ARIMA & Prophet모델비교 2차

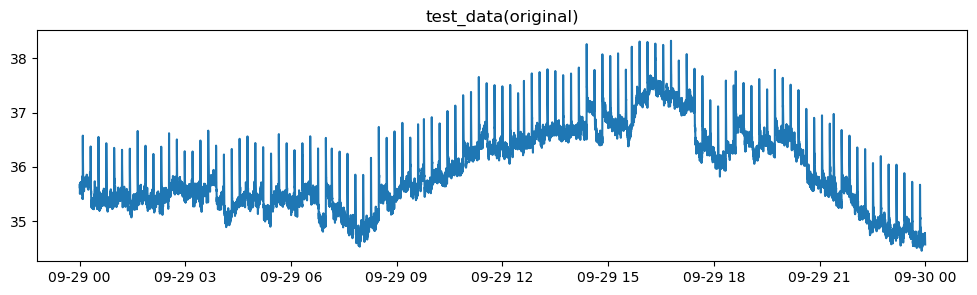
# 개요

* [ARIMA & Prophet 모델비교 1차]에서 나온 결과를 토대로 실제 열화상센서데이터를 사용하여  
  두 모델의 성능을 비교해보고자 함
* 기존에는 ARIMA모델의 특성 때문에 5초단위로 수집된 데이터를 1분간격으로 resampling하여   
  예측을 진행하였는데 Prophet은 resampling없이도 예측이 잘되는지 실험해보고자 함

# 모델비교1차에서 확인한 모델 별 특징

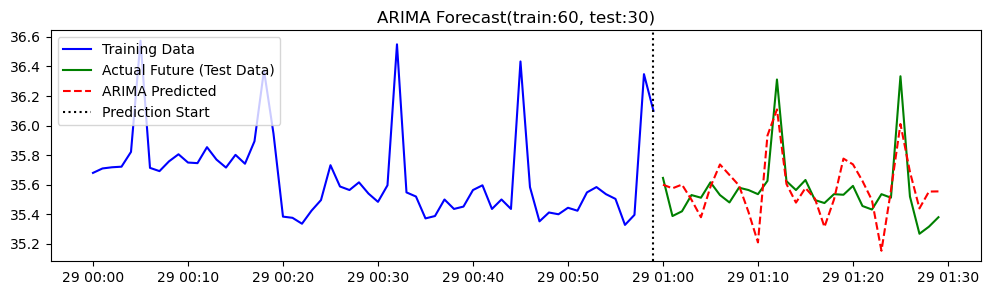
|  |  |  |
| --- | --- | --- |
|  | ARIMA | Prophet |
| 장점 | 적은 데이터로 예측가능  예측에 소요되는 평균시간이 전반적으로 빠름 | 노이즈, 추세를 반영해도 예측이 가능함 |
| 단점 | 학습데이터가 길어진다면 추세를 예측할 수 없음  노이즈가 추가된다면 예측이 불가능해짐 | 데이터가 적을 때 학습시간이 오래걸림  (데이터가 충분하면 ARIMA보다 빠름) |

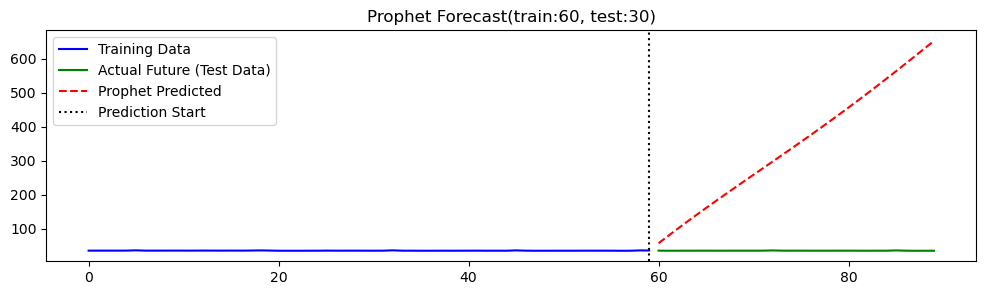
# 모델 성능 비교



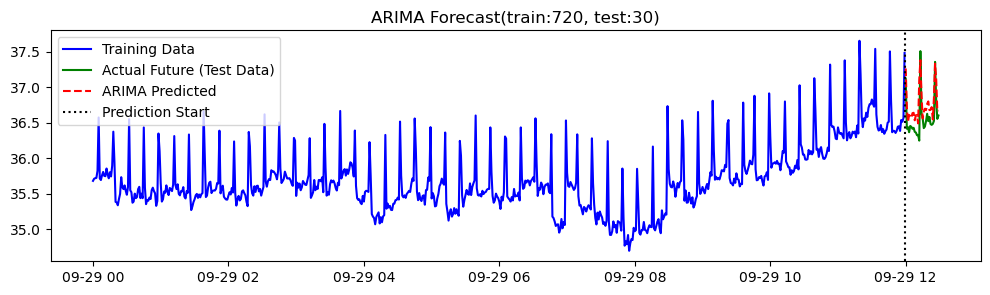
.

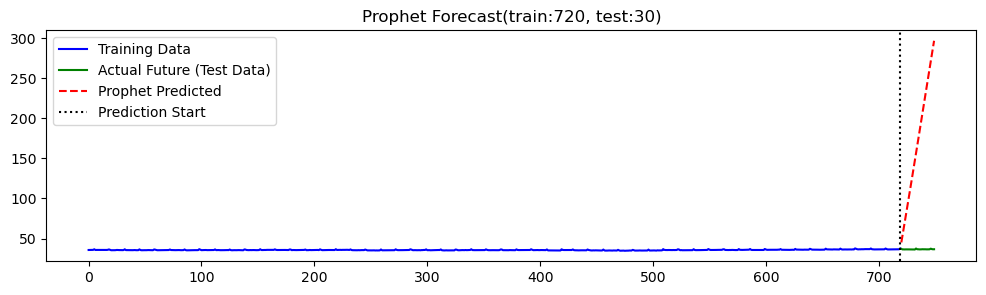
* 다양한 조건에서의 성능을 비교해보고자 현재 사용중인 resampling된 데이터를 통한 모델 성능 비교를 한 후 original데이터에 대한 모델성능 분석을 해보고자 함
* 또한 prophet의 강점인 추세예측이 어느정도 데이터가 요구되는지 확인해보고자 함

**1. resampling데이터를 통한 모델 비교(기존 프로젝트에 사용되는 데이터)1분간격 60개학습 30개예측**



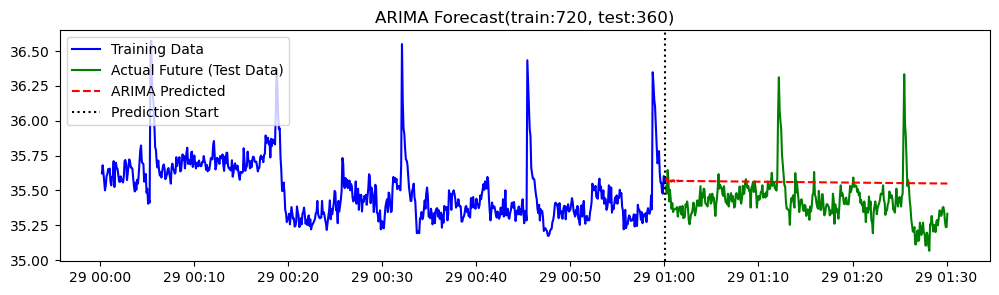
|  |  |  |
| --- | --- | --- |
|  | ARIMA | Prophet |
| 파라미터 | p : 12, d : 2, q : 6 | Yearly, Weekly, Daily바꿔가면서 확인 |
| 소요시간 | 4.9초 | 0.77초 |

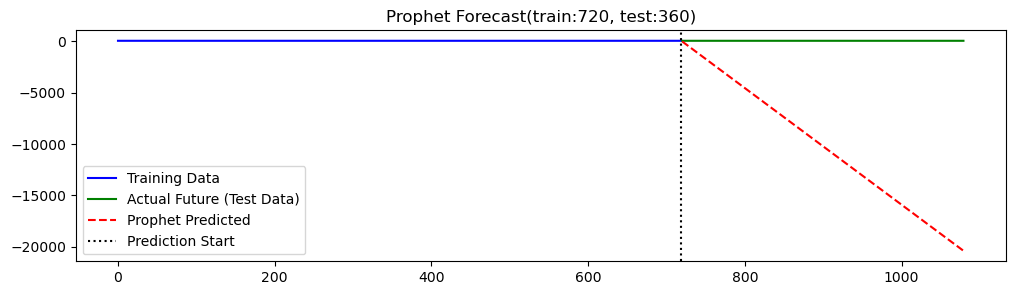
**2. resampling데이터를 통한 모델 비교(학습데이터 추가)1분간격 720개학습 30개 예측**



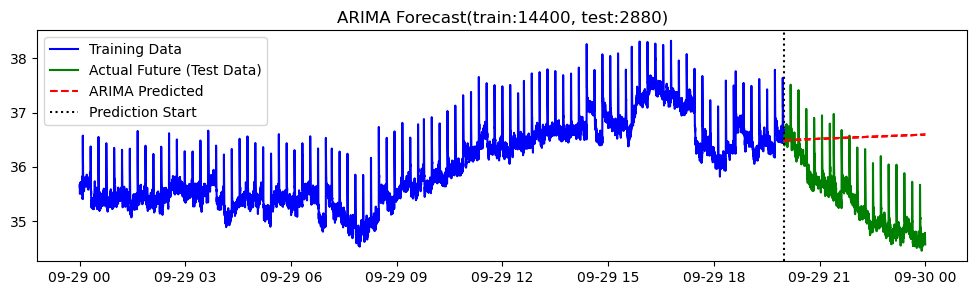
|  |  |  |
| --- | --- | --- |
|  | ARIMA | Prophet |
| 파라미터 | p : 12, d : 2, q : 6 | Yearly, Weekly, Daily바꿔가면서 확인 |
| 소요시간 | 6.2초 | 0.59초 |

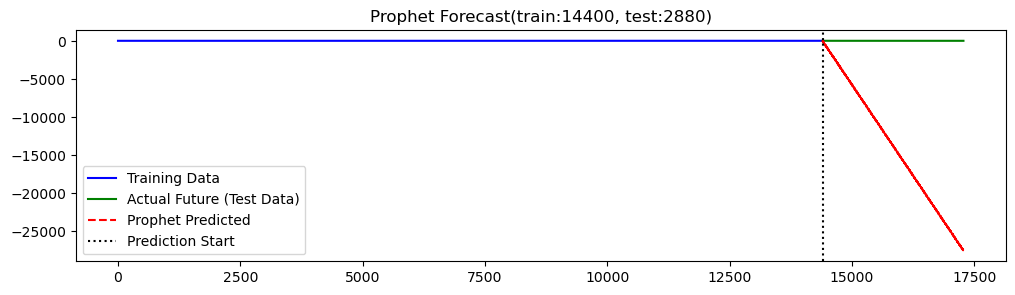
기존처럼 1분간격의 데이터로 예측해본 결과 ARIMA는 추세와 패턴을 잘 예측하지만 Prophet의 경우 학습데이터와 관계없는 예측결과가 나옴, 또한 학습데이터를 충분히 추가하여도 성능이 그대로임

**3. original데이터를 통한 모델 비교5초간격 720개학습 360개예측**



|  |  |  |
| --- | --- | --- |
|  | ARIMA | Prophet |
| 파라미터 | p : 12, d : 2, q : 6 | Yearly, Weekly, Daily바꿔가면서 확인 |
| 소요시간 | 4.7초 | 0.67초 |

**4. original데이터를 통한 모델 비교5초간격 14400개학습 2880개예측**



|  |  |  |
| --- | --- | --- |
|  | ARIMA | Prophet |
| 파라미터 | p : 12, d : 2, q : 6 | Yearly, Weekly, Daily바꿔가면서 확인 |
| 소요시간 | 35.26초 | 13.72초 |

5초간격의 original데이터로 1시간 학습하여 30분을 예측해도 Prophet은 여전히 예측을 못함

# 모델 성능 비교 결과

기존 목표였던 1시간 데이터를 학습하여 30분 후를 예측하는 성능을 확인하기 위해 열화상 데이터로 확인해본 결과, Prophet은 계절성과 추세를 기반으로 예측하는 모델이기 때문에 프로젝트에서 요구하는 짧은 구간의 패턴이 반복되는 형태의 데이터를 예측하기에는 적합하지 않은 모델 같다.   
또한 데이터를 추가해 24시간데이터를 학습하여 30분을 예측해보았지만 해당 데이터에서는 추세나 계절성 학습을 학습하기 어려워 예측 성능이 기대만큼 좋지 않음을 확인할 수 있다.

ARIMA의 경우 데이터의 차분을 활용하여 추세변화를 반영하면서도 계속해서 반복적으로 변화하는(12분마다 1번씩 값이 상승) 패턴을 학습하여 예측하는데 좋은 예측 성능을 가지고 있어  
열화상 프로젝트에는 기존의 데이터를 활용해 ARIMA모델을 적용하는 것이 적절해 보인다.

이번 비교에서 사용한 온도데이터뿐만 아니라 압력데이터나 FLOW계통의 데이터를 사용한다하더라도  
기존의 목표인 1시간데이터로 30분데이터를 예측하는 것은 학습시간이 짧아 추세나 계절성학습이 어려워 Prophet모델 사용은 어려울 것으로 판단된다.