

movie Dataset Report

Train Dataset

Preview

title	label
차기 위성 무궁화 5A·7호...아시아 시장 정조준	0
터키 에르도안 미국에 쿠르드 총사령관 넘겨라종합	4
日방위상 F35A 전투기 추락 원인 일본이 주체가 돼 조사	4
내일날씨 칼바람에 맹추위 기승...서울 12도·대관령 18도	3
새누리당 김성우 김해시장 예비후보 등록 무효 파문	6
게시판 NH투자증권 100세시대 행복리포트 49호	1
전쟁·탄압 딛고 축구로 새 희망...난민 출신 축구스타들	5
국민의당 비례대표 1번에 신용현 한국표준과학연구원 원장속보	6
與지도부 전직대통령·유족 예방...MB 정권 재창출 중요	6
CJ오쇼핑공영홈쇼핑 제휴...중소기업 판로 확대	1

Description

title	label
36542	36542
36542	7
차기 위성 무궁화 5A·7호...아시아 시장 정조준	4
1	6636

Info

Non-Null Count	Null Count	Dtype
36542	0	object
36542	0	category

Validation Dataset

Preview

title	label
한미 연구진 스스로 광합성하는 인공세포 제작 성공	0
터키 IS 가담 후 조직원 모집 외국인 4명 체포	4
사우디 홍해상 폭발 이란 유조선 조력 시도...이란은 부인종합	4
알리 호날두 넘어설 재목...공격포인트 40개 압도적	5
마니커 140억원 제3자배정 유상증자 결정	1
남북정상 핫라인 열렸다	6
미중 무역전쟁서 미국 피해 더 커...CNBC 평가	4
북한날씨 흐리고 곳에 따라 눈	3
문 대통령 적폐청산 정부방침 따랐던 중하위직에 불이익 안돼	6
LG 속절없이 7연패...이틀 연속 롯데에 1점 차 패배	5

Description

title	label
9136	9136
9136	7
한미 연구진 스스로 광합성하는 인공세포 제작 성공	4
1	1684

Info

Non-Null Count	Null Count	Dtype
9136	0	object
9136	0	category

Test Dataset

Preview

title	label
5억원 무이자 융자는 되고 7천만원 이사비는 안된다	2
왜 수소충전소만 더 멀리 떨어져야 하나 한경연 규제개혁 건의	2
항응고제 성분 코로나19에 효과...세포실험서 확인	0
실거래가 가장 비싼 역세권은 신반포역...3.3㎡당 1억 육박	1
기자회견 하는 성 소수자 단체	2
모의선거 교육 불허 선관위·교육부 각성하라	2
뮤지컬 영웅 합류한 안재욱 정성화와 다른 안중근 보여줄것	3
가짜뉴스 징벌적 손해배상제도 도입 변협 토론회	2
MBN 노조 부동산 물적분할 중단하고 소유경영 분리해야	2
포스코건설 11년 만에 더 샵 브랜드 로고 교체	1

Description

title	label
9107	9107
9107	7
5억원 무이자 융자는 되고 7천만원 이사비는 안된다	2
1	3701

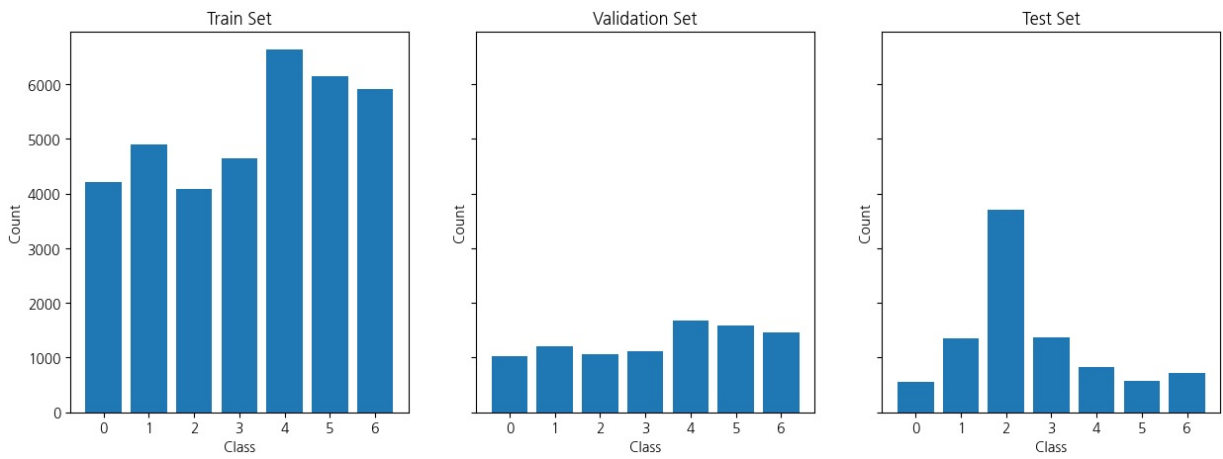
Info

Non-Null Count	Null Count	Dtype
9107	0	object
9107	0	category

Visualizations

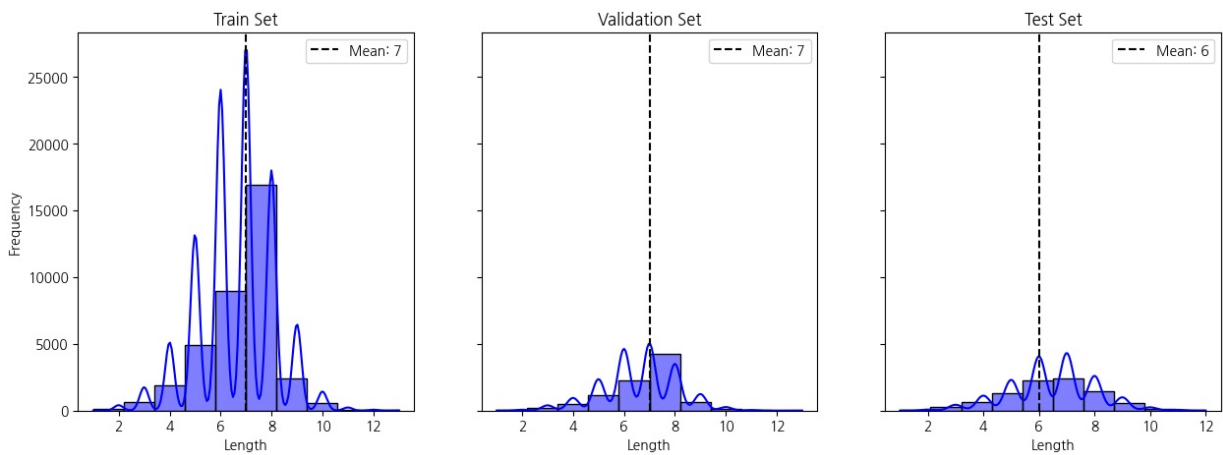
****► All Descriptors**** - `TextLength`, `SentenceCount`, `Sentiment`, `OOV`, `NonLetterCharacterPercentage`

Class Distribution



Text Length Summary

Document length distribution per dataset. Helps determine appropriate input length for models.



Dataset	Longest Sentence	Shortest Sentence	Mean Sentence Length	Sum of Sentences
Train	13	1	7	36542
Validation	13	1	7	9136
Test	12	1	6	9107

Word Cloud

