

(마지막 화면) ===========================

이상 팀 빵빵이들 발표 마치겠습니다. 감사합니다.