

OTT 매체에 대한 뉴스 기사로 긍정/부정 반응 예측

소프트웨어융합학과 4 학년 한주상

학번: 2020105741

초 록 (abstract)

OTT 를 너무 익숙해지게 되어서 사람들은 항상 심심하면 OTT 에 나온 콘텐츠를 보기 마련이며, OTT 의 이슈를 모르고 무분별하게 보는 경향이 나타나는 것을 문제 방향으로 삼았다. 그래서 이번에 뉴스 데이터를 이용하여 뉴스 제목, 내용, 출판 시간대(시간, 년, 월, 일)에 따라서 데이터를 여러 방법으로 분석하고 가설검정을 한 결과 세 가지 가설검정을 실행하였다. 가설 1: OTT 매체에 대한 뉴스 기사로 OTT 매체에 대한 긍정적 부정적 반응 예측 가설 2: OTT 에서 언급된 OTT 와 뉴스 제목 및 뉴스 내용에 언급된 OTT 가 사용된 키워드 빈도에 따라 OTT 가 긍정적인지 부정적인지가 달라진다. 가설 3: OTT 매체들에 언급되는 콘텐츠에 긍정적인지 부정적인지 키워드에 따라 긍정적인지 부정적인지에 따라 달라진다. 연도, 그리고 제목과 내용이 언급되어 있는 ott 매체 키워드에 언급된 수에 따라서 긍정, 부정에 영향을 주는 요인이라는 것을 알 수 있었다. 결론적으로 긍정인지 부정인지 영향력을 확인해 보고 싶었다. 그리고 제목 및 내용이 얼마나 OTT 에 대해서 얼마나 언급되고 있는지 얼마나 강조하느냐에 따라서 영향력을 미치는지에 대한 인사이트를 보여준다.

1. 서 론

OTT 매체에 대해서 요즘 넷플릭스, 유튜브, 티빙과 같이 온라인 동영상 서비스를 이용하여 예전에 비해서 많이 사용되고 있으며, OTT 사용률이 점점 증가할 예정이다. 그만큼 혼자서 사람들이 OTT 를 예전에 비해 많이 보고 있다. 정다준씨에 의하면[1] 2023 년 기준으로 방송매체 이용행태조사를 결과를 통해서 10 ~ 20 대가 가장 많이 시청하고 있으며, 유튜브, 넷플릭스, 티빙이 가장 대표적으로 보고 있는 OTT 매체들이다. 하지만, 이러한 OTT 매체들도 긍정적인 영향인 것도 있고, 부정적인 영향에 미치는 것이 있다. 박수선 씨 외 1 인에 의하면[2] PD 들은 자본력으로 앞세운 OTT 의 콘텐츠 시장 잠식에 대해서 TV 시청자들의 이탈과 방송사 수익 감소, 흥미위주의 콘텐츠 증가, 방송 프로그램 품질 저하의 결과로 이어지기도 한다. 그리고 OTT 의존 심화로 인하여 이를 대응하기가 어려워지고 있다는 것이 지금도 해결되고 있는 과제입니다.

지금 OTT 의 무분별한 시청 및 정보에 대한 이슈에 대해서 긍정적인지 부정적인지 파악 못하고 OTT 에 의존하는 경향이 늘어나고 있다. Prasad, S. S 에 의하면[3] OTT 가 직접적으로 친구, 부모와 같이 가족에게 어떠한 웰빙에 영향을 미치는지, 사회에 어떠한 것을 기여하는지, 개인의 교육에 건강에 영향을 미치는지 가설 설정하여 검증을 하지만, 다 가설에 기각되며, 부정적인 영향에 미친다는 것을 확인할 수 있다. Mulla, T.에 의하면[4] 미디어 소비의 변화는 기존 유통 채널에 우회하여서 미디어 분야에서 스트리밍 플랫폼의 부상으로 콘텐츠, 가격, 유연성, 편의성, 인지도, 유용성, 인지도, 즐거움, 모든 제약으로부터 자유로움에 대한 욕구, 오락적 가치, 사회화, 문화적 포용성, 정주행, 도입으로 주요 요인으로 분석을 한다. 기존 문제 해결을 위해서 Shiri.C 에 의하면[5] 문제를 보완하고자 OTT 커뮤니케이션 서비스에 영향을 미치는 요인을 체계적으로 통합하고, 이를 바탕으로 연구 프레임워크를 제안하고, 향후 연구자들에게 미래 연구 로드맵을 제공합니다.

Prasad, S.S 에 의하면[6] OTT 미디어가 가정과 사회에 미치는 영향을 조사하고, 개인의 직업, 건강에 대해서 영향을 규명하고자 한다. 조해령님이 말하기[7]를 OTT 사용에 대한 문제에 대한

거래조건 실태조사를 확인한 결과, 요금제 가격, 이용약관에 따라서 나타내어 있다. OTT 관련하여서 해결하는 과정을 통해서 대체적으로 뉴스 및 이슈에 고려하고 있는지 명확하지 않으며, OTT 에 대해서 문제를 해결하는 데에 있어 소프트웨어적 관점에서 봤을 때 현실적으로 접근하기가 어렵다. 또한, OTT 에 대해서 긍정/부정으로 감성 분석하는 데에 있어서 고객 관점에서 더욱 고려하여서 하되, 사람들이 익숙하지 않은 업데이트를 시행도 하고 있어. 이것에 대해서 적응하는 데 시간 효율적으로 낭비될 수 있다는 것이 한계이다. OTT 에서 무분별하게 시청도 하고 있고, 모든 사람들이 OTT 매체에 대해서 접하다 보니, OTT 에 업로드된 콘텐츠를 무분별하게 접하다 보니, OTT 에 대해서 이슈가 어떤 것이 올라와 있고, 이 OTT 가 지금 이슈적으로 어떠한 상황인지 일일이 뉴스를 찾아보았을 때 일일이 읽어보기도 하여 정보가 나에게 OTT 가 긍정적인지 부정적인지 한눈에 확인하고 분별하는데 오래 걸립니다. 그래서 OTT 의 무분별한 사용을 벗어나고, OTT 에 있는 이슈들을 한눈에 비교하여 OTT 를 분별하여 볼 수 있도록 방향을 잡아주고 싶어서 분석하려고 한다. 이를 위해서 저가 가설 3 가지를 세웠습니다. 첫째, OTT 에서 시간대에 따라서 년/월/일에 따라서 OTT 가 긍정적인지 부정적인지가 달라진다. 둘째, OTT 에서 언급된 OTT 와 뉴스 제목 및 뉴스 내용에 언급된 OTT 가 사용된 키워드 빈도에 따라 OTT 가 긍정적인지 부정적인지가 달라진다. 셋째, OTT 매체들에 언급되는 콘텐츠에 긍정적인지 부정적인지 키워드에 따라 긍정적인지 부정적인지가 달라진다. 분석하고자 하는 것은 OTT 에 속한 뉴스에 해당하는 제목들, 내용, 그리고 뉴스를 출판한 시간에 따라서, 이 뉴스가 긍정적인지 부정적인지 분석하는 것이다. 각 가설별 분석 방법을 활용하였습니다. OTT 에 대해서 이슈를 이용하여 긍정적인지 부정적인지 모델로 예측하여 OTT 를 사용할 때에도, 긍정적이면 평가에 따라 사람들이 OTT 의 신뢰도를 확인하여 더욱 분별하여 나아가도록 하기 위함에 의의가 있습니다.

2. 관련/연구

2.1. OTT 의 방향성

한국정보화진흥원에 의하면¹ OTT 가 종류가 다각화된 만큼 데이터 분석을 통한 커머스의 개인화가 증가하여 사용자별로 고도화하여 제공이 되어질 것이다. OTT 다각화를 통해서 성공 요인으로 전략, 기술, 콘텐츠의 변화로 새로운 팬층들을 강화하기 위해서 관심을 유도하기 위한 전략을 세운다. OTT 에 대해서 영향에 대해서 사람들에게 유익하기도 하고, 유해하기도 하여 양날의 검처럼 양면성을 띄고 있어 OTT 을 사용하는 데에 신중하게 결정합니다.. 또한, 홍태호님에 의하면[8] 여러가지 토픽에 따라서 감성 분석을 실험을 하여 LDA 를 통해서 수집된 데이터 내에서 키워드로 분류하도록 합니다. 분석을 통해서 모델을 실험을 한다. 토픽별 키워드에 대해서 선정을 하는데에 있어서 키워드를 선정하여 나타냅니다.

2.2. 정보 감성 분석(Sentiment Analysis)

Medhat, W 외 1 인에 의하면[9] 감정 감지(ED), 자원 구축(BR), 전이 학습 (TL)이 포함합니다. 감정은 감정을 추출하고 분석하는 것을 목표로 합니다. 감정은 문장에 명시적이거나 암묵적일 수 있습니다. 전이 학습 또는 교차 도메인 분류는 한 도메인의 데이터를 분석한 다음 대상 도메인에서 결과를 사용하는 것과 관련 있습니다. 자원 구축은 어휘집, 의견 표현이 극성에 따라 주석이 달린 코퍼스, 그리고 때로는 사전을 만드는 것을 목표로 합니다. Taboada, M 에 의하면[10] 감정 분석은 언어학과 컴퓨터 과학의 교차점에서 성장하는 분야로, 텍스트에 포함된 감정을 자동으로 파악하려는 시도입니다. 감정은 언어를 통해 표현되는 긍정적 또는 부정적 평가로 특징지을 수 있습니다. 감정 분석의 일반적인 응용 분야에는 온라인에 게시된 리뷰(영화, 책 또는 소비재)가 리뷰 대상 항목에 대해 긍정적 또는 부정적인지 자동으로 판단하는 것이 포함됩니다. 감정 분석은 이제 기업, 마케터 및 정치 분석가가 수행하는 소셜 미디어 분석 레퍼토리에서 일반적인 도구입니다. 감정 분석에 대한 연구는 텍스트의 긍정적 및 부정적 단어, 해당 단어의 맥락 및 텍스트의 언어적 구조에서 정보를 추출합니다. 이 간략한 리뷰에서는 특히 언어적 지식이 감정을 자동으로 파악하는 작업에 기여할 수 있는지 살펴봅니다. 이러한 연구와 다르게 뉴스의 OTT 의 제목 및 내용을 통하여 최근에 사회에 이슈가 되고 있는 것이 무엇인지에 대해서 확인 및 시간대로 뉴스 데이터들을 모아서 사람들의 사용자 취향 추세 및 여론이 긍정적인 구간과 부정적인 구간을 서로 비교하는 관점에서 접근합니다. 또한, 뉴스에 OTT 키워드가 언급되고 있는 빈도에 따라서 여론이 긍정적인지 부정적인지 분석을 하기 위해 가설을 접근합니다.

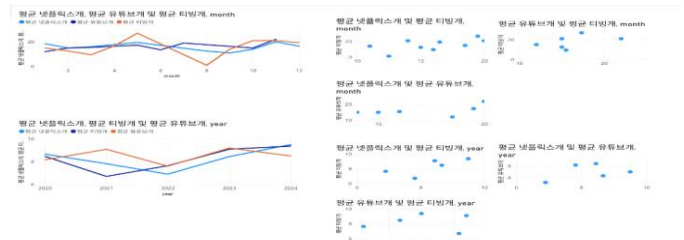
3. 본 론

ATTRIBUTE	데이터 설명
ott_type	OTT 의 종류 이름(넷플릭스, 티빙, 유튜브)(object)
year	뉴스 공개 날짜의 년(2020 ~ 2024 년)(int64)
month	뉴스 공개 날짜의 월(1 ~ 12 월)(int64)
day	뉴스 공개 날짜의 일(1 ~ 31 일)(int64)
source	뉴스 출처 회사(object)
title	뉴스 제목(object)
comment	뉴스 내용(object)
label	뉴스가 긍정이면 1, 부정이면 0

[표 1] 준비한 데이터의 ATTRIBUTE 와 그에 대한 설명

위의 표와 같이 ott 에 대한 이슈에 필요한 요인들을 확인하기 위해서 ott 에 해당하는 미디어의 종류, 뉴스가 보도되었던 시간대(년/월/일), 뉴스 출처, 제목, 내용을 수집을 하여 2020 년 ~ 2024 년 동안 검색어로 ‘넷플릭스’, ‘유튜브’, ‘티빙’ 각각을 검색하여 사용합니다. 한국어로 검색, 필터를 이용하여 2020, 2021, 2022, 2023, 2024 년을 통해서 필터링합니다. Google 홈페이지에서 ‘한국어 웹’, 2024 년을 입력하면 ‘기간 설정’ → 시작일 ‘2024년 1월 1일’ 종료일 ‘2024년 12월 31일’로 필터링을 한 다음, 이 홈페이지를 Listly 로 이용하여 부분 추출을 하여 모든 페이지를 일일이 모읍니다. 그리고, 이 뉴스 제목 및 내용에서 이 내용이 긍정적인 내용인지 부정적인 내용인지 직접 일일이 데이터 라벨링(긍정이면: 1, 부정이면: 0)을 하였습니다.

3.1. 첫 번째 가설: OTT 에서 시간대에 따라서 년/월/일에 따라서 OTT 가 긍정적인지 부정적인지가 달라질 것이다.



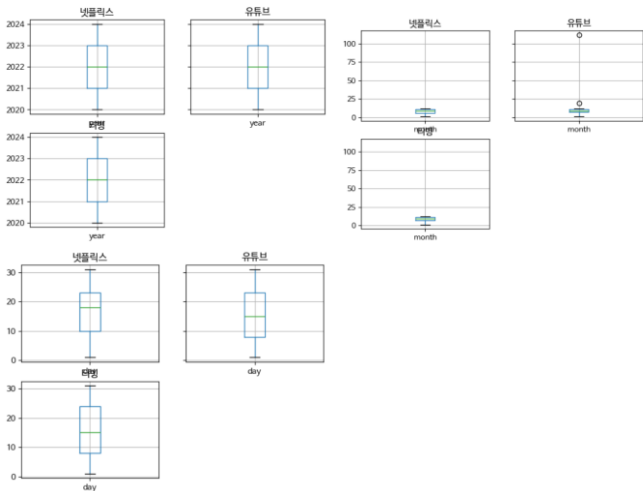
[그림 1] OTT 별로 year, month, day 에 대한 상관 관계 및 그래프

	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.857	0.857	1.0	0.923
DecisionTreeClassifier				
SVM Classifier				
Random Forest Classifier				
GaussianNB (Naive	0.142	0.0	0.0	0.0

1

Bayes)				
BaggingClassifier – DecisionClassifier				
BaggingClassifier–SVMClassifier				
Stacking base_model : LogisticRegression, RandfomForest, Naive Bayes stacked_model: SVC	0.857	0.857	1.0	0.923

[표 2] 분류 방법을 이용하여 시간대(year, month, day)의 요인에 따라서 긍정/부정 분류 분석 성능 평가표



[그림 2] label 의 값에 따라 year, month, day 에 대한 분포 boxplot 시각화

Label 값에 따라서 각각 시간대 Year, Month, Day 의 차이점을 시각화로 나타내개 위해서 다음과 같은 결과가 나옵니다. 독립적으로 label 의 값에 0, 1 에 따라서 분포가 어떻게 나타내는지 boxplot 그래프로 표현하였습니다.

유의 확률(0.05)		One-side less	One-side greater	Two-side
One Sample T-test	H0: mean of year = 2022 H1: mean of year ≠ 2022	0.9483	0.0516	0.1032
	H0: mean of month =	2.412e-21	1.0	4.824e-21

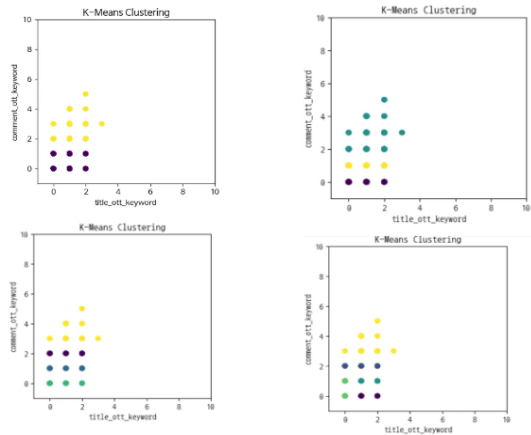
Two Sample Independent T-test	9 H1: mean of month ≠ 9			
	H0: mean of day = 15 H1: mean of day ≠ 15	0.9999	1.7745e-15	3.5491e-15
	label 이 0 인 year 의 표본과 label 이 1 인 year 차이에 대한 유의	3.3739e-05	0.9999	6.7479e-05
	label 이 0 인 month 의 표본과 label 이 1 인 month 의 차이에 대한 유의	0.0736	0.9263	0.1473
	label 이 0 인 day 의 표본과 label 이 1 인 day 의 차이가 유의	0.4846	0.5153	0.9692

[표 3] One Sample, Two sample Year, Month, Day 에 따른 T-test Statistics 분석

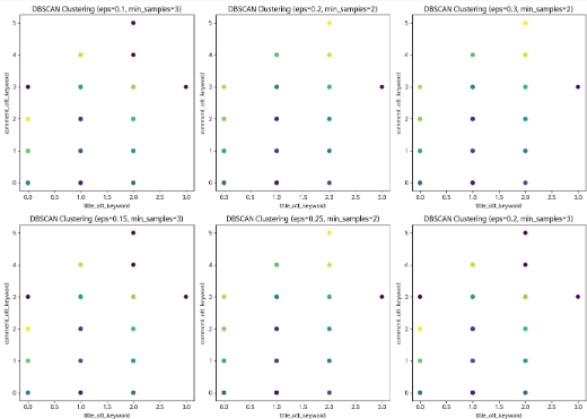
가설 검정을 한 결과 유의 확률을 0.05 로 잡았을 때 One-sample Test 일 때는 단측 검정 및 양측 검정을 하였을 때 year 의 경우에는 단측 검정 양측검정 모두 가설에 기각됩니다. 그리고 Two sample Independent T-test 일 때 label 이 다를 때 마다 유의하게 차이가 있으며, 데이터 편향이 치우쳐져 있기도 하다는 것을 알게 되었습니다. 분류 방법으로 metrics 를 비교하였을 때, 성능적으로 다 비슷하지만, Roc-Auc score 에서 점수가 1 에 가까운 모델은 없으며 다 거의 0.5 와 거의 가까운 수치를 다다르는 데 공통적이다. 따라서, 가설과 달리 이 결과로 봤을 때, 가설 설정으로는 연도에 따라 유의미합니다. ROC-AUC Score 가 0.5 에 가까울 경우 거의 랜덤으로 모델링을 이용하여 랜덤 샘플링을 하여 모델을 분석하는 기능과 거의 비슷하다. 그리고 연도에 따라서 이 데이터를 통해 감정적으로 긍정인지 부정인지 가장 영향이 있는 요인이라고 알게 되었다. 그리고 연도에 따라 ott 에 대해서 내용이 긍정인지 부정인지 시간대에서 가장 영향이 작용된다는 것을 알 수 있다.

3.2. 두 번째 가설: OTT 에서 언급된 OTT 와 뉴스 제목 및

뉴스 내용에 언급된 OTT 가 사용된 키워드 빈도에 따라 OTT 가 긍정적인지 부정적인지가 달라진다



[그림 3] K-mean 기법을 k 값에 따라 군집 시각화



[그림 4] DBSCAN 기법으로 도출한 군집 시각화

	Sihouette Score	Homogeneity	Completeness	V-measure
K-mean(k = 2)	0.6739	0.0001	9.1728e-05	0.0001
K-mean(k = 3)	0.8259	0.0040	0.0016	0.0023
K-mean(k = 4)	0.8776	0.0041	0.0014	0.0021
K-mean(k = 5)	0.9275	0.0103	0.0033	0.0050
DBSCAN	0.9997	0.0225	0.0064	0.0100

[표 4] K-mean, DBSCAN 에 대한 Metrics

이것에 대해서 metrics 에 대해서 DBSCAN 이 K-means(k = 2)일 때보다 더 지표가 높다는 것을 알게 되었습니다.

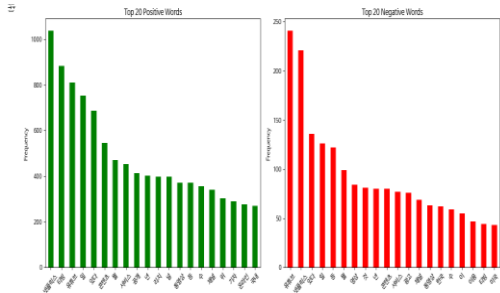
		One-	One-	Two-
--	--	------	------	------

		side less	side great	side
One-sample T-test	H0: mean of title ott keyword = 1 H1: mean of title ott keyword ≠ 1	1.8539 e-12	0.9999	3.707 e-12
	H0: mean of content ott keyword = 1 H1: mean of content ott keyword ≠ 1	1.8539 e-12	1.8539 e-12	0.0004
Two-sample Independent T-test	title_ott_keyword 의 차이가 유의	3.5853 e-11	0.9999	7.1707 e-11
	comment_ott_keyword 의 차이가 유의	0.0094	0.9905	0.0188

[표 5] One Sample, Two sample Title, Comment 언급한 횟수에 따른 T-test Statistics 분석

k_cluster 의 개수가 높아질수록 훨씬 모델 성능이 높아진다는 것을 알 수 있다. 그리고 제목, 내용에 언급된 ott 매체의 키워드에 언급하는 데에 있어서 유의미한 차이가 있으며, ott 매체에 있는 키워드가 긍정적인지 부정적인지에 따라 차이가 있으며, 영향이 있습니다. DBSCAN 도 마찬가지로 epsilon, minPts 를 각각 0.3, 2 로 가정하였을 때 DBSCAN 기법을 이용하여 군집끼리 하여 comment_ott_keyword 의 수가 군집 수가 2 개보다 넘게 분포되어 있습니다.

3.3. 세 번째 가설: OTT 매체들에 언급되는 콘텐츠에 긍정적인지 부정적인지 키워드에 따라 긍정적인지 부정적인지에 따라 달라진다



[그림 4] Text Analysis 에 따른 빈도가 많은 20 개 순으로 나온 키워드의 시각화

	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.8535	0.8823	0.9550	0.9172
DecisionTree Classifier	0.8128	0.8831	0.8988	0.8909

SVM Classifier	0.8224	0.8901	0.9026	0.8963
Random Forest Classifier	0.8614	0.8675	0.9878	0.9238
KNeighbors Classifier	0.8503	0.8600	0.9840	0.9179
Bagging Classifier Decision Classifier	0.8367	0.8795	0.9363	0.9070
Bagging Classifier SVM Classifier	0.8519	0.8550	0.9943	0.9194

[표 6] Text Analysis & Classification 으로 명사 형용사 분석하여 분석한 결과

이와 같은 경우에는 Text 분석 요소들을 분석하는 데에 있어 가장 빈도가 높은 상위 단어들을 분석해본 결과 나온 단어들 중에서 긍정 부정 둘 다 뜻이 긍정, 부정의 뜻을 가진 단어들은 없었으며, 공통적으로 ‘ott 이름’, ‘있다’, ‘일’, ‘년’, ‘월’, ‘콘텐츠’, ‘서비스’, ‘광고’와 같은 표현이 많이 분포되어 있어서 기존 가설과 다르게 긍정적인 단어, 부정적인 단어가 생각보다 부족하다는 사실을 알게 되었습니다. 하지만, 품사 태깅을 통해 분리된 명사, 형용사를 모아서 모델링을 해본 결과 기본 적인 모델인, Decision Tree Classifier, SVM Classifier, Random Forest Classifier KNeighbors Classifier 가 가장 정확한 성능을 향상시킨다는 것을 확인할 수 있다. 그리고 ROC_AUC score 에 대해서 첫 번째 가설과 달리 점수가 높아지긴 했지만, 0.5 ~ 0.6 값 사이로 미미하게 성능이 향상이 되었다는 것을 알 수 있습니다.

Text Analysis 을 이용해서 하였을 때 추출한 품사 태깅으로 명사를 추출하였을 때 수치를 측정한 결과 거의 1 과 가까운 지표로 나오게 되어있지만, 가장 빈도가 있는 단어들이 긍정, 부정에 관련된 단어가 아니라는 점에서 이 가설을 결론 내는 데에 있어 빈도가 가장 큰 단어들에 속하지 않지만, 키워드를 통해서 긍정 부정을 판별할 수 있다는 것은 사실입니다.

4. 결론

본론 내용을 통해서 ott 매체에 대해서 지금 어떻게 뉴스에서 이슈가 어떤지에 대해서 긍정적 반응인지 부정적인 반응에 작용되는 요인으로 시간대 중 연도, 그리고 제목과 내용이 언급되어 있는 ott 매체 키워드에 언급된 수에 따라서 긍정, 부정을 군집화해도 어느 정도로 영향에 작용된다는 것을 알 수 있었다. 텍스트 분석을 통해서 명사들을 추출한 결과 아쉽게도, 가장 빈도가 많이 나온 키워드들이 긍정, 부정에 직접적으로 의미가 있는 키워드들이 가장 많이 언급이 안 되어 있다는 것이 이 연구에 있어서 한계점입니다. 하지만, 이를 분석한 결과 가장

빈도가 많이 나온 키워드들로 학습이 되어서 긍정 부정을 판별할 수 있다는 것임을 알 수 있습니다.

이 본론의 결과를 통해서 이 분석은 OTT 에 대한 이슈들을 모아서 사람들에게 무분별하게 OTT 를 보지 않고 OTT 에 대해서 시간이 지나면서 긍정, 부정의 흐름이 변화에 영향에 미친다는 것을 보여주었다. 그리고 OTT 에 대해서 많이 언급하거나 적게 언급할수록 더욱 강조를 해주어 이 OTT 가 얼마나 중요하게 평가로 인식 받고 있는지도 보여줄 수 있다는 것을 알 수 있었다. 이 분석을 통해서 나중에 OTT 에 대한 이슈로 감성 분석하는 데에 있어서 텍스트 키워드에 대해 긍정 부정에 직접적인 단어 및 직접적으로 작용되지 않은 단어가 얼마나 감성 분석에 대해서 영향에 미치는지에 대해서 이 연구를 통해 알고 싶었다. 또한, 언론에 있는 시간대 (년, 월, 일)들도 OTT 에 대한 이슈에 긍정인지 부정인지 영향력을 확인해 보고 싶었다. 그리고 제목 및 내용이 얼마나 OTT 에 대해서 얼마나 언급되고 있는지 얼마나 강조하느냐에 따라서 영향력을 미치고 있는지 알고 싶었습니다.

5. 참고 문헌

- [1] 정다준. “올해 OTT 이용률 77%… 유튜브·넷플릭스·티빙 순”. 천지일보. 2023.
- [2] 박수선, 이준엽.” PD 들이 보는 OTT, “방송시장 잠식” · “자극제 될 것” 흔재”.Pd.JOURNAL.2020
- [3] Prasad, S. S. “Impact of OTT media on the society: insights from path analysis”. Asia-Pacific Journal of Business Administration. 14(2). pp.185~198. 2022.
- [4] Mulla, T. “Assessing the factors influencing the adoption of over-the-top streaming platforms: A literature review from 2007 to 2021”. Telematics and Informatics, 69, p.101797. 2022.
- [5] Shri, C., Kishnani, S., Kishnani, S., Joshi, G., & Gupta, V. “FACTORS INFLUENCING OTT COMMUNICATION SERVICES: AN INTEGRATED REVIEW AND A RESEARCH ROADMAP FOR THE FUTURE”. Journal of Content, Community & Communication.18. 2023
- [6] Prasad, S. S. “Impact of OTT media on the society: insights from path analysis”. Asia-Pacific Journal of Business Administration. 14(2). pp.185~198. 2022.
- [7] 조해령. “OTT 서비스 관련 소비자 문제 실태조사”. 한국소비자원. pp. 1~81. 2024.
- [8] 홍태호. “LDA 를 이용 LDA 를 리뷰의 다중 토픽별 감성분석- TripAdvisor 사례를 중심으로”. 정보시스템연구. 27(1), pp.89~110.2018.
- [9] Medhat, W., Hassan, A., & Korashy, H. “Sentiment analysis algorithms and applications: A survey”. Ain Shams engineering journal. 5(4), pp.1093~1113. 2014.
- [10] Taboada, M. “Sentiment analysis: An overview from linguistics”. Annual Review of Linguistics. 2(1), 325-347. 2016.