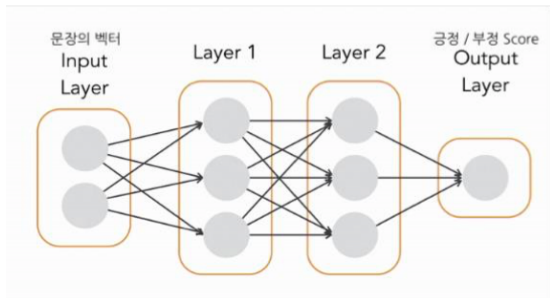


## 댓글 감성 분석 - 긍정/중립/부정 분류

### 사용 모델 : Keras의 MLP(다층 퍼셉트론 모델)

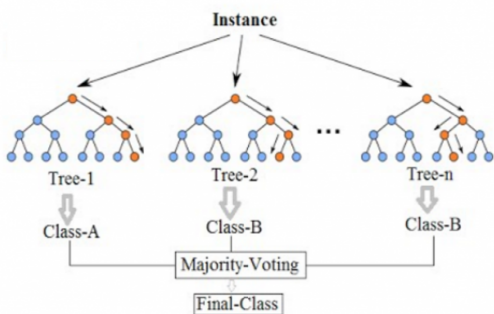
Keras는 파이썬 기반의 각 모듈들이 짧고 간결하며 최소한의 제약으로 서로 연결된다. Keras의 다층 퍼셉트론 모델에서는 Dense 레이어를 사용한다. Dense 레이어는 입력과 출력을 모두 연결해준다. 다른 딥러닝 모델보다 간결해 본 프로젝트 처리에 적합했고 LSTM 모델 보다 빠른 속도를 보여 댓글을 긍정/중립/부정으로 분류할 때 사용되었다.



위 사진과 같이, 입출력을 모두 연결해주는 64개의 유닛을 가진 Dense Layer 2개를 사용하고 활성화 함수로는 relu를 사용한다. 최종 output 을 내기 위해서는 이진 분류 문제의 출력층 활성화 함수인 Sigmoid 함수를 사용해 긍정의 리뷰일 확률을 출력한다. 손실함수로는 2진 분류에 적합한 binary\_crossentropy를 사용하고 RMSProp optimizer를 통해 경사하강법을 진행한다.

## 댓글 감성 분석 - 6가지 감정 분류

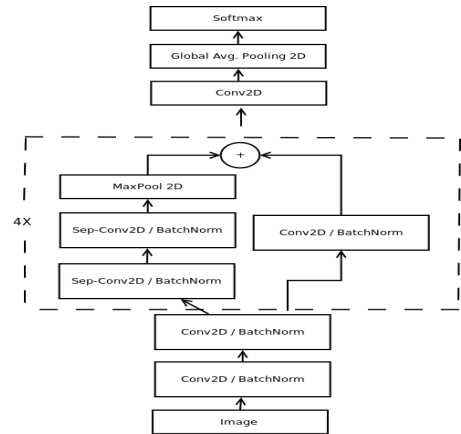
### 사용 모델 : Xgboost



Xgboost 알고리즘은 gradient boosting 알고리즘으로 다른 gradient boosting 알고리즘의 실행에 비해 빠르다. Xgboost는 Python 기반의 라이브러리로서 기본적으로 Boosting 기술을 사용한다. Boosting은 약한 분류기를 세트로 묶어서 정확도를 예측하는 방법이다. 이번 프로젝트에서 사용한 Xgboost 알고리즘의 주요 파라미터 중 booster는 gbtrees, 트리의 최대 깊이를 정의하는 파라미터인 max\_depth는 6, validation set에 적용되는 목적 함수인 objective는 다중 분류에 적합한 multi-class softmax, 학습률은 0.3으로 정해 6가지(중립, 행복, 슬픔, 놀람, 화남, 두려움) multi-class 분류를 사용해 학습시켰다.

## 영상 속 인물 표정 감성 분석

### 사용 모델 : Mini Xception model



GoogLeNet의 Inception module에 영감을 받아 컨볼루션 레이어 사이를 나눌 수 있을 만큼 전부 나눠서 사용하는 Xception architecture의 모습을 개량한 Mini Xception model은 사용자 얼굴에서 성별 혹은 표정의 감정을 분류하는데 효과적이다. 위 신경망의 가장 큰 특징은 매개변수의 수다. 비슷한 성능의 다른 CNN의 매개변수 수 5백만 개에 비해 이 모델의 매개변수는 58,432개로 100분의 1 정도로 수가 줄었다. 이런 특징은 학습뿐 아니라 전방 계산도 빠르기 때문에 웹상에서 빠르게 서비스를 제공해야 하는 현재 상황에 잘 맞다.

## 정확도 측정

댓글 분석은 네개의 유튜브 영상을 지정해 직접 라벨링을 해 정확도를 측정한다.

### 1) 댓글 - 긍정/중립/부정 분류의 정확도

영상 제목	맞은 댓글 수/총 댓글 수	확률(%)
나혼자산다 화사 트러플짜파게티 벤츠 먹방	397/539	73%
지오디 어머니께 유비 외국인반응	321/440	72%
ENG2박 3일 일본 vlog 돈키호테쇼핑	147/188	78%
BTS 신곡 Boy With Luv 외국인 리액션	41/47	87%
통계	906 /1214	74.6%

### 2) 댓글 - 6가지 감정 분류의 정확도

영상 제목	맞은 댓글 수/총 댓글 수	확률(%)
나혼자산다 화사 트러플짜파게티 벤츠 먹방	422/539	78%
지오디 어머니께 유비 외국인반응	357/440	81%
ENG2박 3일 일본 vlog 돈키호테쇼핑	151/188	80%
BTS 신곡 Boy With Luv 외국인 리액션	44/47	93%
통계	974 /1214	80.2%

### 3) 영상 - 7가지 감정 분류의 정확도

유튜브 영상 중 '나혼자산다 화사씨의 트러플 짜파게티 과연 맛은!?' 외 9 개의 영상에서 랜덤하게 사진 400개를 추출한 후 직접 라벨링을 해 .csv파일로 만든 후 정확도를 측정한다.

감정	정확도
Angry	66.66%
Disgust	18.18%
Scared	20.83%
Happy	61.98%
Sad	39.70%
Surprised	58.97%
Neutral	64.70%
통계	52.75%