

Web Traffic Time Series Forecasting

Forecast future traffic to Wikipedia pages

<http://www.ewha.ac.kr>



NAME

이영주 임소현 이은진

Department

이화여자대학교 통계학과



INDEX



SUBJECT & DATA

MODEL



RESULTS

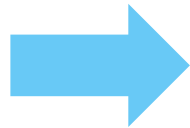
CODE





약 14만개 정도의 데이터셋

시계열 데이터이므로 결측치가 없는 200개 데이터만 추출



사이트의 일별 조회수 예측





2015년 7월 1일 ~ 2016년 12월 31일
위키피디아 기사의 일별 조회수 데이터

	2015-07-01	2015-07-02	2015-07-03	2015-07-04	2015-07-05	2015-07-06	2015-07-07	2015-07-08	2015-07-09	2015-07-10
0	18.0	11.0	5.0	13.0	14.0	9.0	9.0	22.0	26.0	24.0
1	11.0	14.0	15.0	18.0	11.0	13.0	22.0	11.0	10.0	4.0
2	35.0	13.0	10.0	94.0	4.0	26.0	14.0	9.0	11.0	16.0
3	118.0	26.0	30.0	24.0	29.0	127.0	53.0	37.0	20.0	32.0
4	5.0	23.0	14.0	12.0	9.0	9.0	35.0	15.0	14.0	22.0

5 rows × 550 columns

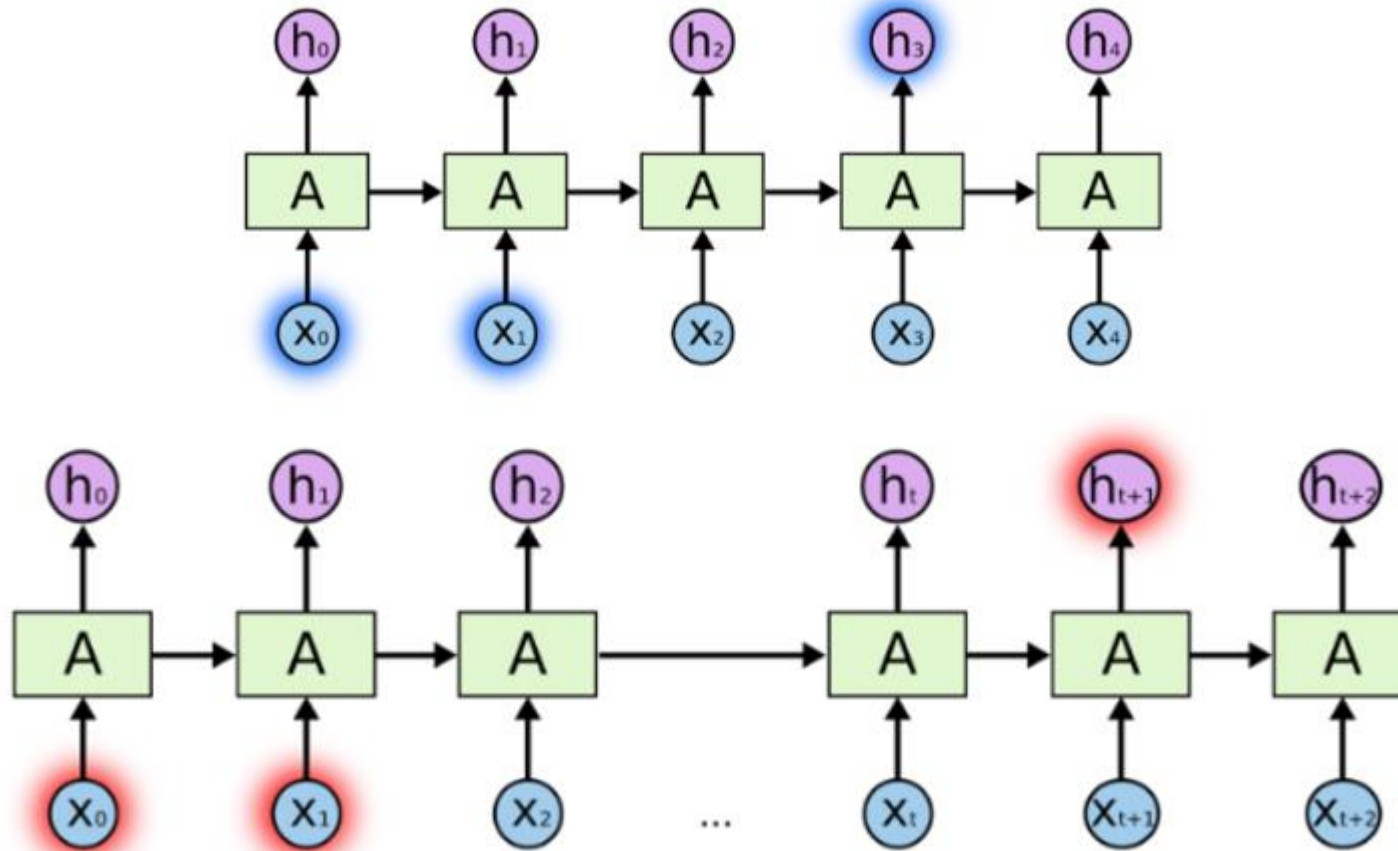




About Model

RNN & LSTM

Web Traffic Time Series Forecasting Forecast future traffic to Wikipedia pages



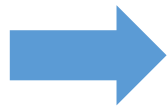
관련 정보와 그 정보를 사용하는 지점 사이 거리가 멀 경우 RNN 학습능력 저하



Example

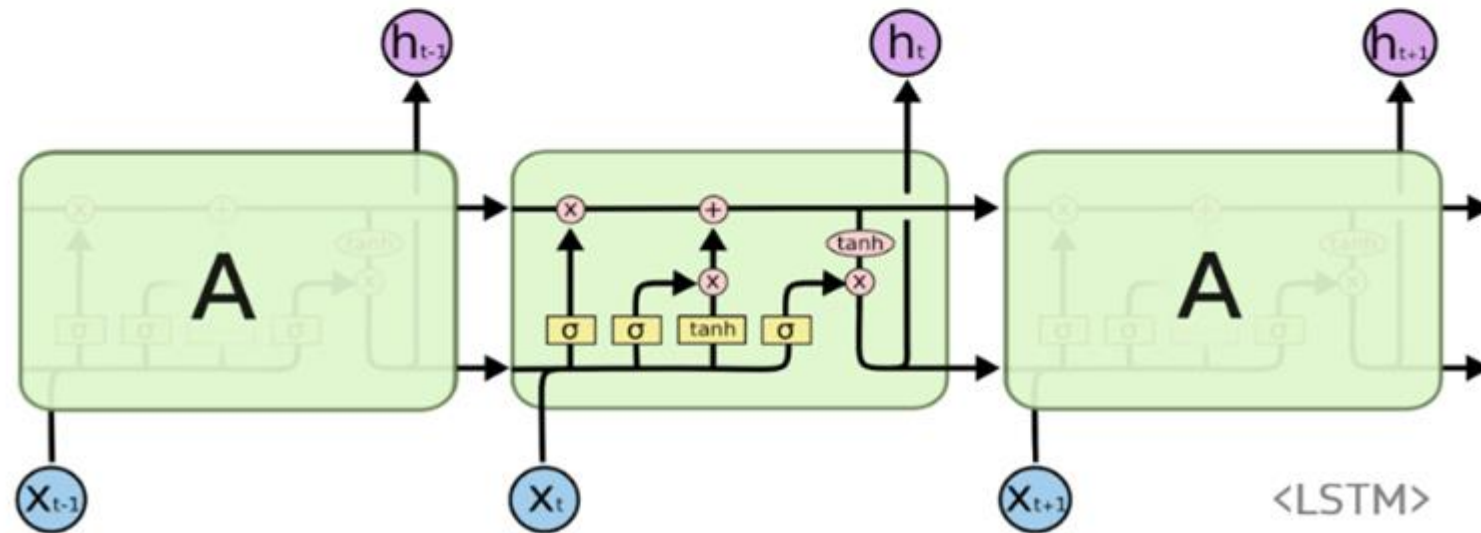
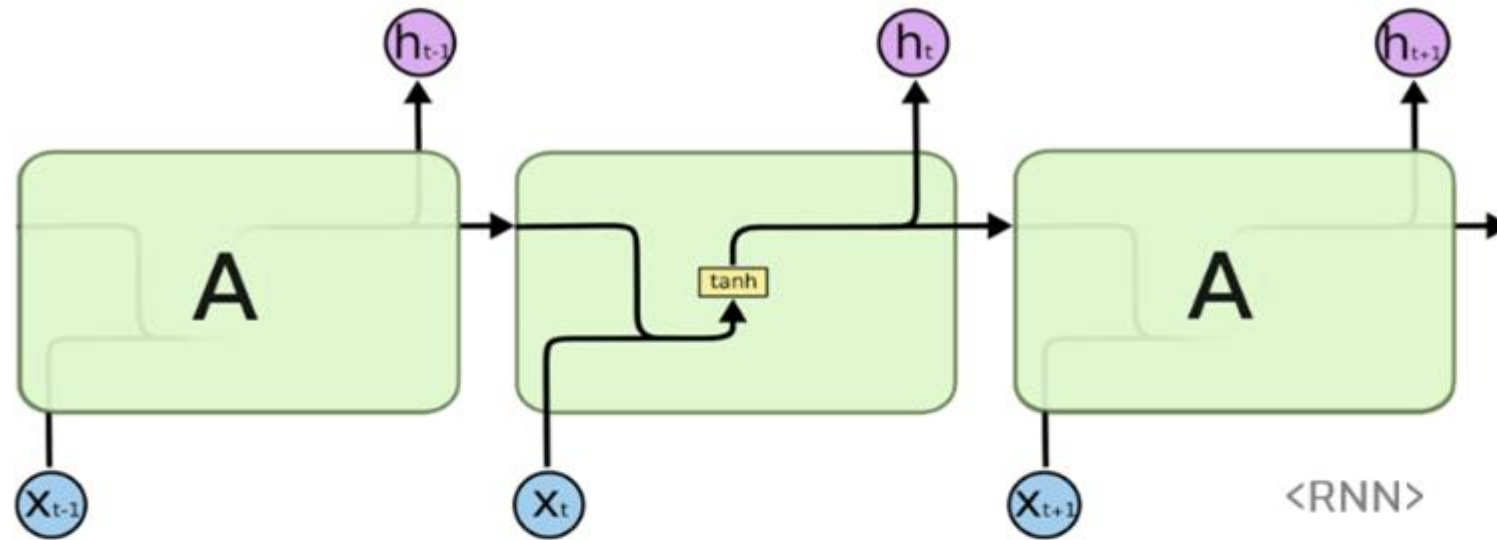
The clouds are in the sky.

I grew up in France. I speak fluent French.



이런 식으로 문장표현의 순서상 갭이 커질수록,
RNN은 두 정보의 문맥을 연결하기 힘들다.

LSTM





Results

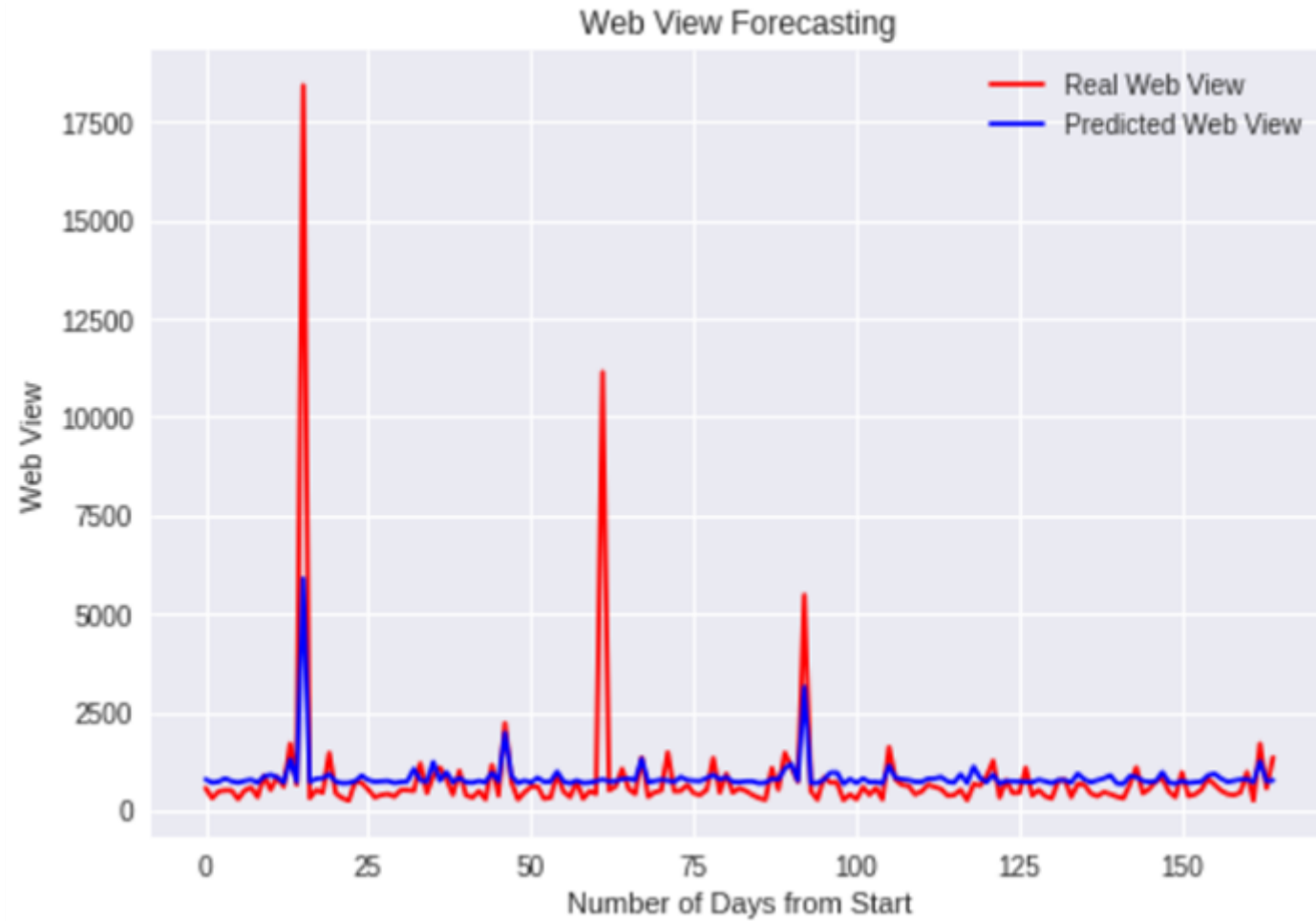
4 examples

RESULTS



1) Train과 Test로 나누어 (7:3) 예측

Irina Shayk (러시아 슈퍼모델) 기사

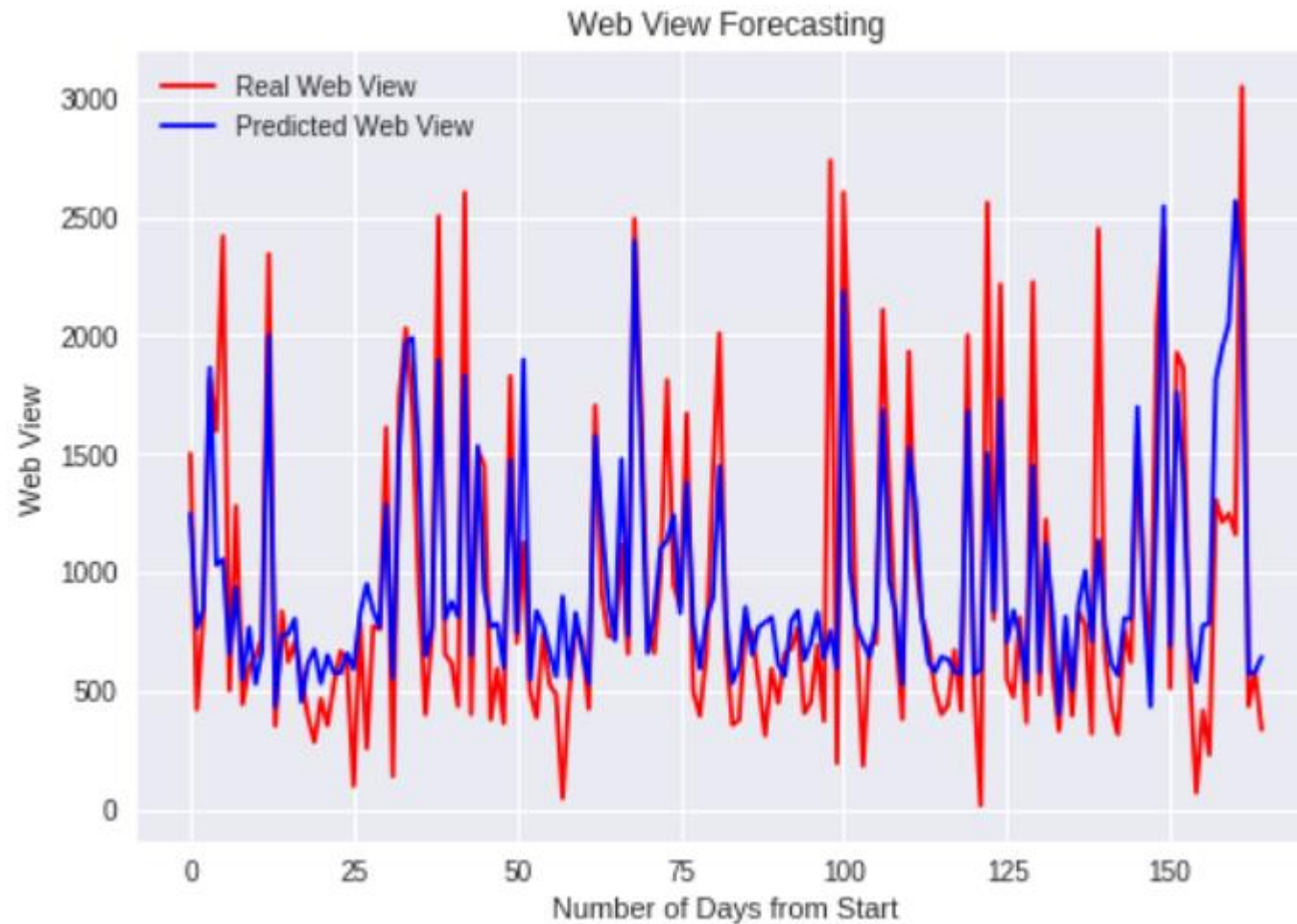


RESULTS



1) Train과 Test로 나누어 (7:3) 예측

Delvin Ndinga (축구선수) 기사

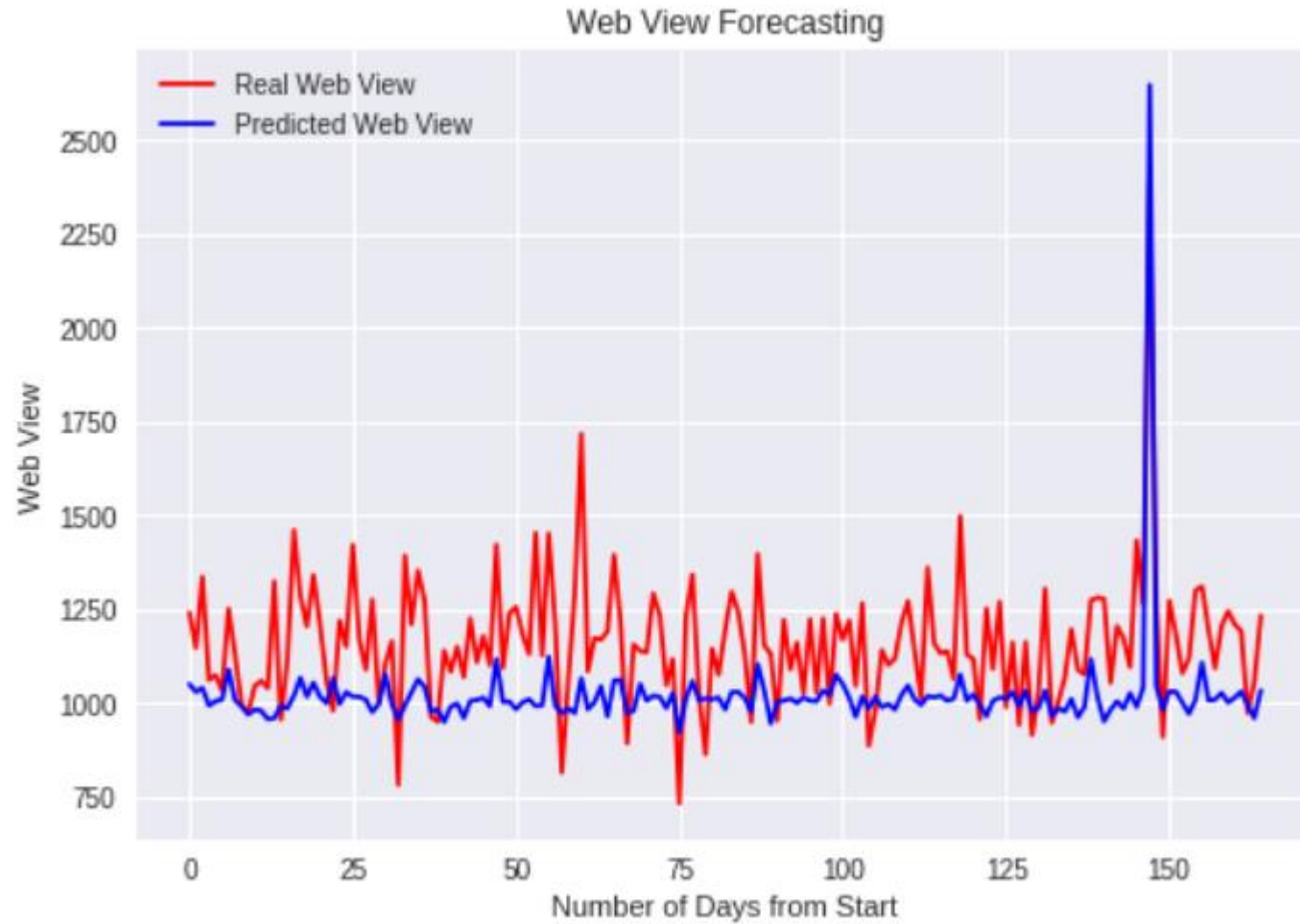


RESULTS



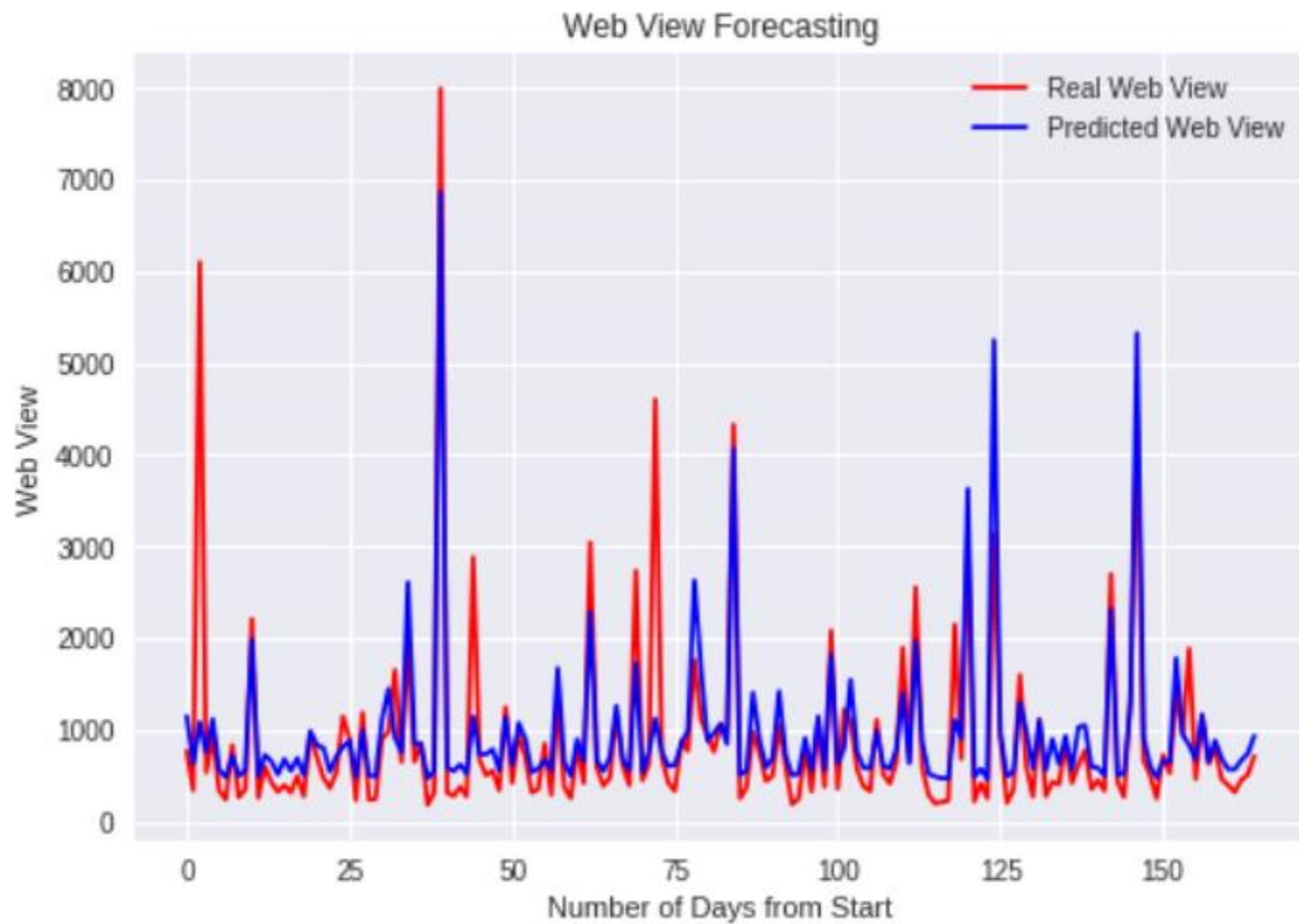
1) Train과 Test로 나누어 (7:3) 예측

Mainz (독일의 도시) 기사



RESULTS

1) Train과 Test로 나누어 (7:3) 예측

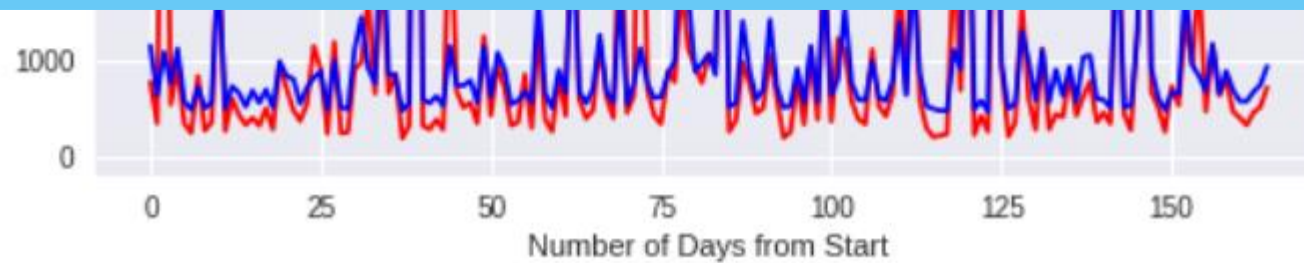


RESULTS

1) Train과 Test로 나누어 (7:3) 예측



예측을 잘 못하는 이유 → 시계열 데이터가 깨지기 때문

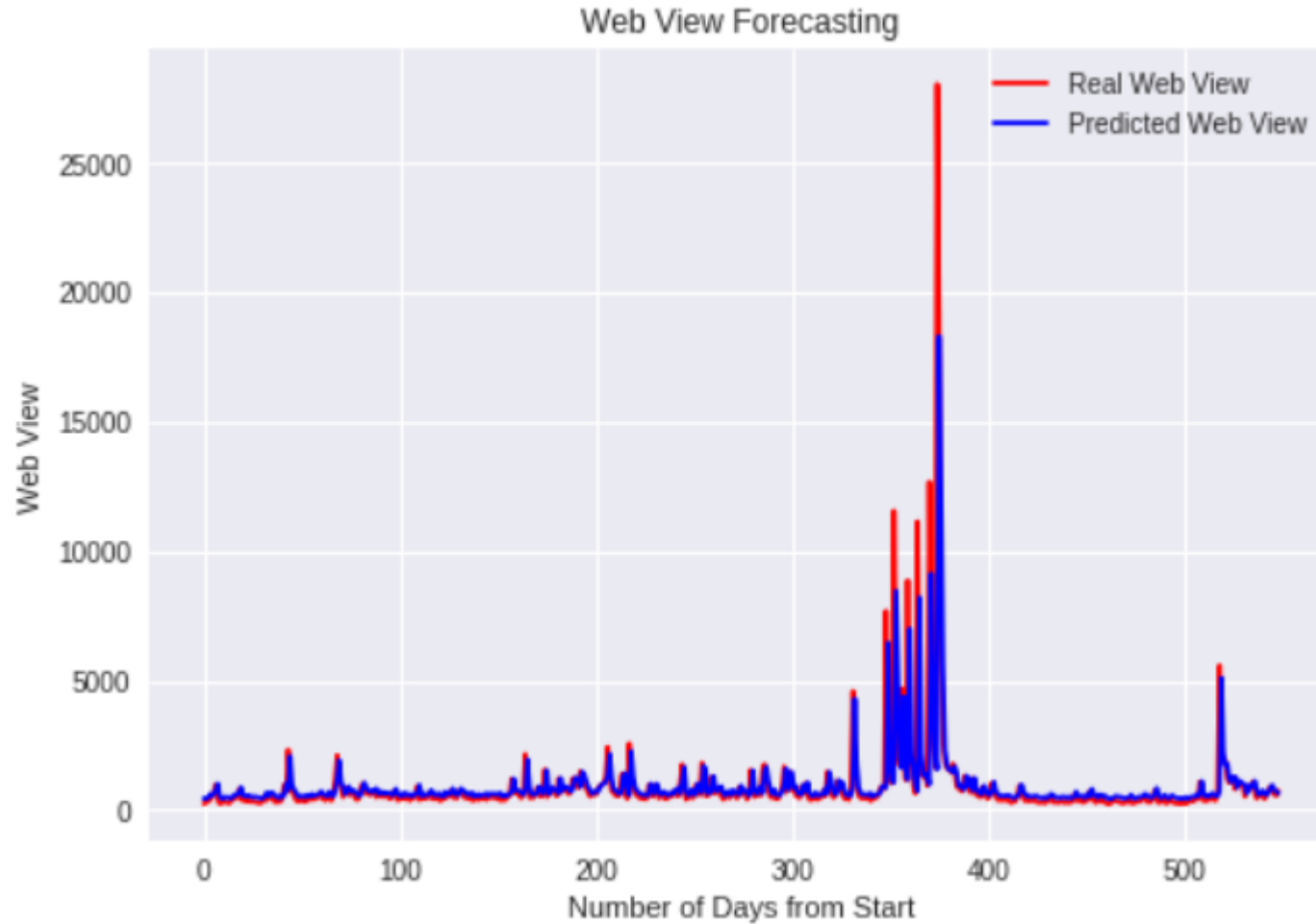


RESULTS



2) 모델 fitting 후 예측

Train : Delvin Ndinga (축구선수) 기사
Test : Irina Shayk (러시아 슈퍼모델) 기사

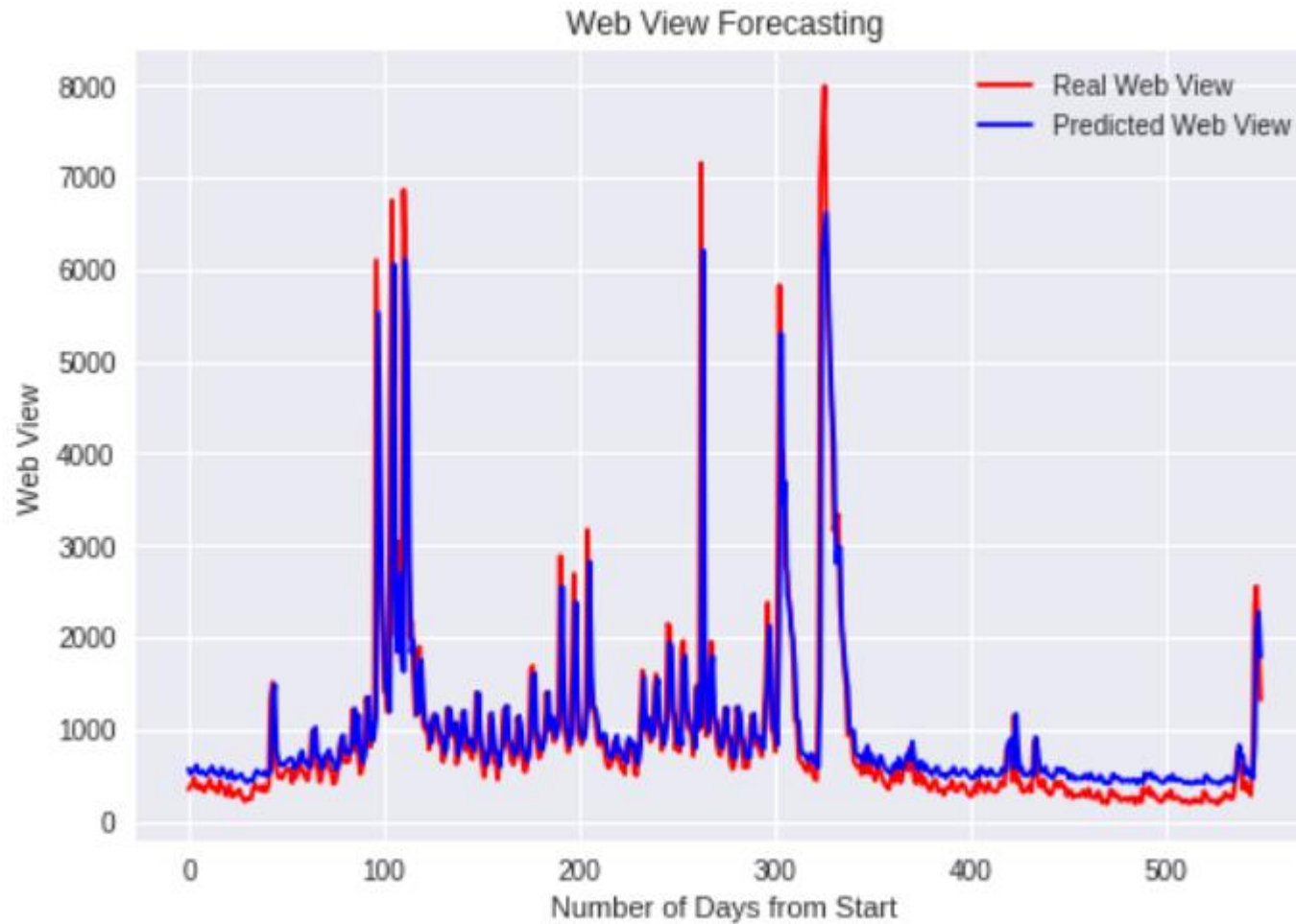


RESULTS



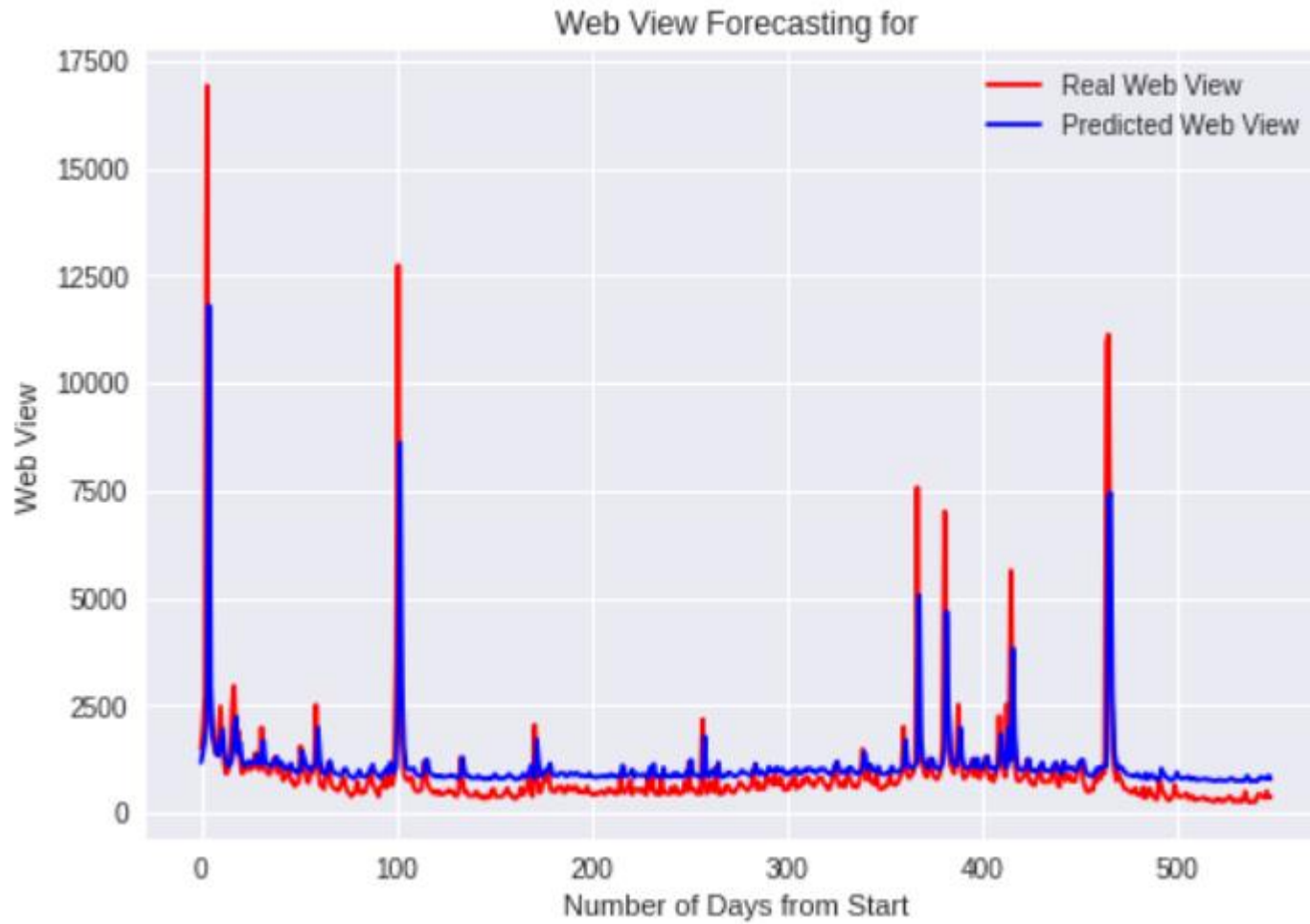
2) 모델 fitting 후 예측

Train : Delvin Ndinga (축구선수) 기사
Test :



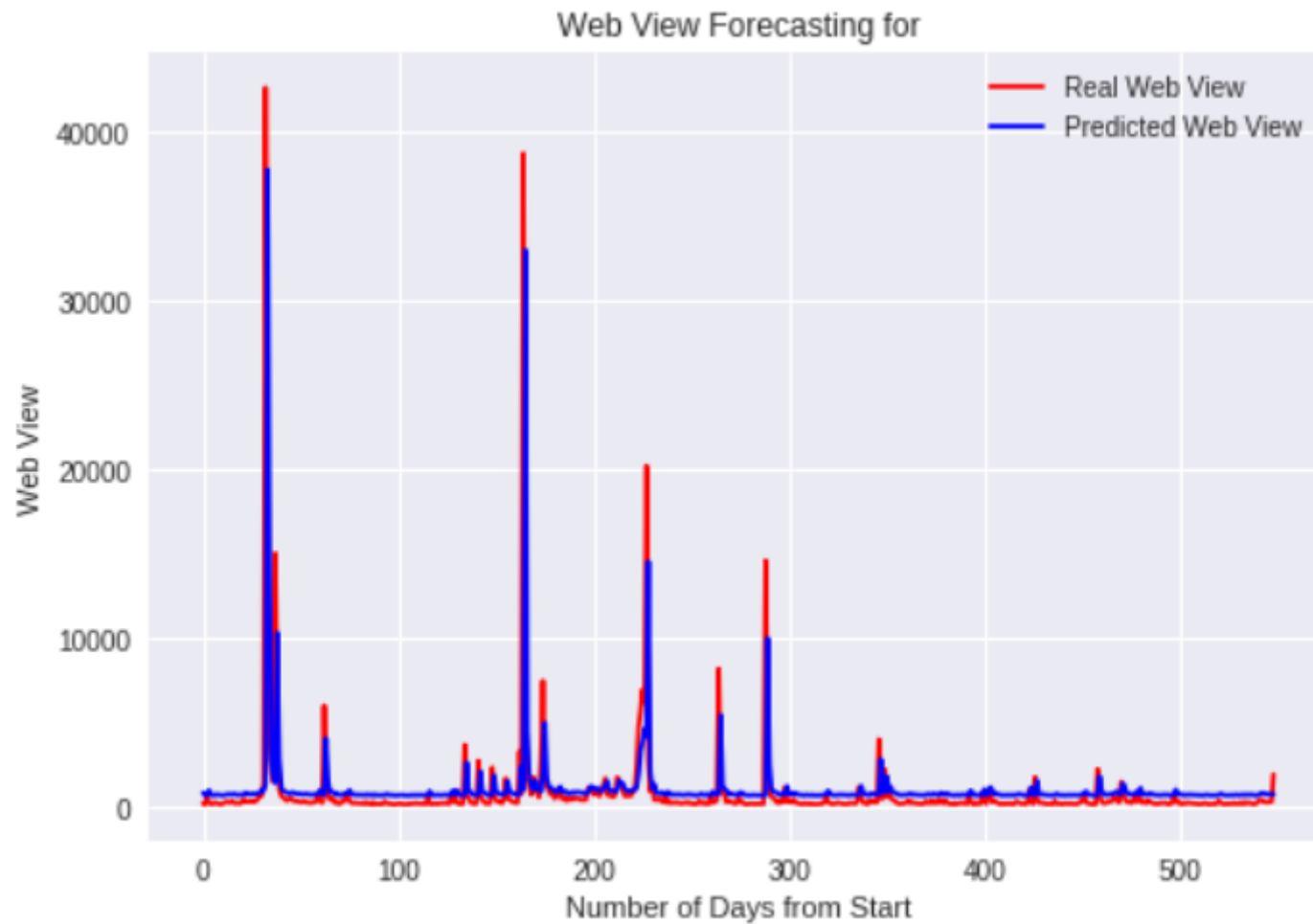
RESULTS

3) 일반화 : 모델 fitting 후 예측



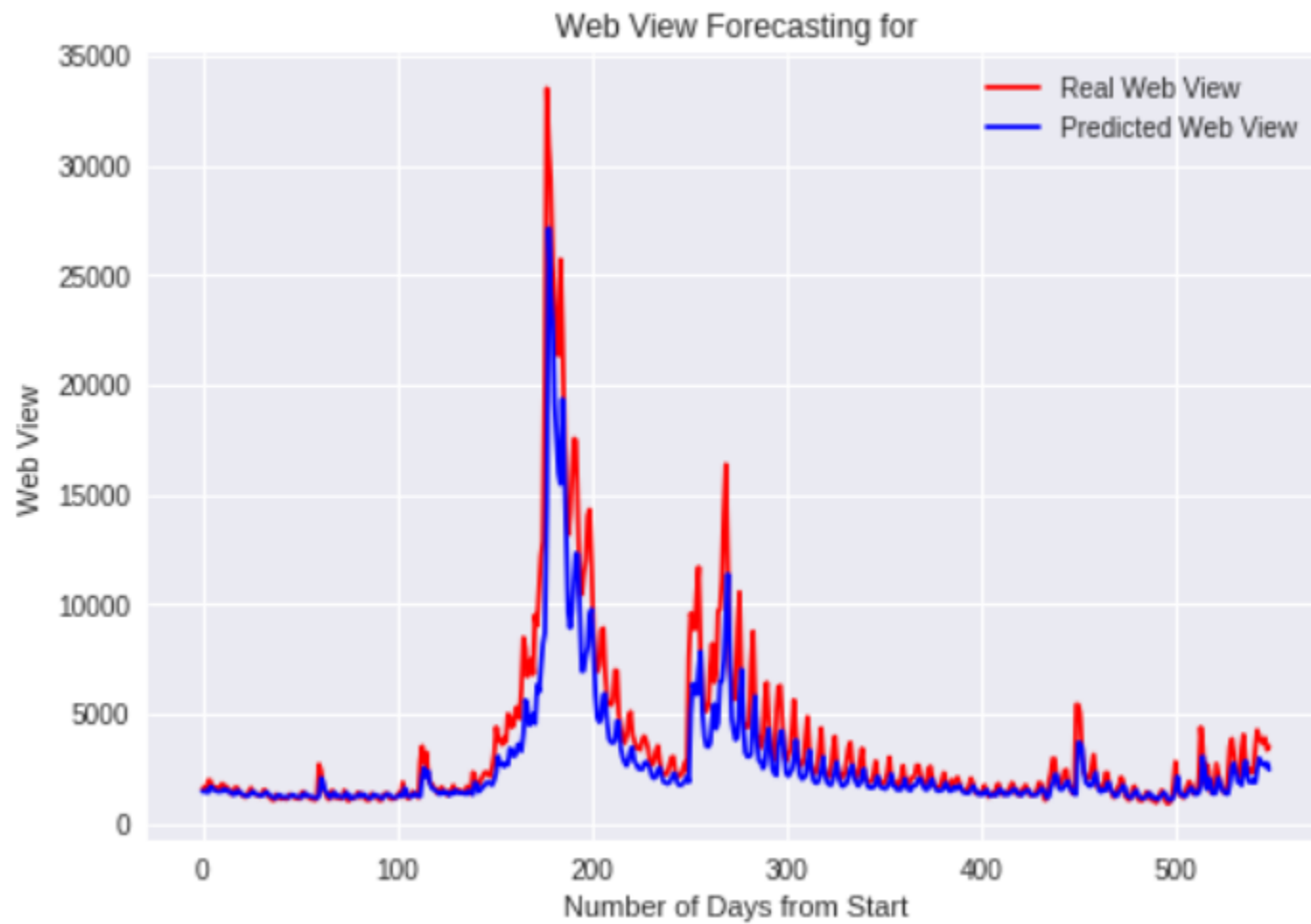
RESULTS

3) 일반화 : 모델 fitting 후 예측



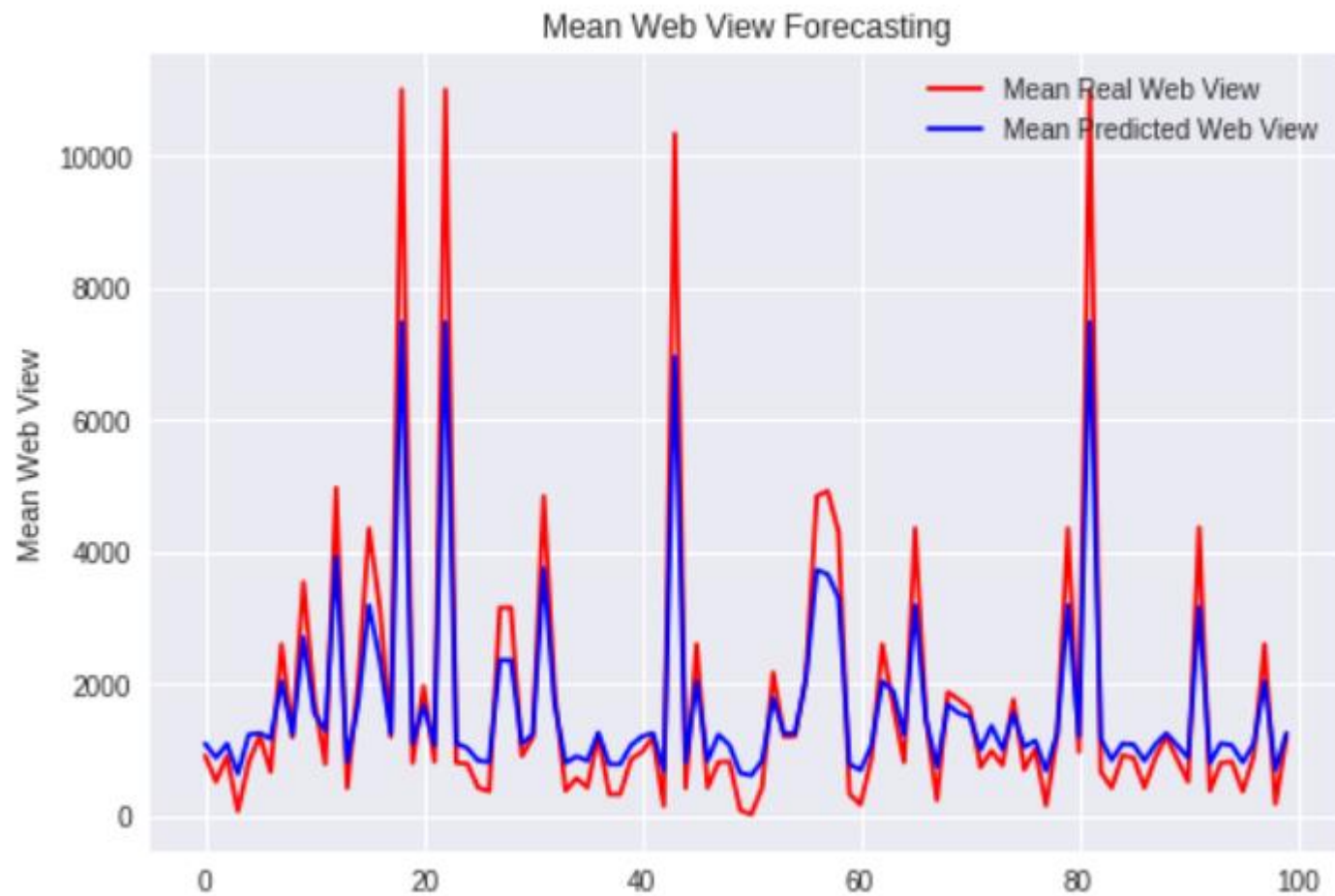
RESULTS

3) 일반화 : 모델 fitting 후 예측



RESULTS

4) 100개 데이터의 평균 예측





CODE
with Colab (feat. 영주)

Web Traffic Time Series Forecasting Forecast future traffic to Wikipedia pages