

2019년 2학기 컨퍼런스

Text2Moji

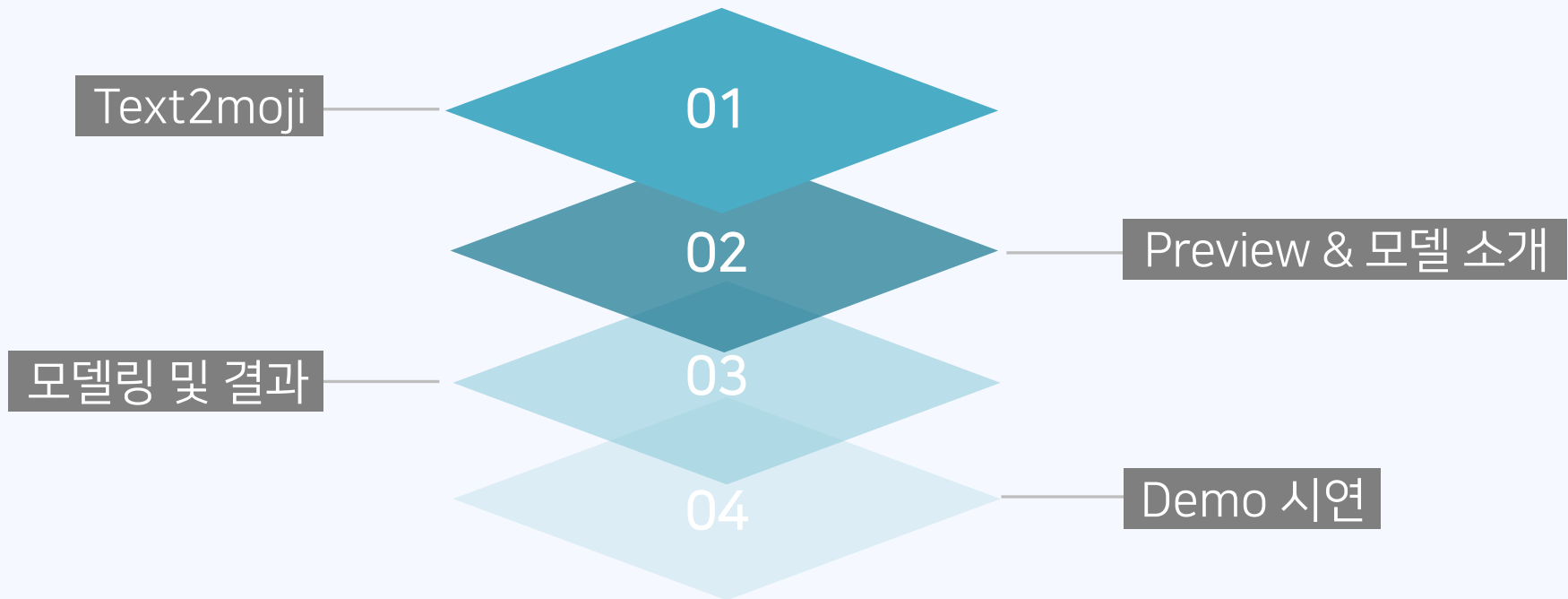
Final Presentation

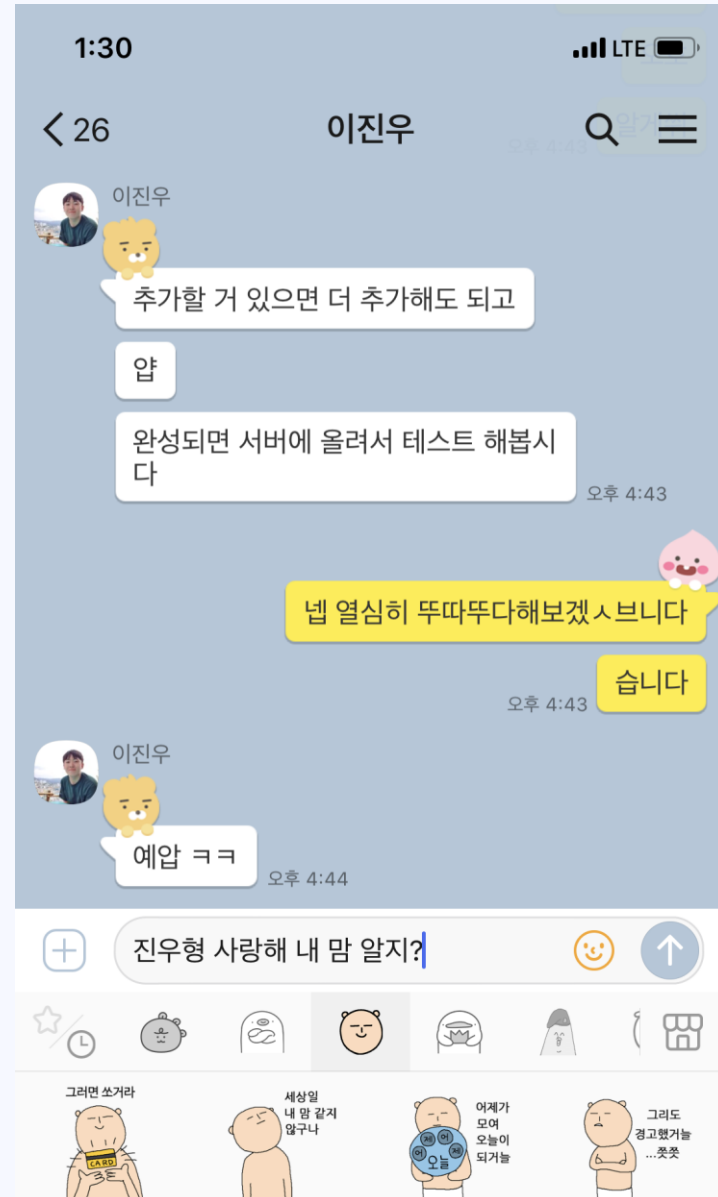


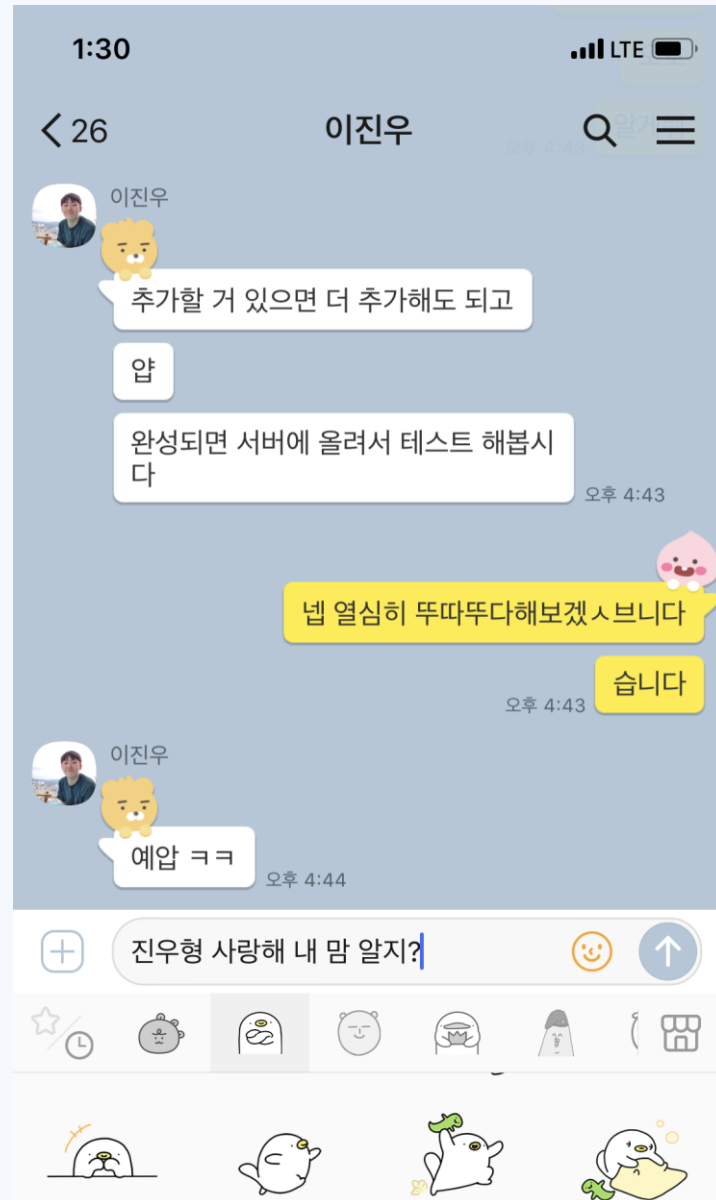
이진우 강소영 김용우
박현지 염정운

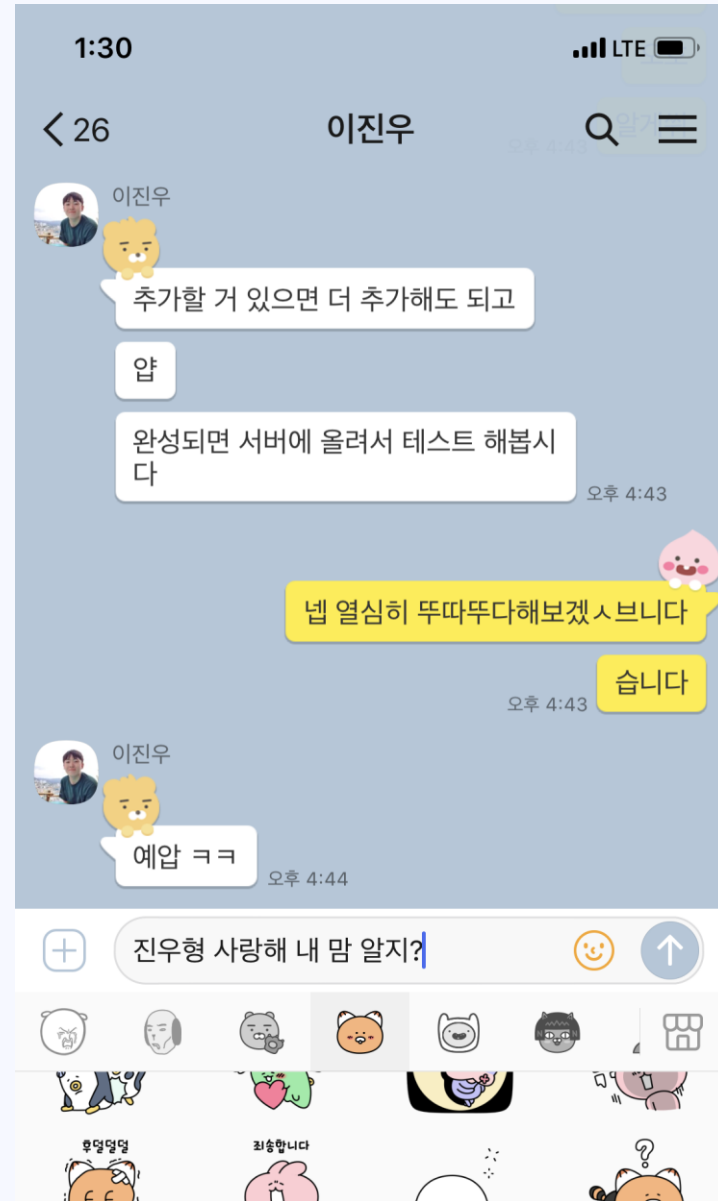


CONTENTS



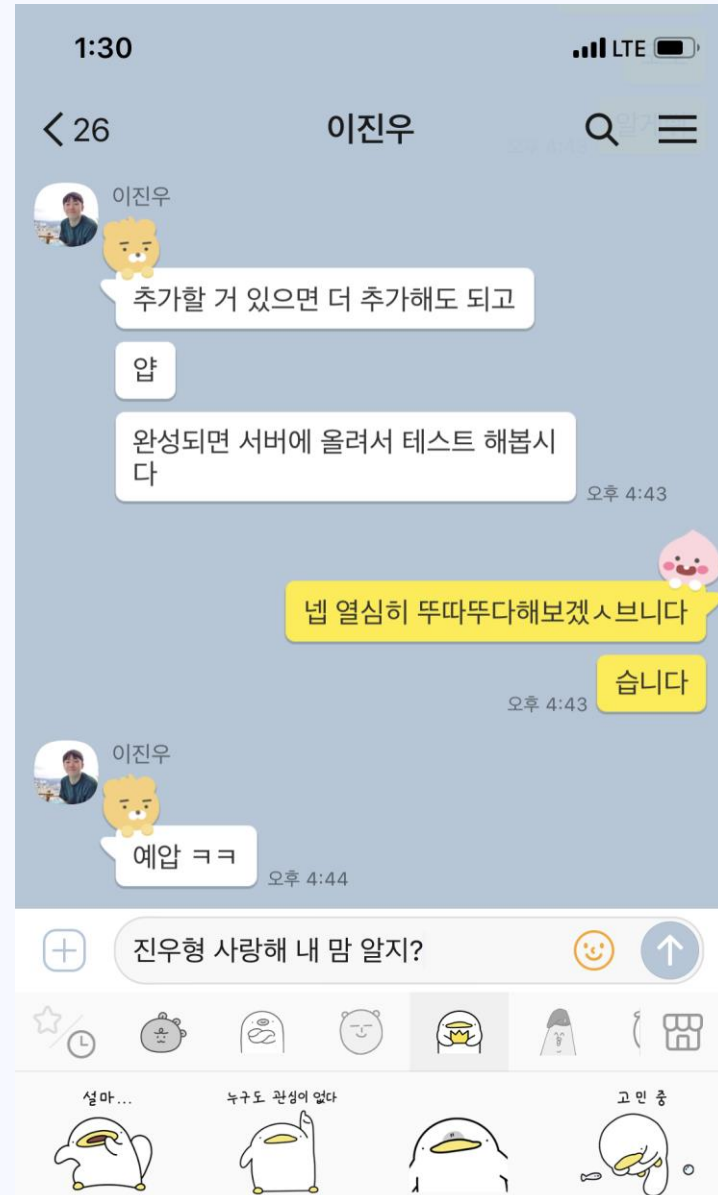


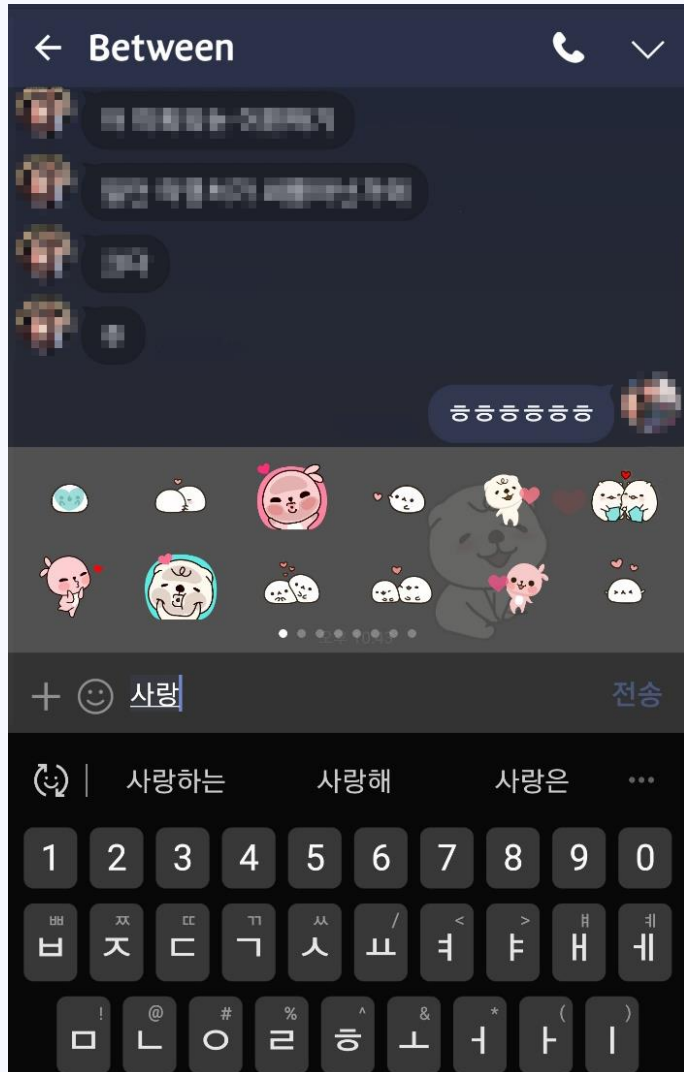




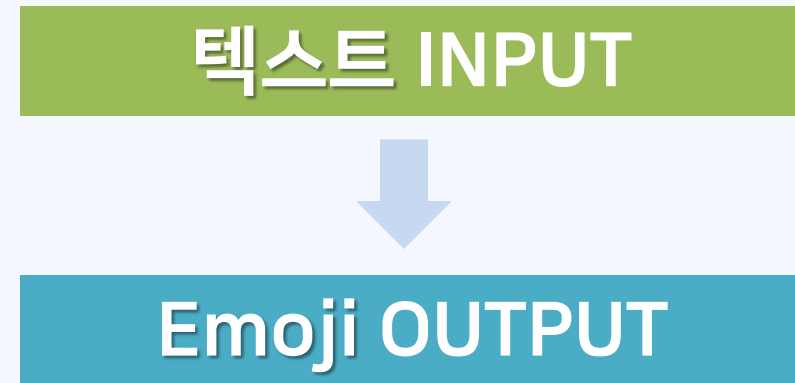


어디 있는거야..
하트하는 오구야..ㅠㅠ





* Between 메신저에 일부 구현된 기능 (키워드 단위)



키워드 뿐 아니라 문맥에 맞는 이모티콘을
출력하는 모델

인공지능



전문가 시스템

로봇 공학

HI

⋮

Machine Learning 과거의 데이터에 숨어있는 규칙, 패턴을 찾아 예측 / 분류 / 회귀
학습



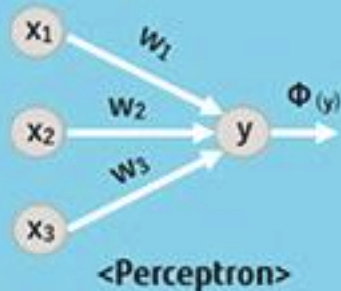
선형회귀

의사결정나무

SVM

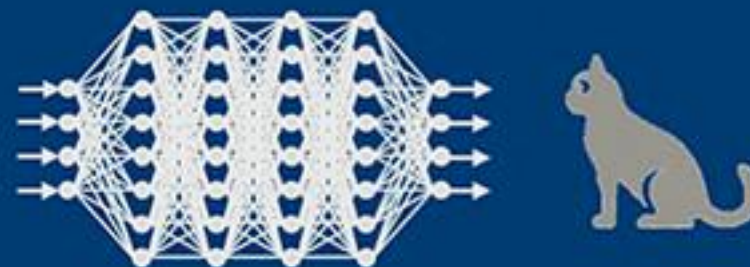
⋮

인공신경망 (Artificial Neural Network)



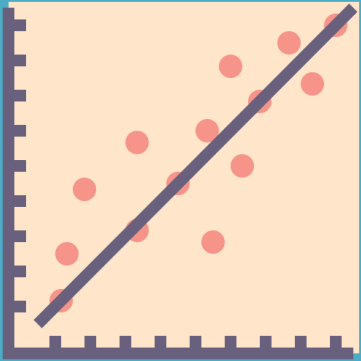
Deep Learning

여러 층의 인공 신경망을 이용한 학습



인간의 지적 활동을 기계를 통해 구현

Machine Learning

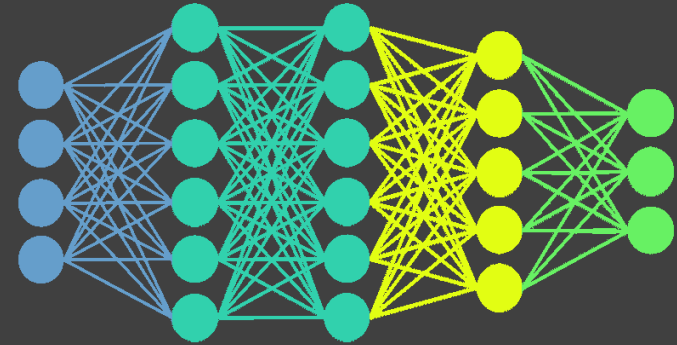


※ Machine Learning 특징

- 기본 모델의 형태가 비교적 직관적이고 해석이 용이함
- 인간이 데이터를 모델 형태에 맞게 가공해서 제공해야 함 (정규분포 등의 통계적 가정)

“알기 쉽게 정리해줄 테니
예쁘게 정리해서 가져와!”

Deep Learning



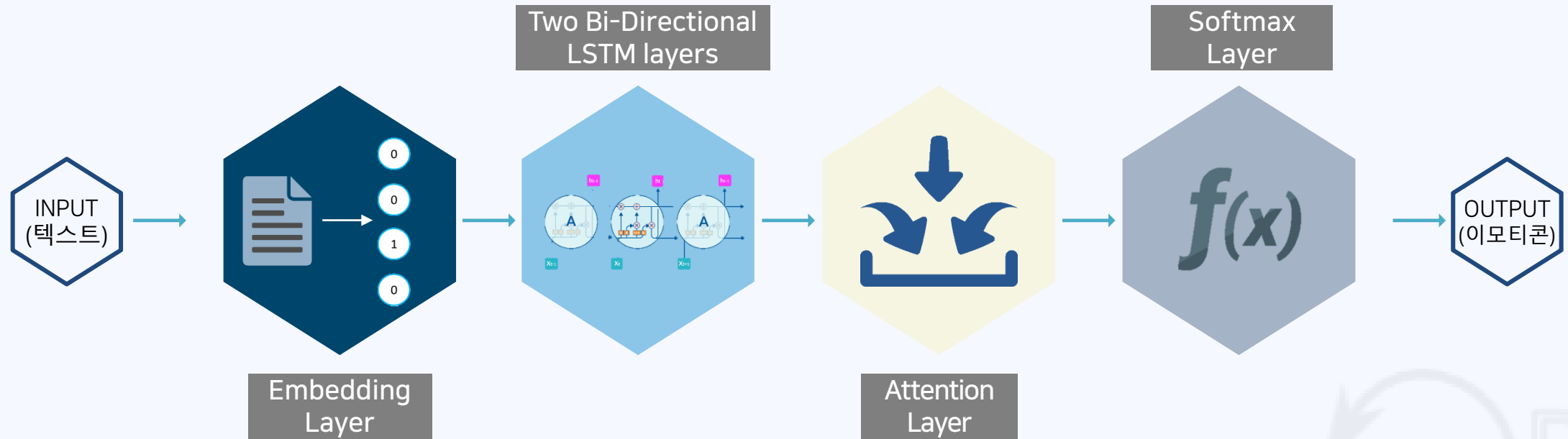
※ Deep Learning 특징

- 신경망이 복잡하게 얽혀 있어 해석이 어려움 (블랙박스)
- 데이터에 대한 통계적 가정이 크게 필요 없음
- 대신, 많은 데이터와 높은 연산량을 요구함
- 비정형 데이터(이미지, 텍스트 등)를 다루기에 용이함

“다 알아서 할 테니 너 결과만 봐!”

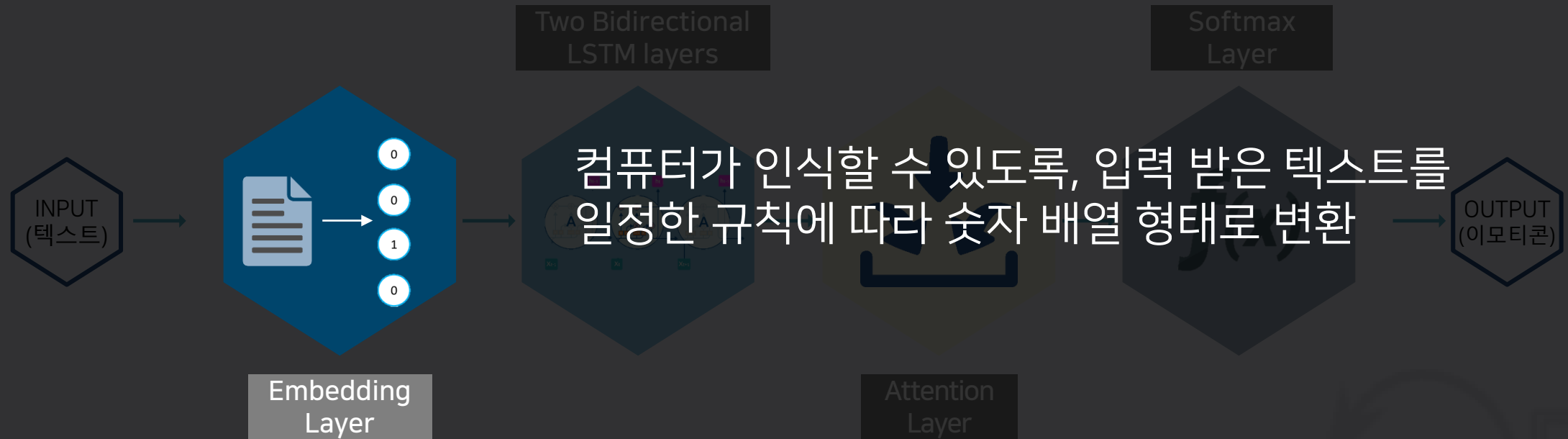
Bi-Directional LSTM Attention Model

텍스트 데이터를 다루기에 용이한 인공지능망 모델 RNN을 개선한 것으로, 총 4가지 층을 통해 데이터를 가공한다.



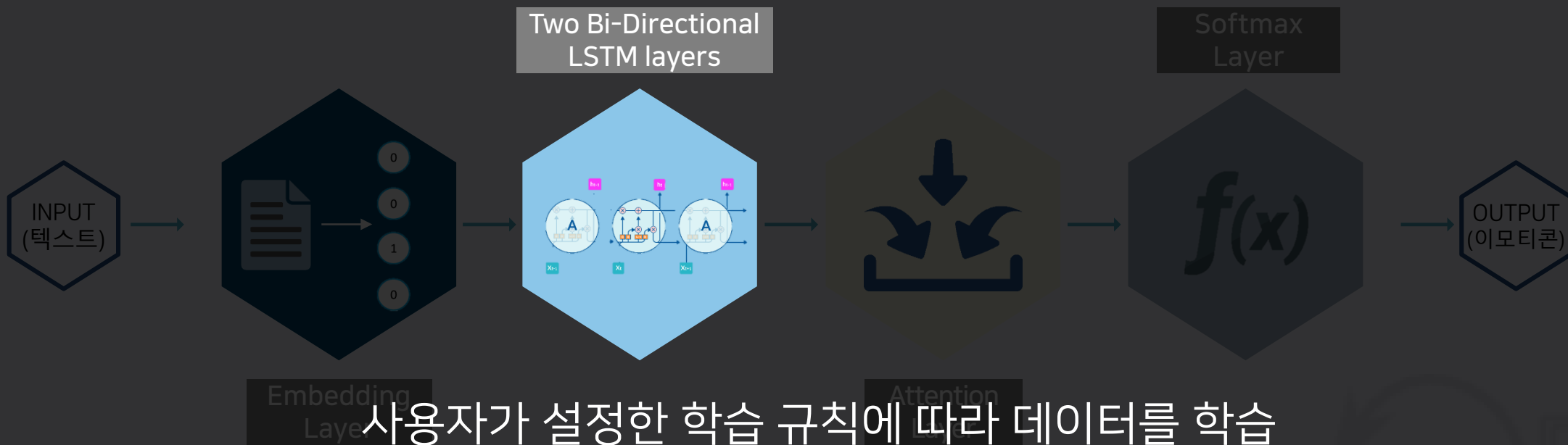
Bi-Directional LSTM Attention Model

텍스트 데이터를 다루기에 용이한 인공지능망 모델 RNN을 개선한 것으로, 총 4가지 층을 통해 데이터를 가공한다.



Bi-Directional LSTM Attention Model

텍스트 데이터를 다루기에 용이한 인공지능망 모델 RNN을 개선한 것으로, 총 4가지 층을 통해 데이터를 가공한다.



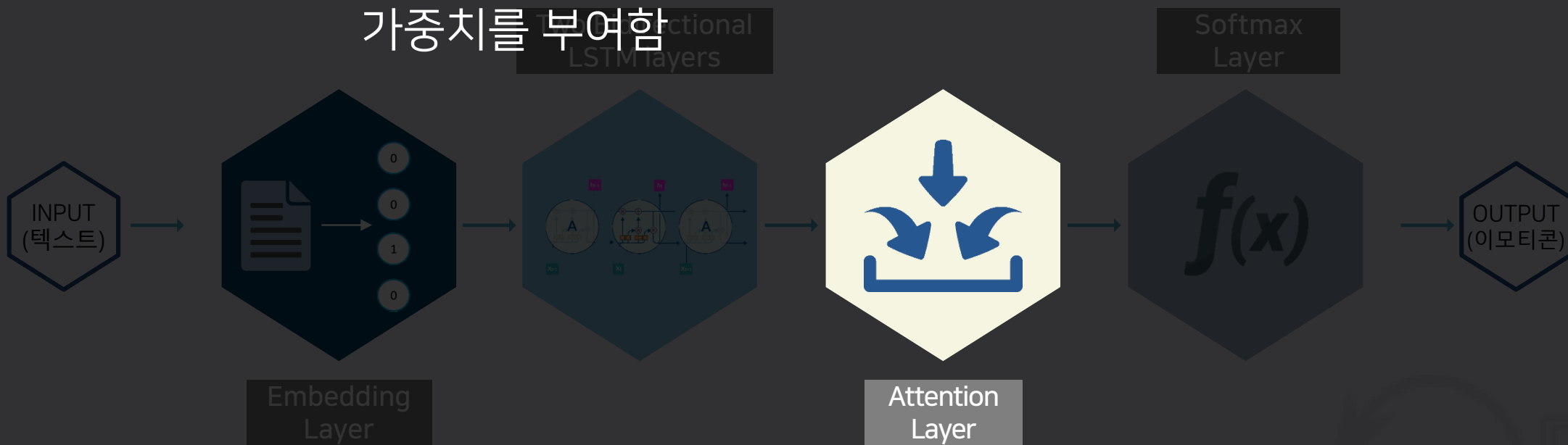
사용자가 설정한 학습 규칙에 따라 데이터를 학습

* Bi-Directional: 단방향이 아닌 양방향으로 학습을 한다는 의미로, 전후 문맥의 의미를 파악하는데 적합한 형태

Bi-Directional LSTM Attention Model

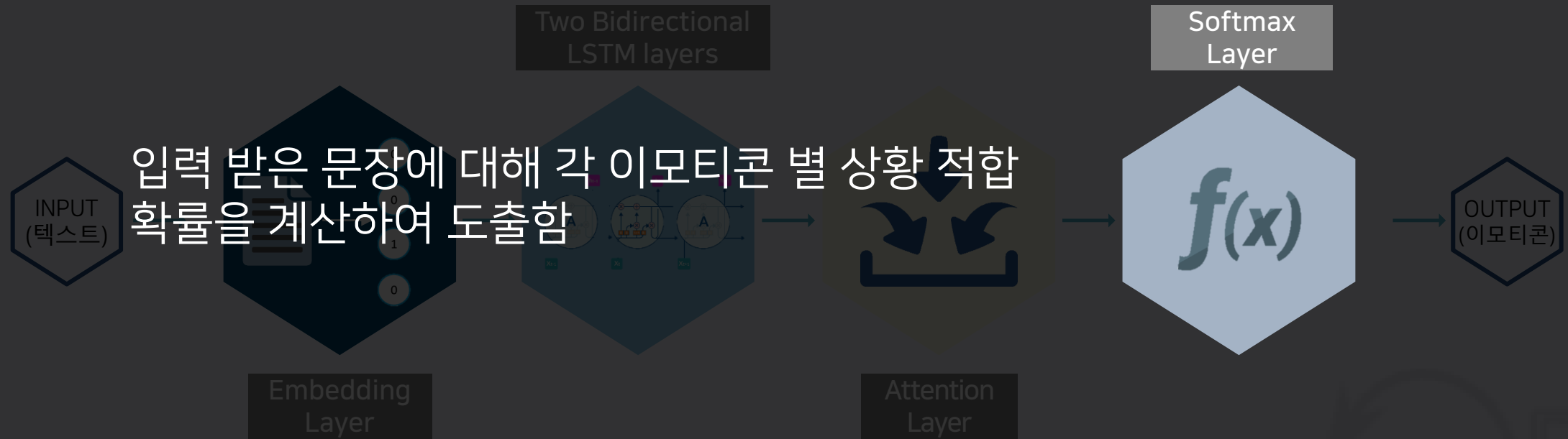
텍스트 데이터를 다루기에 용이한 인공지능망 모델 RNN을 개선한 것으로, 총 4가지 층을 통해 데이터를 가공한다.

문장 내에서 중요한 단어에 주목해서 해당 단어에
가중치를 부여함



Bi-Directional LSTM Attention Model

텍스트 데이터를 다루기에 용이한 인공지능망 모델 RNN을 개선한 것으로, 총 4가지 층을 통해 데이터를 가공한다.



1. 사용할 이모티콘 선정

2. Data Set 수집

3. Tokenizing

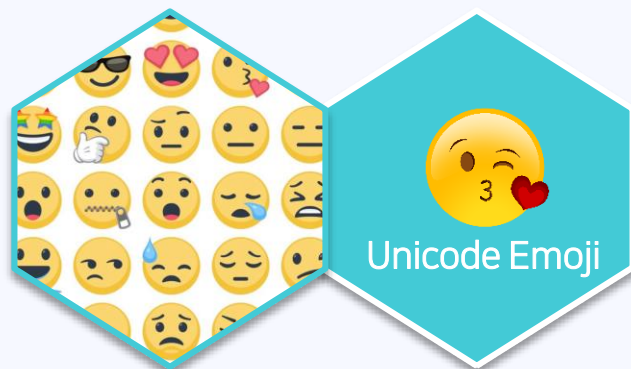
4. 모델 튜닝 및 학습

5. 웹 데모 구현

Back-End

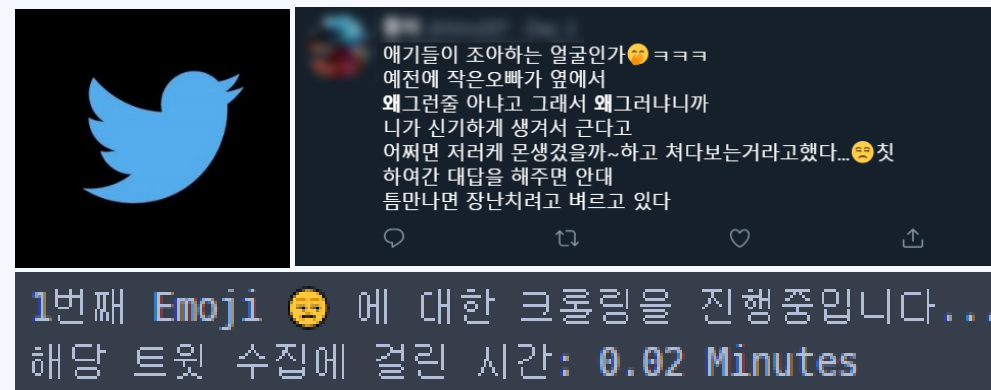
Front-End

1. 사용할 이모티콘 선정



범용성이 높고 어느 플랫폼에서도 검색 가능한
Unicode Emoji 64개

2. Data Set 수집



웹크롤링을 통해 해당되는 이모티콘이 담긴
트윗을 유니코드 별로 수집

총 Data Set 16,944,430개

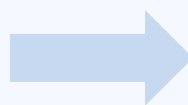
3. Tokenizing

text_stripped	token
비령님 어제 핸드폰 던져놓고 기절해서 이제봤구 쏘아침이구여 217년 첫멘션이네여 새...	[비령, 님, 어제, 핸드폰, 던지다, 기절, 하다, 이제, 보다, 쏘, 아침, 이...
그 너가 누구일까요	[그, 너, 가, 누구, 이다, 요]
이제 출근 다하섯으로냐 저도 모리님도 오늘 파이팅임미다	[이제, 출근, 다, 하, 섯, 으로, 나, 저, 도, 모리, 님, 도, 오늘, 파...
설록 본방사수하다가 잘 타이밍을 놓쳤어요	[설록, 본방, 사수, 하, 다가, 자다, 타이밍, 을, 놓치다]

변형 형태가 많은 한국어를 형태소 단위로 나누고,
원형으로 바꿔주는 Tokenizing을 진행

Konlpy 패키지의 Okt tag를 이용

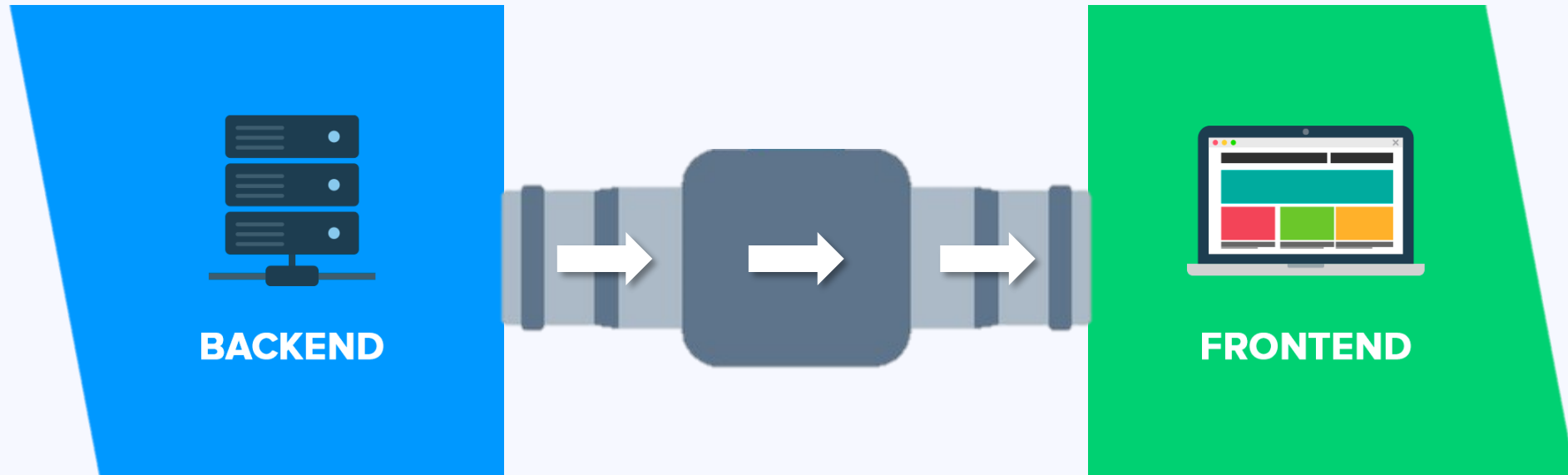
잘 타이밍을 놓쳤어요



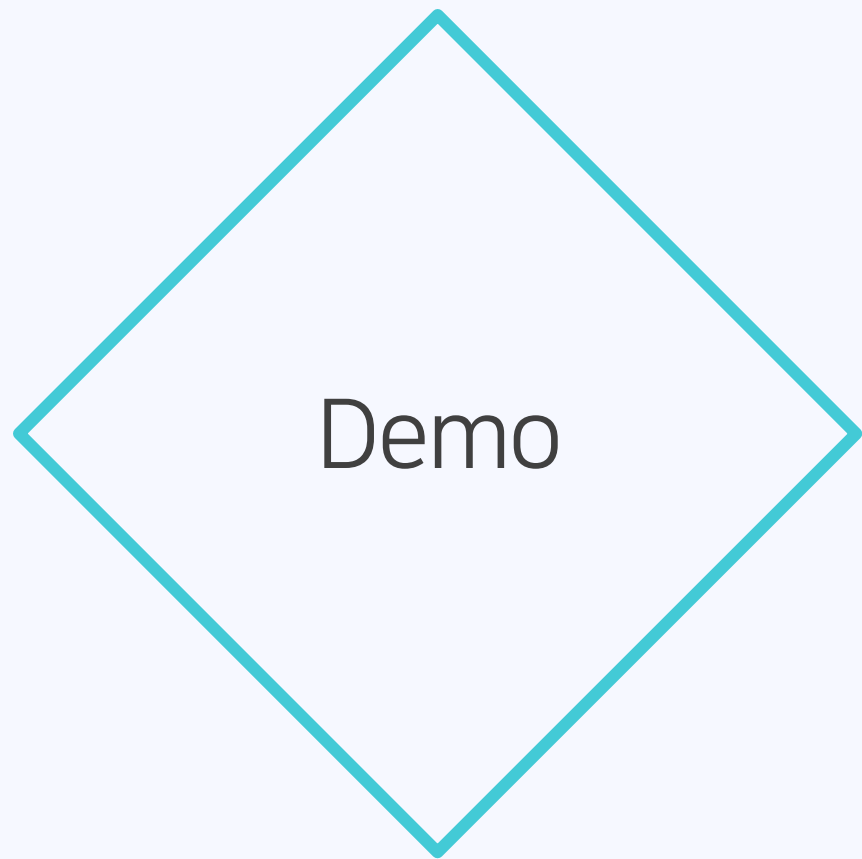
자다, 타이밍, 을, 놓치다

4. 모델 튜닝 및 학습

5. 웹 데모 구현



웹 서버 프레임워크 Flask를 이용해,
학습 완료한 모델을 활용하기 위한 웹 형태로 구현



Demo