

Youtube 댓글을 활용한 동영상 분석

11기 - 4팀
노희섭, AAA, BBB

목차

- 01. 프로젝트 개요
- 02. 프로젝트 팀 구성 및 역할
- 03. 프로젝트 수행 절차 및 방법
- 04. 프로젝트 수행 결과
- 05. 자체 평가 의견

1. 프로젝트 개요



▶ 프로젝트의 배경

- Youtube는 한국인이 카카오톡에 이어 **가장 많이 사용**하고, **사용시간이 가장 긴** 서비스
- Youtube 채널 중 광고를 게재하는 채널은 국가별로 보았을 때 **한국이** 약 10만 개로 **세계 7위**
- 피식대학, 숏박스 등 큰 인기를 끌고 있는 Youtube 채널은 **재생목록 별로** 서로 다른 세계관을 지닌 동영상이 제공

▶ 프로젝트의 목적

- **재생목록에 따른 시청자의 댓글** 반응을 분석
- 결과에 따른 의사결정으로 크리에이터가 자신의 **채널을 성장**

▶ 프로젝트의 구현 내용

- 웹 스크래핑을 통한 데이터 수집
- 데이터프레임 제작 및 전처리
- DL 모델 설계 및 학습
- 데이터 시각화 진행
- DB 구축

▶ 훈련 내용과의 관련성

- Section1 Pandas DataFrame, 데이터 전처리 등
- Section3 웹 스크래핑, API, DB 등
- Section4 자연어처리, 딥러닝 모델 설계 및 학습 등
- Section5 Python 프로그래밍 등

▶ 활용 장비 및 개발 환경

- Windows 10 64bit
- Mac OS
- Visual Studio Code
- Git Hub
- Anaconda
- Python == 3.8

▶ 프로젝트 구조

- Youtube 동영상 링크 입력
- 데이터 수집, 전처리, 감성 예측, 데이터 시각화, DB전송
- HTML 결과 페이지 출력

2. 프로젝트 팀 구성 및 역할



훈련생	역할	담당 업무
노희섭	팀장	<ul style="list-style-type: none">▶ 데이터 수집 및 전처리 진행▶ WordCloud 시각화
AAA	팀원	<ul style="list-style-type: none">▶ DL 학습용 데이터 수집▶ DL 모델 설계 및 훈련
BBB	팀원	<ul style="list-style-type: none">▶ 서버 Api 개발▶ DB 연동

3. 프로젝트 수행 절차 및 방법



구분	기간	활동	비고
사전 기획	▶ 6/24(금) ~ 6/27(월)	▶ 프로젝트 기획 및 주제 선정	▶ 아이디어 선정
역할 분담 및 기술 스택 보완	▶ 6/27(월) ~ 6/28(화)	▶ 역할 분담 ▶ 역할에 필요한 기술 스택 보완	▶ 역할 지원 및 보완
데이터 수집 및 전처리 코드 작성	▶ 6/28(화) ~ 7/2(토)	▶ 데이터 수집 코드 작성 ▶ DL모델 예측에 맞는 전처리 코드 작성	▶ Git Hub를 통한 공유
DL 모델링	▶ 6/30(목) ~ 7/5(화)	▶ DL모델 구현	▶ Git Hub를 통한 공유
서비스 구축	▶ 6/30(목) ~ 7/5(화)	▶ Web API 설계 ▶ MongoDB 연동 ▶ jinja 를 활용한 결과 페이지 구현	▶ 최적화, 오류 수정
총 개발기간	▶ 6/24(월) ~ 7/6(수) (약 2주)		

4. 프로젝트 수행 결과

결과 제시 ① DataFrame 설명



▶ DataFrame

1 test_data
✓ 0.2s Python

	list_name	name	views	youtube_id	comment_ori	like_num	comment_del	tokenized	comment_cloud
0	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	thanks for coming.	□허혈,, 무한상사에 진심인 나,,재즈 대신 무한상사 틀어놓고 자는데,,	1500	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데	[허, 혈, 무한, 상사, 진심, 나, 재즈, 대신, 무한, 상사, 틀, 어, 놓,,...	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데
1	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	쏘토닉	와....예능으로 이렇게 엄청난 구성이 가능하구나. 진짜 대단하다 존경을 넘어 경외...	155	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들	[와예, 능, 으로, 이렇게, 엄청난, 구성, 가능, 구나, 진짜, 대단, 존경, ...	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들
2	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	이진현	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음..그...	41	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...	[진짜, 에피소드, 실패, 든, 성공, 든, 이렇게, 여러, 개, 도전, 해, 보,,...	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...
3	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	kim david	김태호는 유재석 데리고 하고 싶은 거 다하는 구만 ㅋㅋㅋ	1700	김태호는 유재석 데리고 하고 싶은 거 다하는 구만	[김태호, 유재석, 데리, 싶, 거, 구만]	김태호는 유재석 데리고 하고 싶은 거 다하는 구만
4	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	신영웅	진짜 김태호PD의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 ㅎㅎ ...	446	진짜 김태호의 제작력이란 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도많이...	[진짜, 김태호, 제작, 력, 이란, 참, 스펙트럼, 넓, 정말, 포인트, 잘, 잡...	진짜 김태호의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에...

- list_name
 - 재생목록 이름

4. 프로젝트 수행 결과

결과 제시 ① DataFrame 설명



▶ DataFrame

1 test_data
✓ 0.2s Python

	list_name	name	views	youtube_id	comment_ori	like_num	comment_del	tokenized	comment_cloud
0	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	thanks for coming.	□허혈,, 무한상사에 진심인 나,,재즈 대신 무한상사 틀어놓고 자는데,,	1500	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데	[허, 혈, 무한, 상사, 진심, 나, 재즈, 대신, 무한, 상사, 틀, 어, 놓,...	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데
1	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	쏘토닉	와....예능으로 이렇게 엄청난 구성이 가능하구나. 진짜 대단하다 존경을 넘어 경외...	155	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들	[와예, 능, 으로, 이렇게, 엄청난, 구성, 가능, 구나, 진짜, 대단, 존경, ...	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들
2	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	이진현	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그...	41	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...	[진짜, 에피소드, 실패, 든, 성공, 든, 이렇게, 여러, 개, 도전, 해, 보,...	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...
3	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	kim david	김태호는 유재석 데리고 하고 싶은 거 다하는 구만 ㅋㅋㅋ	1700	김태호는 유재석 데리고 하고 싶은 거 다하는 구만	[김태호, 유재석, 데리, 싶, 거, 구만]	김태호는 유재석 데리고 하고 싶은 거 다하는 구만
4	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	신영웅	진짜 김태호PD의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 ㅎㅎ	446	진짜 김태호의 제작력이란 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도많이...	[진짜, 김태호, 제작, 력, 이란, 참, 스펙트럼, 넓, 정말, 포인트, 잘, 잡...	진짜 김태호의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에...

- name
 - 개별 동영상 이름

4. 프로젝트 수행 결과

결과 제시 ① DataFrame 설명



▶ DataFrame

1 test_data
✓ 0.2s Python

	list_name	name	views	youtube_id	comment_ori	like_num	comment_del	tokenized	comment_cloud
0	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	thanks for coming.	ㅎ허혈,, 무한상사에 진심인 나,,재즈 대신 무한상사 틀어놓고 자는데,,	1500	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데	[허, 혈, 무한, 상사, 진심, 나, 재즈, 대신, 무한, 상사, 틀, 어, 놓,,...	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데
1	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	쏘토닉	와....예능으로 이렇게 엄청난 구성이 가능하구나. 진짜 대단하다 존경을 넘어 경외...	155	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들	[와예, 능, 으로, 이렇게, 엄청난, 구성, 가능, 구나, 진짜, 대단, 존경, ...	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들
2	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	이진현	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그...	41	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...	[진짜, 에피소드, 실패, 든, 성공, 든, 이렇게, 여러, 개, 도전, 해, 보,,...	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...
3	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	kim david	김태호는 유재석 데리고 하고 싶은 거 다하는 구만 ㅋㅋㅋ	1700	김태호는 유재석 데리고 하고 싶은 거 다하는 구만	[김태호, 유재석, 데리, 싶, 거, 구만]	김태호는 유재석 데리고 하고 싶은 거 다하는 구만
4	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	신영웅	진짜 김태호PD의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 ㅎㅎ	446	진짜 김태호의 제작력이란 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도많이...	[진짜, 김태호, 제작, 력, 이란, 참, 스펙트럼, 넓, 정말, 포인트, 잘, 잡,,...	진짜 김태호의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에...

- views
 - 개별 동영상의 조회수

4. 프로젝트 수행 결과

결과 제시 ① DataFrame 설명



▶ DataFrame

1 test_data
✓ 0.2s Python

	list_name	name	views	youtube_id	comment_ori	like_num	comment_del	tokenized	comment_cloud
0	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	thanks for coming.	ㅎ허혈,, 무한상사에 진심인 나,,재즈 대신 무한상사 틀어놓고 자는데,,	1500	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데	[허, 혈, 무한, 상사, 진심, 나, 재즈, 대신, 무한, 상사, 틀, 어, 놓,,...	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데
1	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	쏘토닉	와...예능으로 이렇게 엄청난 구성이 가능하구나. 진짜 대단하다 존경을 넘어 경외...	155	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들	[와예, 능, 으로, 이렇게, 엄청난, 구성, 가능, 구나, 진짜, 대단, 존경, ...	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들
2	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	이진현	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그...	41	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...	[진짜, 에피소드, 실패, 든, 성공, 든, 이렇게, 여러, 개, 도전, 해, 보,,...	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...
3	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	kim david	김태호는 유재석 데리고 하고 싶은 거 다하는 구만 ㅋㅋㅋ	1700	김태호는 유재석 데리고 하고 싶은 거 다하는 구만	[김태호, 유재석, 데리, 싶, 거, 구만]	김태호는 유재석 데리고 하고 싶은 거 다하는 구만
4	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	신영웅	진짜 김태호PD의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 ㅎㅎ ...	446	진짜 김태호의 제작력이란 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도많이...	[진짜, 김태호, 제작, 력, 이란, 참, 스펙트럼, 넓, 정말, 포인트, 잘, 잡,,...	진짜 김태호의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에...

- youtube_id
 - 동영상에 댓글을 남긴 유튜브 아이디

4. 프로젝트 수행 결과

결과 제시 ① DataFrame 설명



▶ DataFrame

1 test_data
✓ 0.2s Python

	list_name	name	views	youtube_id	comment_ori	like_num	comment_del	tokenized	comment_cloud
0	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	thanks for coming.	□허혈,, 무한상사에 진심인 나,,재즈 대신 무한상사 틀어놓고 자는데,,	1500	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데	[허, 혈, 무한, 상사, 진심, 나, 재즈, 대신, 무한, 상사, 틀, 어, 놓,,...	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데
1	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	쏘토닉	와...예능으로 이렇게 엄청난 구성이 가능하구나. 진짜 대단하다 존경을 넘어 경외...	155	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들	[와예, 능, 으로, 이렇게, 엄청난, 구성, 가능, 구나, 진짜, 대단, 존경, ...	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들
2	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	이진현	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음..그...	41	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...	[진짜, 에피소드, 실패, 든, 성공, 든, 이렇게, 여러, 개, 도전, 해, 보,,...	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...
3	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	kim david	김태호는 유재석 데리고 하고 싶은 거 다하는 구만 ㅋㅋㅋ	1700	김태호는 유재석 데리고 하고 싶은 거 다하는 구만	[김태호, 유재석, 데리, 싶, 거, 구만]	김태호는 유재석 데리고 하고 싶은 거 다하는 구만
4	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	신영웅	진짜 김태호PD의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 ㅎㅎ	446	진짜 김태호의 제작력이란 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도많이...	[진짜, 김태호, 제작, 력, 이란, 참, 스펙트럼, 넓, 정말, 포인트, 잘, 잡,,...	진짜 김태호의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에...

- comment_ori
 - 동영상에 남겨진 댓글의 원본

4. 프로젝트 수행 결과

결과 제시 ① DataFrame 설명



▶ DataFrame

1 test_data
✓ 0.2s Python

	list_name	name	views	youtube_id	comment_ori	like_num	comment_del	tokenized	comment_cloud
0	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	thanks for coming.	ㅎ혈,, 무한상사에 진심인 나,, 재즈 대신 무한상사 틀어놓고 자는데,,	1500	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데	[허, 혈, 무한, 상사, 진심, 나, 재즈, 대신, 무한, 상사, 틀, 어, 놓,,...	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데
1	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	쏘토닉	와...예능으로 이렇게 엄청난 구성이 가능하구나. 진짜 대단하다 존경을 넘어 경외...	155	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들	[와예, 능, 으로, 이렇게, 엄청난, 구성, 가능, 구나, 진짜, 대단, 존경, ...	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들
2	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	이진현	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음..그...	41	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...	[진짜, 에피소드, 실패, 든, 성공, 든, 이렇게, 여러, 개, 도전, 해, 보,,...	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...
3	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	kim david	김태호는 유재석 데리고 하고 싶은 거 다하는 구만 ㅋㅋㅋ	1700	김태호는 유재석 데리고 하고 싶은 거 다하는 구만	[김태호, 유재석, 데리, 싶, 거, 구만]	김태호는 유재석 데리고 하고 싶은 거 다하는 구만
4	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	신영웅	진짜 김태호PD의 제작력이란... 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 ㅎㅎ	446	진짜 김태호의 제작력이란 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도많이...	[진짜, 김태호, 제작, 력, 이란, 참, 스펙트럼, 넓, 정말, 포인트, 잘, 잡,,...	진짜 김태호의 제작력이란... 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도...

- like_num
 - 댓글을 누른 좋아요의 숫자

4. 프로젝트 수행 결과

결과 제시 ① DataFrame 설명



▶ DataFrame

1 test_data
✓ 0.2s Python

	list_name	name	views	youtube_id	comment_ori	like_num	comment_del	tokenized	comment_cloud
0	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	thanks for coming.	ㅎ허혈,, 무한상사에 진심인 나,,재즈 대신 무한상사 틀어놓고 자는데,,	1500	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데	[허, 혈, 무한, 상사, 진심, 나, 재즈, 대신, 무한, 상사, 틀, 어, 놓,...	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데
1	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	쏘토닉	와...예능으로 이렇게 엄청난 구성이 가능하구나. 진짜 대단하다 존경을 넘어 경외...	155	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들	[와예, 능, 으로, 이렇게, 엄청난, 구성, 가능, 구나, 진짜, 대단, 존경, ...	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들
2	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	이진현	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그...	41	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...	[진짜, 에피소드, 실패, 든, 성공, 든, 이렇게, 여러, 개, 도전, 해, 보,...	진짜 에피소드가 실패하든 성공하든 이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...
3	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	kim david	김태호는 유재석 데리고 하고 싶은 거 다하는 구만 ㅋㅋㅋ	1700	김태호는 유재석 데리고 하고 싶은 거 다하는 구만	[김태호, 유재석, 데리, 싶, 거, 구만]	김태호는 유재석 데리고 하고 싶은 거 다하는 구만
4	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	신영웅	진짜 김태호PD의 제작력이란... 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 ㅎㅎ	446	진짜 김태호의 제작력이란 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도많이...	[진짜, 김태호, 제작, 력, 이란, 참, 스펙트럼, 넓, 정말, 포인트, 잘, 잡...	진짜 김태호의 제작력이란... 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에...

- comment_del
 - 이모티콘, 문장부호, 특수문자 등을 제거한 댓글

4. 프로젝트 수행 결과

결과 제시 ① DataFrame 설명



▶ DataFrame

1 test_data
✓ 0.2s Python

	list_name	name	views	youtube_id	comment_ori	like_num	comment_del	tokenized	comment_cloud
0	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	thanks for coming.	□허혈,, 무한상사에 진심인 나,,재즈 대신 무한상사 틀어놓고 자는데,,	1500	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데	[허, 혈, 무한, 상사, 진심, 나, 재즈, 대신, 무한, 상사, 틀, 어, 놓,...]	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데
1	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	쏘토닉	와...예능으로 이렇게 엄청난 구성이 가능하구나. 진짜 대단하다 존경을 넘어 경외...	155	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들	[와예, 능, 으로, 이렇게, 엄청난, 구성, 가능, 구나, 진짜, 대단, 존경, ...]	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들
2	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	이진현	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그...	41	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...	[진짜, 에피소드, 실패, 든, 성공, 든, 이렇게, 여러, 개, 도전, 해, 보,...]	진짜 에피소드가 실패하든 성공하든 이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...
3	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	kim david	김태호는 유재석 데리고 하고 싶은 거 다하는 구만 ㅋㅋㅋ	1700	김태호는 유재석 데리고 하고 싶은 거 다하는 구만	[김태호, 유재석, 데리, 싶, 거, 구만]	김태호는 유재석 데리고 하고 싶은 거 다하는 구만
4	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	신영웅	진짜 김태호PD의 제작력이란... 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 ㅎㅎ	446	진짜 김태호의 제작력이란 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도많이...	[진짜, 김태호, 제작, 력, 이란, 참, 스펙트럼, 넓, 정말, 포인트, 잘, 잡,...]	진짜 김태호의 제작력이란... 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도...

- tokenized
 - 형태소 분석기(Mecab)를 활용한 토큰화 → DL 모델로 감성 예측할 때 사용

4. 프로젝트 수행 결과

결과 제시 ① DataFrame 설명



▶ DataFrame

1 test_data
✓ 0.2s

	list_name	name	views	youtube_id	comment_ori	like_num	comment_del	tokenized	comment_cloud
0	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	thanks for coming.	□허혈,, 무한상사에 진심인 나,,재즈 대신 무한상사 틀어놓고 자는데,,	1500	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데	[허, 혈, 무한, 상사, 진심, 나, 재즈, 대신, 무한, 상사, 틀, 어, 놓,...	허혈 무한상사에 진심인 나재즈 대신 무한상사 틀어놓고 자는데
1	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	쏘토닉	와...예능으로 이렇게 엄청난 구성이 가능하구나. 진짜 대단하다 존경을 넘어 경외...	155	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들	[와예, 능, 으로, 이렇게, 엄청난, 구성, 가능, 구나, 진짜, 대단, 존경, ...	와예능으로 이렇게 엄청난 구성이 가능하구나 진짜 대단하다 존경을 넘어 경외감이 들
2	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	이진현	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그...	41	진짜 에피소드가 실패하든 성공하든이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...	[진짜, 에피소드, 실패, 든, 성공, 든, 이렇게, 여러, 개, 도전, 해, 보,...	진짜 에피소드가 실패하든 성공하든 이렇게 여러개 도전해보는 예능은놀워뿐인것 같음그래서...
3	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	kim david	김태호는 유재석 데리고 하고 싶은 거 다하는 구만 ㅋㅋㅋ	1700	김태호는 유재석 데리고 하고 싶은 거 다하는 구만	[김태호, 유재석, 데리, 싶, 거, 구만]	김태호는 유재석 데리고 하고 싶은 거 다하는 구만
4	유본부장	[놀면뭐하니? 예고] 바쁜 도심 속.. 유부장의 새로운 시작 ✨ MBC210529방송	600312	신영웅	진짜 김태호PD의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 ㅎㅎ	446	진짜 김태호의 제작력이란 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도많이...	[진짜, 김태호, 제작, 력, 이란, 참, 스펙트럼, 넓, 정말, 포인트, 잘, 잡...	진짜 김태호의 제작력이란,,, 참 스펙트럼도 넓고 정말 포인트를 잘 잡는듯 이번에도...

- comment_cloud
 - WordCloud를 위해 전처리 한 댓글

4. 프로젝트 수행 결과

결과 제시 ② DL 모델 학습

▶ 학습 데이터 소개

- AIHub: 한국어_단발성_대화 데이터 셋
 - Context: 38,594개
 - Label: 7개(행복, 놀람, 분노, 공포, 혐오, 슬픔, 중립)

		Sentence	Emotion
0		언니 동생으로 부르는게 맞는 일인가요	0
1		그냥 내 느낌일뿐겠지	0
2		아직너무초기라서 그런거죠	0
3		유치원버스 사고 났다던데	0
4		근데 원래이런거맞나요	0
...	
38589		솔직히 예보 제대로 못하는 데 세금이라도 아끼게 그냥 폐지해라	6
38590		재미가 없으니 망하지	6

행복	0.156423
놀람	0.152822
분노	0.146784
공포	0.141680
혐오	0.140670
슬픔	0.136472
중립	0.125149

4. 프로젝트 수행 결과

결과 제시 ② DL 모델 학습

▶ 학습 데이터 소개

- Tokenizer: Mecab(형태소 분류), 특수문자, 의성어 등 제거
- Word Embedding : Fast Text를 활용, 각 단어를 character 단위 n-gram으로 쪼갬다
 - Word2Vec 한계 보완

공포리뷰 평균 길이 7.7873079736649595

놀람리뷰 평균 길이 6.770261105459478

분노리뷰 평균 길이 8.286142983230361

슬픔리뷰 평균 길이 7.827226124928802

중립리뷰 평균 길이 7.8648033126294

행복리뷰 평균 길이 7.430511843630942

혐오리뷰 평균 길이 9.114754098360656

```
TOP_10 [('안', 490), ('겠', 392), ('네요', 347), ('같', 326), ('없', 294), ('보', 287), ('해야', 284), ('아', 260), ('어', 255), ('어떨  
게', 251)]
```

```
TOP_10 [('보', 312), ('진짜', 301), ('안', 297), ('아니', 271), ('냐', 269), ('어', 269), ('에서', 249), ('왜', 242), ('했', 240),  
('있', 239)]
```

```
TOP_10 [('냐', 492), ('아', 431), ('안', 373), ('라', 339), ('없', 320), ('으로', 296), ('나라', 288), ('로', 288), ('어', 279),  
('말', 270)]
```

```
TOP_10 [('네요', 568), ('안', 517), ('너무', 448), ('없', 403), ('겠', 351), ('보', 334), ('아', 323), ('어', 299), ('같', 297), ('어  
요', 278)]
```

```
TOP_10 [('안', 321), ('겠', 281), ('없', 277), ('보', 240), ('으로', 238), ('로', 226), ('에서', 214), ('것', 210), ('같', 209),  
('할', 188)]
```

```
TOP_10 [('좋', 557), ('감사', 520), ('합니다', 507), ('잘', 431), ('보', 427), ('주', 402), ('님', 391), ('습니다', 379), ('해', 363),  
('네요', 358)]
```

```
TOP_10 [('안', 460), ('없', 397), ('아', 391), ('냐', 378), ('어', 322), ('겠', 304), ('말', 298), ('니', 289), ('으로', 271), ('로',  
271)]
```

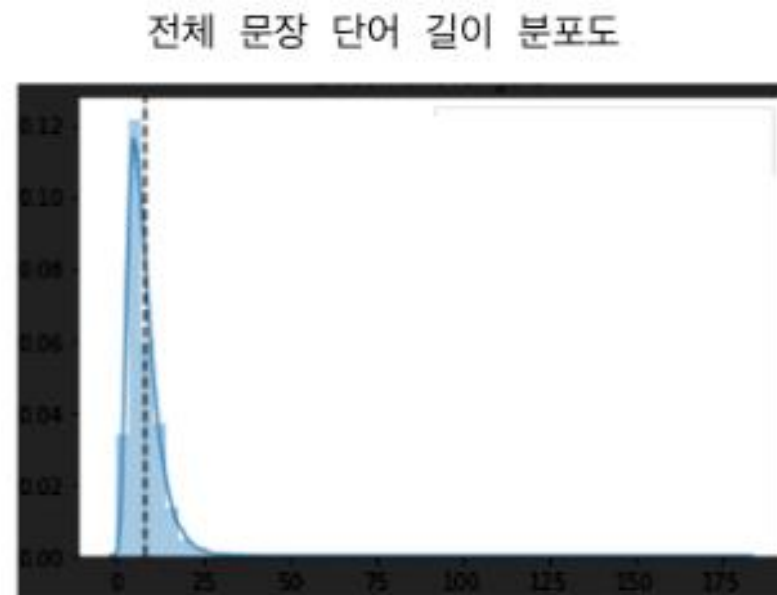
4. 프로젝트 수행 결과

결과 제시 ② DL 모델 학습

▶ 학습 데이터 소개

- Tokenizer: Mecab(형태소 분류), 특수문자, 의성어 등 제거
- Word Embedding : Fast Text를 활용, 각 단어를 character 단위 n-gram으로 쪼갬다
 - Word2Vec 한계 보완

공포리뷰 평균 길이 7.7873079736649595
 놀람리뷰 평균 길이 6.770261105459478
 분노리뷰 평균 길이 8.286142983230361
 슬픔리뷰 평균 길이 7.827226124928802
 중립리뷰 평균 길이 7.8648033126294
 행복리뷰 평균 길이 7.430511843630942
 혐오리뷰 평균 길이 9.114754098360656



패딩 처리시 위 시각화를 참조하였다

없', 294), ('보', 287), ('해야', 284), ('아', 260), ('어', 255), ('어떨
 ('냐', 269), ('어', 269), ('에서', 249), ('왜', 242), ('했', 240),
 ('', 320), ('으로', 296), ('나라', 288), ('로', 288), ('어', 279),
 ('겠', 351), ('보', 334), ('아', 323), ('어', 299), ('같', 297), ('어
 '로', 238), ('로', 226), ('에서', 214), ('것', 210), ('같', 209),
 ('보', 427), ('주', 402), ('님', 391), ('습니다', 379), ('해', 363),
 ' ', 322), ('겠', 304), ('말', 298), ('니', 289), ('으로', 271), ('로',

4. 프로젝트 수행 결과

결과 제시 ③ 모델 개요

▶ 사용한 모델 (RNN, LSTM, 다층Bi-LSTM)

- RNN
 - 연속형 데이터를 처리하기 위해 고안된 신경망.
 - 어떤 길이의 Sequential 데이터라도 처리 가능.
 - 하지만 현재 노드 위치와 먼 상태를 사용한 문맥 처리가 어렵다
 - (장기 의존성 문제 - Long-Term Dependency)
- LSTM
 - RNN은 거슬러 올라가는 단계수가 많아질수록 기울기 소실 문제가 발생한다.
 - 기울기 소실 문제를 해결한 신경망
- Bi-LSTM
 - RNN이나 LSTM은 입력 순서를 시간 순대로 입력하기 때문에 결과물이 직전 패턴을 기반으로 수렴하는 경향을 보인다는 한계가 있다.
 - 그러나 데이터 길이가 길고 층이 깊으면, 과거의 정보가 손실되는 단점이 있다. 이를 극복하기 위해 제안된 알고리즘이 양방향 LSTM이다. (Ko et al., 2018)

4. 프로젝트 수행 결과

결과 제시 ③ 모델 개요

▶ 공통적으로 적용된 초기 학습 방법

- Embedding 정수 인덱싱 한 후 padding처리한 x_train을 밀집 벡터로 mapping
 - 0으로 padding처리 된 인덱스는 의미 없음
- 각 신경망 구성
- 각 신경망 계층에서 나오는 결과를 softmax 적용, 이를 통해 각 결과 값이 7개의 감정에 할당될 확률

```
model_1 = Sequential()  
model_1.add(Embedding(vocab_size, dim, mask_zero=True, input_length=max_length))  
model_1.add(SimpleRNN(128))  
model_1.add(Dense(7, activation='softmax'))  
model_1.compile(loss='sparse_categorical_crossentropy',  
                optimizer='adam',  
                metrics=['accuracy'])  
_1 = model_1.fit(x_train_pad, y_train,  
                epochs=8,  
                batch_size=256,  
                validation_data=(x_test_pad, y_test))
```

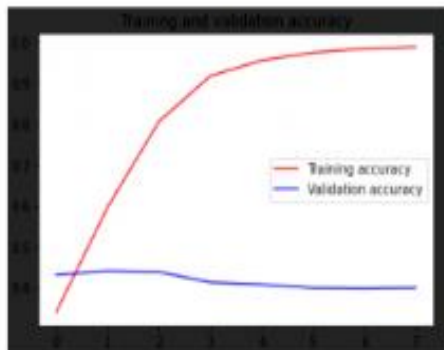
4. 프로젝트 수행 결과

결과 제시 ④ 모델 평가 및 개선



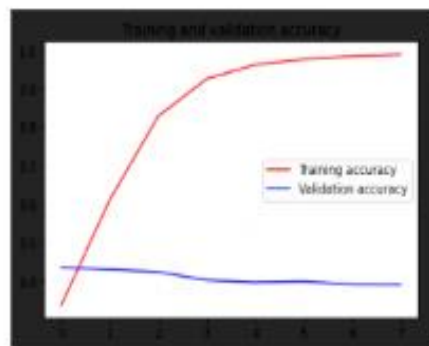
단층 RNN

acc:0.98 , val_acc:0.39



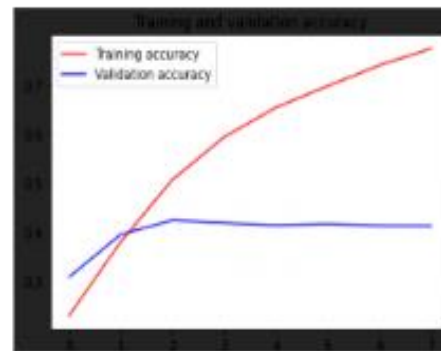
다층 RNN

acc:0.98, val_acc:0.38



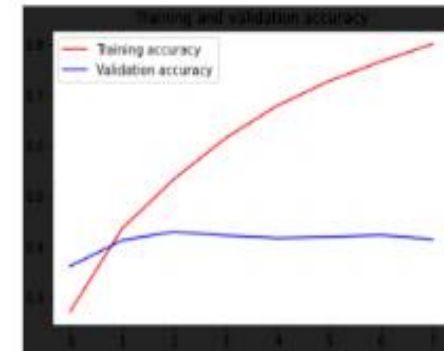
단층 LSTM

acc:0.77, val_acc:0.41



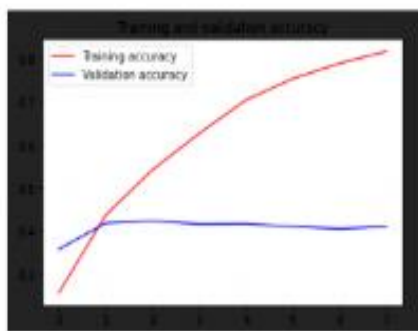
다층 LSTM

acc:0.80, val_acc:0.41



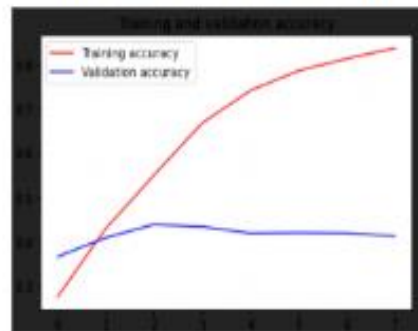
단층 Bi-LSTM

acc:0.81, val_acc:0.40



최종 다층 Bi-LSTM

acc:0.83, val_acc:0.41

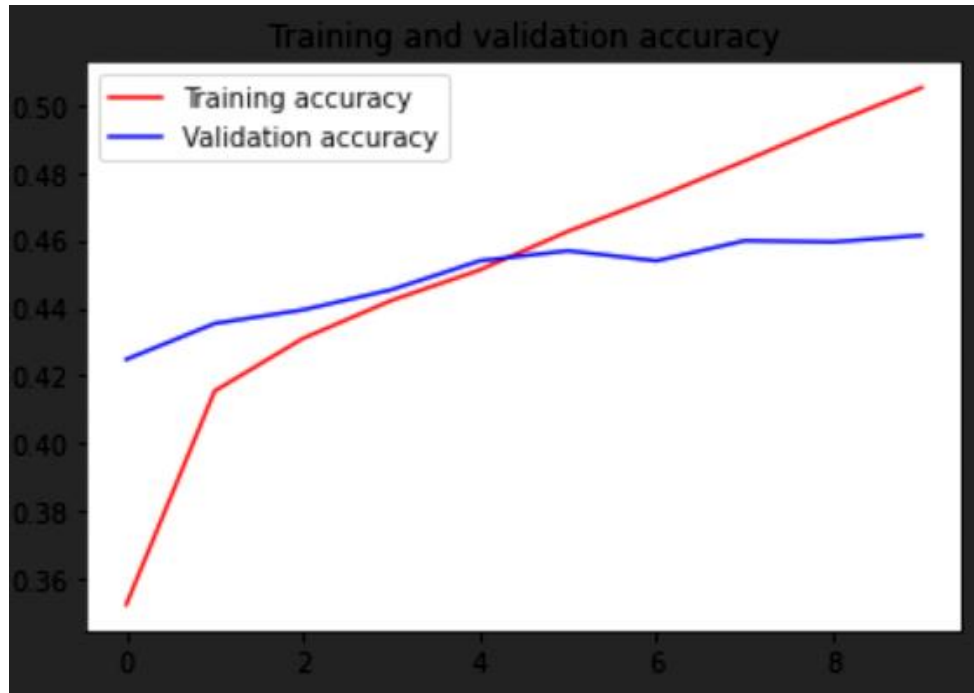


4. 프로젝트 수행 결과

결과 제시 ④ 모델 평가 및 개선

▶ 최종 모델 선정

- 최종적으로 다층 bi-LSTM 사용
- 추가적으로 Embedding 벡터는 사전 학습된 fast text를 사용
- 최종: 다층 bi-LSTM
- 하이퍼 파라미터 수정 : acc: 0.505, val_acc: 0.4615



```
Fast Text 임베딩 사전학습 모델

In [58]: fasttext_model = KeyedVectors.load_word2vec_format('cc.ko.300.vec', encoding='utf-8', limit=1000000)

In [59]: print(f"Type of Model: {type(fasttext_model)}")
print(fasttext_model.vectors.shape)
print(fasttext_model.most_similar('장남'))
len(fasttext_model[0])

...

In [60]: vocab_size = len(word_index) + 1

In [61]: embedding_matrix = np.zeros((vocab_size, 300))

In [62]: def get_vector(word):
    """
    해당 word가 word2vec에 있는 단어일 경우 임베딩 벡터를 반환
    """
    if word in fasttext_model:
        return fasttext_model[word]
    else:
        return None

for word, i in word_index.items(): # 훈련 데이터의 단어 집합에서 단어와 정수 인덱스를 1개씩 꺼내준다.
    temp = get_vector(word)       # 단어(key) 해당되는 임베딩 벡터의 300개의 값(value)을 임시 변수에 저장
    if temp is not None:          # 만약 None이 아니라면 임베딩 벡터의 값을 리턴받은 것이므로
        embedding_matrix[i] = temp # 해당 단어 위치의 행에 벡터의 값을 저장한다.

In [ ]:

In [63]: vocab_size = len(word_index) + 1
dim = 300
max_length = 25

In [73]: model_7 = Sequential()
model_7.add(Embedding(vocab_size, dim, weights=[embedding_matrix], input_length=max_length, trainable=False))
model_7.add(Bidirectional(LSTM(128, dropout=0.2, return_sequences=True)))
model_7.add(Bidirectional(LSTM(128, dropout=0.2, return_sequences=True)))
model_7.add(Bidirectional(LSTM(128, dropout=0.2, return_sequences=True)))
model_7.add(Bidirectional(LSTM(64)))
model_7.add(Dense(7, activation='softmax'))

In [74]: model_7.summary()

...

In [75]: model_7.compile(loss='sparse_categorical_crossentropy', # 'sparse_categorical_crossentropy'
optimizer='adam',
metrics=['accuracy'])

In [76]: history_7 = model_7.fit(x_train_pad, y_train,
epochs=10,
batch_size=256,
validation_data=(x_test_pad, y_test))
```


4. 프로젝트 수행 결과

결과 제시 ⑤ Flask API

▶ 서버

- flask-Restfull 프레임워크 사용
 - API 매서드와 엔드포인트를 class 단위로 만들 수 있음
 - 로컬에서만 실행 (배포 X)

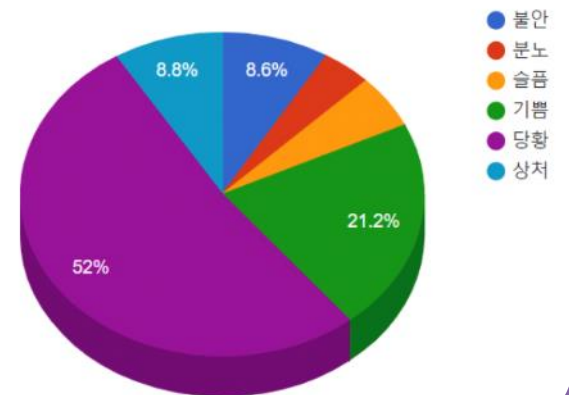
```
class HelloWorld(Resource):  
    def get(self):  
        return {'hello': 'world'}  
  
api.add_resource(HelloWorld, '/')
```

- 구현 API 엔드포인트:
 - 유튜브 동영상 댓글 저장 & 감정 통계 가져오기
 - 저장된 동영상 리스트별 댓글의 감정 통계 가져오기
 - 저장된 동영상 댓글의 많이 사용된 단어 가져오기

```
api.add_resource(CC, "/cc/<string:videoname>") #비디오 별,  
api.add_resource(list_name, "/bylist/<string:listname>") # 리스트 별 파이차트  
api.add_resource(Username, "/byuserandlist") # 리스트 + 유저 별 파이차트
```

- Response
 - jinja 템플릿을 활용해 html page에 데이터 값을 넣고 사용자에게 전달

해당 영상 댓글의 감정분포



4. 프로젝트 수행 결과

결과 제시 ⑥ 데이터베이스

▶ 데이터베이스

- **Mongodb Atlas :**
 - 클라우드 데이터 베이스로 로컬에 설치할 필요 없이
 - cluster를 생성하고 user를 만든 뒤 서버에 바로 활용할 수 있다.
- **서버와 연동:**
 - 서버에서 mongoengine 라이브 러리를 이용해서 연동
 - class 형태로 document 청사진을(collection에 저장될 doc 양식) 그려볼 수 있다.

서버 document 양식

실제 데이터베이스에 저장된 모습

4. 프로젝트 수행 결과

결과 제시 ⑥ 데이터베이스

▶ 데이터베이스

- MongoDB Atlas :
 - 클라우드 데이터 베이스로 로컬
 - cluster를 생성하고 user를 만

- 서버와 연동:

```
class Keywords(mongo.Document):
    words = mongo.ListField()
    counts = mongo.ListField()
    list_name = mongo.StringField()
    video_name = mongo.StringField()
```

서버 document 양식

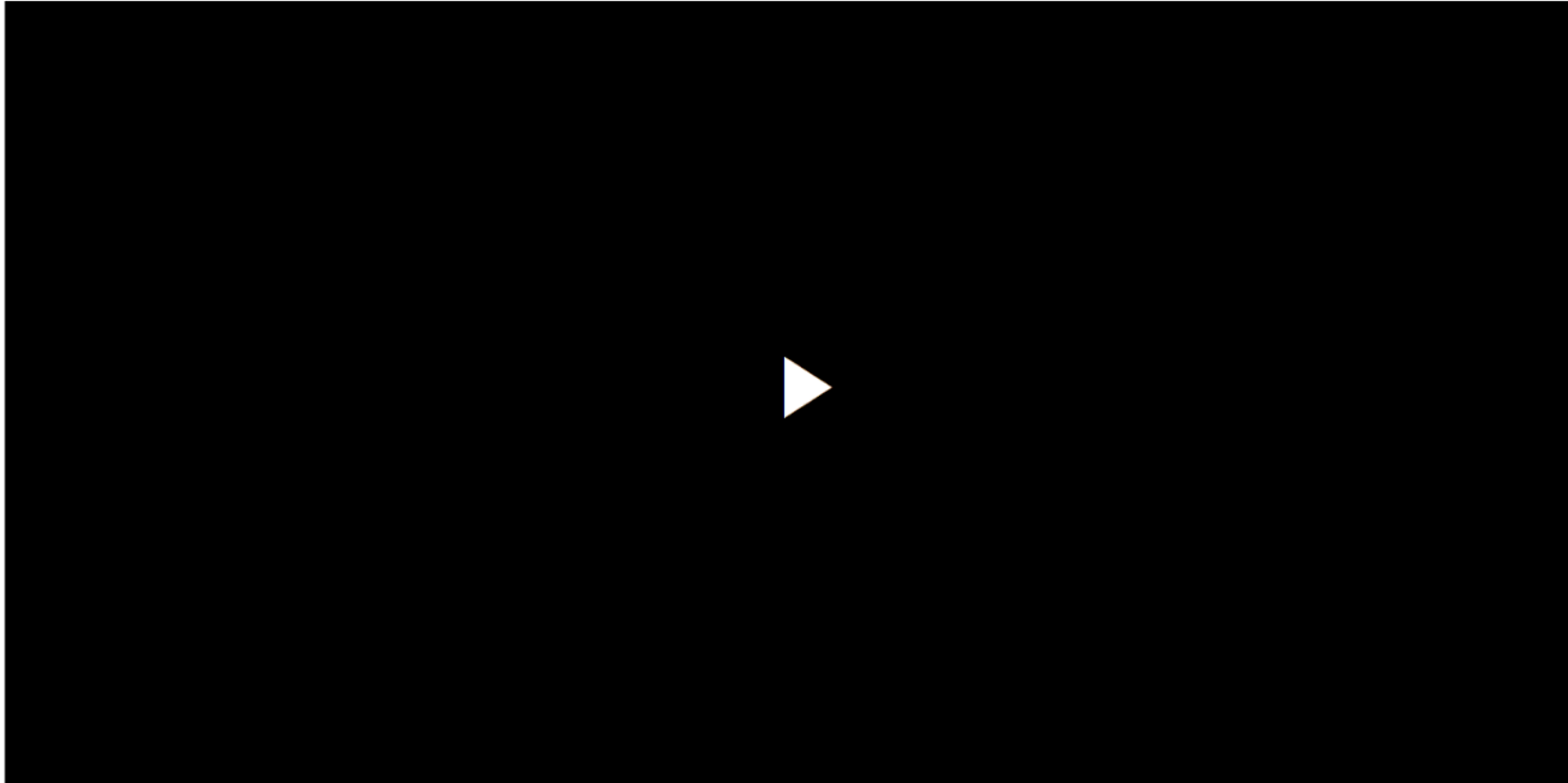
```
2: "ㅋㅋㅋ"
3: "아"
4: "너무"
5: "ㅋㅋㅋㅋ"
6: "카니발"
7: "ㅋㅋㅋㅋㅋ"
8: "와"
9: "것"
10: "왜"
11: "저"
12: "신도시"
13: "정신"
14: "나갈"
```

```
> counts: Array
list_name: "05학번이즈히어🔥"
video_name: "[05학번이즈히어] 신도시 아재는 무엇을 타는가"
```

실제 데이터베이스에 저장된 모습

4. 프로젝트 수행 결과

결과 제시 ⑦ 시연 동영상



시연 동영상 별도 제출

5. 자체 평가 의견



- **익명의 아르마딜로:**
 - 모델 학습에 사용된 데이터와 예측에 사용할 유튜브 댓글 데이터가 괴리감이 있었다.
 - 직접 크롤링하여 라벨링을 했으면 어땠을까 하는 아쉬움이 있다.
 - 한글 댓글 토큰화 및 임베딩 과정에서 형태소나 자모 분리, 단어기반, 문장기반 임베딩 등 여러 방법이 있었지만 시간의 한계로 제대로 사용하지 못했다. 추후 그 부분에 대한 개선이 필요하다.
 - 사전 학습된 임베딩 벡터 말고 유튜브 댓글 감성 분류 주제에 맞는 유튜브 댓글로 fast text 등을 학습하여 사용할 것이다.
- **익명의 기린:**
 - 초기의 계획은 사용자가 영상의 url을 입력해주면 서버에서 크롤링과 감정 분석도 진행하려고 했습니다.
 - 하지만 여러 문제로 인해 서버 외부에서 수행하게 되었습니다. 이 부분이 조금 아쉬웠습니다.
 - 하지만 진행을 하다보니 그 방법이 조금 더 api에 어울리는 것 같다는 생각이 들었습니다.
 - 이번 프로젝트를 진행하며 Flask, Restful api에 대해 많이 알게되었고 MongoDB, MetaBase 등도 한번 더 공부하며 데이터 흐름을 만들어 볼 수 있어서 좋았습니다.
- **익명의 향고래:**
 - 프론트 엔드 부분이 있어서 좀 더 사용자가 보기에 부드럽고 매끄럽게 진행되었으면 더욱 좋았을 것 같습니다.
 - 갈수록 RedOcean화 되어가는 Youtube 크리에이터 시장에서 색다른 방식으로 데이터를 분석하여 채널의 정체성을 확실히 갖춘다면 경쟁력이 생길 수 있는 만큼 충분히 매력적인 프로젝트 주제였습니다.
 - Youtube 댓글을 수집할 때 Youtube api를 사용하고, 전반적인 전처리, 정규화, DataFrame 함수를 매끄럽고, 연산속도를 올린다면 더욱 좋을 것이라 생각합니다.
 - NLP, DataFrame, 데이터 수집 등 코드스테이츠에서 배운 내용을 전반적으로 활용하고자 노력했습니다.

감사합니다.