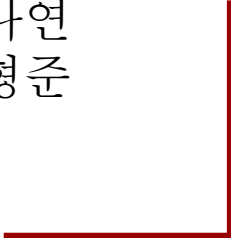




# Team NLP Project Plan

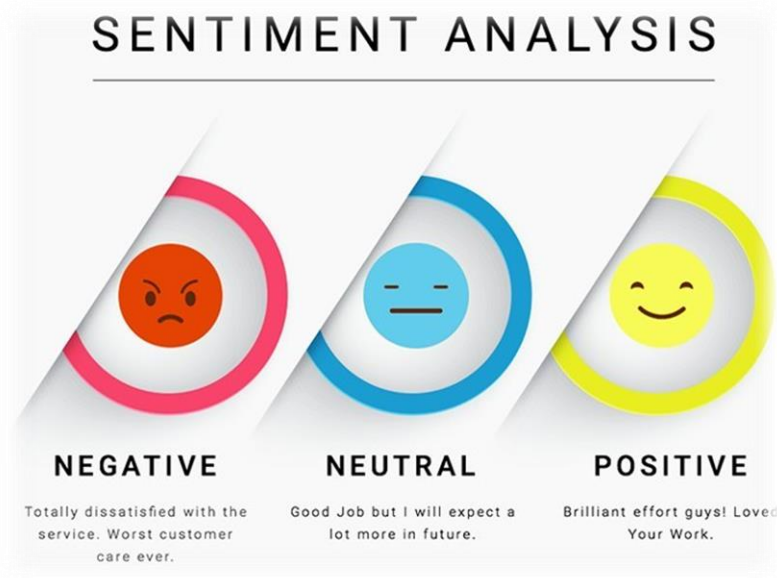
강유정 김근호 김나연  
김정현 박진우 임형준



# Index

1. Sentiment Analysis
2. Talking to The Little Prince
3. Study Plan

# 1. Sentiment Analysis



[ google image) Sentiment Analysis ]

## Product Review

- Movie review data analysis is most popular example
- 'Polarity' of the comment about the movie is estimated by the machine

테스트 정확도: 0.950

영문으로 리뷰를 작성하세요: i love this movie

예측: 긍정적 의견

확률: 92.753%

영문으로 리뷰를 작성하세요: hated movie

예측: 부정적 의견

확률: 99.738%

## 2. Talking to The Little Prince



[ google image) The Little Prince ]

### Seq2Seq Modeling

- Using 25 dialogs between the little prince and the fox from [The Little Prince]
- Chat data(data/chat.log) and Vocabulary data(data/chat.voc) have been used to train the model
- Very small sized data has been used > not precise but still make sense.

1 안녕.  
 2 안녕.  
 3 여기 있어. 사과나무 밑에...  
 4 넌 누구지? 참 예쁘구나...  
 5 난 여우야.  
 6 이리 와서 나하고 놀자.  
 7 난 너하고 놀 수 없어.  
 8 아! 미안해.  
 9 넌 무얼 찾고 있지?  
 10 난 사람들을 찾고 있어.  
 11 너는 닭을 찾고 있니?  
 12 아니. 난 친구들을 찾고 있어.  
 13 길들여진다는 게 무슨 뜻이야?  
 14 그건 관계가 생긴다는 뜻이야.  
 15 관계가 생긴다구?  
 16 그래.  
 17 꽃이 하나 있어... 그 꽃이 나를 길들인 것 같아...  
 18 그럴 수 있겠지. 지구 위엔 온갖 것이 모두 있으니까...  
 19 그럼 다른 별이란 말이야?  
 20 그래.  
 21 그 별에도 사냥꾼이 있어?  
 22 아니 없어.  
 23 그거 대단하군! 그럼 닭은?  
 24 없지.  
 25 그래서 난 좀 지겨워.  
 26 네가 날 길들이면 내 생활은 햇빛처럼 눈부시게 될 거야.  
 27 저기, 밑발이 보이지?  
 28 금빛 밑발을 보면, 네가 생각날 거야.  
 29 제발... 나를 길들여 줘!  
 30 그러구 싶은데... 난 시간이 없어.  
 31 그래서 사람들은 친구가 없어.

```

04 - ChatBot — python chat.py — python — Python chat.py...
[→ 04 - ChatBot git:(master) ✕ python chat.py
깨어나는 중 입니다 . 잠시만 기다려주세요 ...

> 안녕 .
안녕 .

> 난 여우야 .
이리 와서 나하고 놀자 .

> 어떻게 해야 하는데 ?
참을성이 있어야 해 .

> 난 너하고 놀 수 없어 .
아 ! 미안해 .

> 난 좀 지겨워 .
그래서 목요일은 무척 신나는 날이지 !

> 대단하군 !
그러구 . . . . 하나 시간이 . .

> 시간이 뭐 ?
넌 어떤 길들인 것에 대해 다르게 책임이 있어 .

> 알았어 ..
잘 대단하군 . 사과나무 밑에 . . .

> 밑에 뭐가 있어 ?
이리 와서 나하고 놀자 .

> 싫어 .
여기 하나 좀 있어 .

> 잘 있어 .
잘 가 .

> █
  
```

# 3. Study Plan

---

## Word Representation

Transforming the Natural Language into digits and vectors for further analysis with the machine

## Text Classification

Classifying the text data into pre-classified categories.  
Ex) Sentiment Analysis, Spam mail detector, News Report Classification

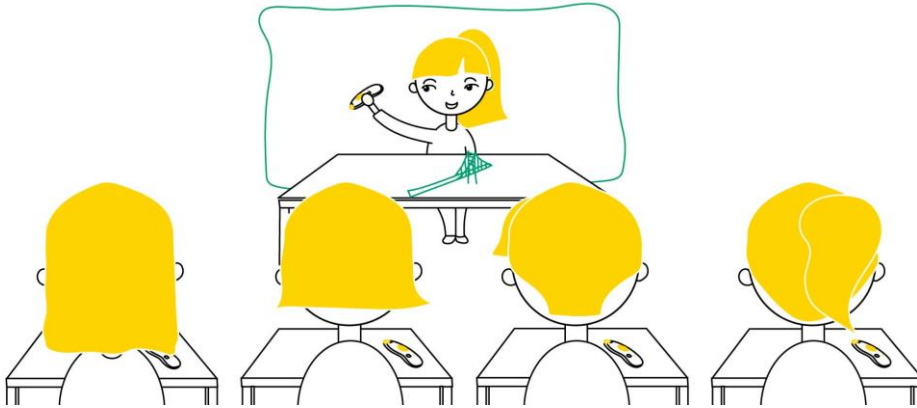
## Text Similarity

Measuring the similarity between the texts.  
cf) Jaccard, Euclidean, Manhattan, Cosine Similarities.

## Building Models

Realizing the actual communication.  
cf) MaLSTM model, Seq2Seq model, Transformer Network.

### 3. Mutual Peer Tutoring



- **Mutually sharing** the knowledge every week
- Everyone taking part in **teaching** process

- **Knowledge first** before diving into the coding process
- Hope to **make results on the way**

텐서플로와  
머신러닝으로 시작하는  
자연어처리

로지스틱  
회귀부터  
트랜스포머  
넷까지  
전달력, 이해도, 응용력 지향

DS 데이터 사이언스 시리즈 000

