

모닝스타 해외주식 태깅 모델 개발

자산관리전략부 / 2021. 08. 26

진정민



본 자료는 NH투자증권이 제작한 것이며, 투자권유를 위한 광고물로 활용될 수 없고, 투자자에게 배포될 수 없습니다. 본 자료에 수록된 내용은 신뢰할만한 자료 및 정보로부터 얻어진 것이나, 당사는 그 정확성이나 안정성을 보장할 수 없습니다. 따라서 어떠한 경우에도 본 자료는 고객의 투자 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다.

Contents

1. 과제 개요
2. 활용 데이터
3. 분석(개발) 방법론
4. 분석 결과
5. 결론 및 활용

1. 과제(개발) 개요

✓ 해외주식 리서치 리포트 내 키워드를 추출하는 태깅 모델 개발

분석(개발) 관련 주요 사항

- 모닝스타 리서치 애널리스트 리포트 데이터를 활용한 리테일 고객 대상 서비스 준비 중
- 향후 모닝스타 해외주식 리포트의 키워드 추출 니즈가 있을 것이라 판단하여, 인턴 프로젝트 겸 당 부서 파일럿 프로젝트로 진행
- 모닝스타에서 미국, 중국, 홍콩, 일본의 770여개 종목의 상승 하락 요인에 대한 코멘트 및 적정주가, 애널리스트 투자전망 코멘트 제공



키워드 추출(태깅) 모델

분석 기간

- 과제 : 2021. 08 .02~ 2021.08.25(약 4주 진행)
- 조예린 사원 이어 받아 추후 분석 진행 예정

종목명: 엔비디아 (NVDA)

- #ARM인수준비
- #적정가격 상향
- #인공지능
- #딥러닝
- #자율주행
- #그래픽 칩

1. 과제(개발) 개요 – 결과 시연 영상

1 티커 입력 -> 해당 종목의 키워드 추출

2 키워드 입력 -> 관련 종목 확인 가능

* soykeyword 이용 : 파이썬 내장 함수

```
dtm_frame_tri = pd.DataFrame(data = dtm_np_tri, columns = feature_names_tri)
dtm_frame_tri = dtm_frame_tri.join(data['MOS_TCK_CD'], how='left')
dtm_frame_tri.set_index('MOS_TCK_CD', inplace = True)
```

In [*]: TCK = input('TCK를 입력하세요: ')

TCK를 입력하세요:

```
In [172]: #unigram
# 단어 빈도수 측정하기
c = Counter(data_set_index['new_nouns'].loc[TCK])
word_common_list = c.most_common(15)
word_data = pd.DataFrame(word_common_list, columns = ['word', 'count'])
count_data = pd.DataFrame(word_data['count'])
scaler = MinMaxScaler().fit(count_data)
count_scaled = scaler.transform(count_data)
word_data['count_scaled'] = count_scaled

tf_frame_uni = pd.DataFrame(data = dtm_frame_uni, columns = feature_names_uni).loc[TCK].sort_values(ascending=False)
tf_frame_uni = pd.DataFrame(tf_frame_uni).reset_index()
tf_frame_uni.columns = ['word', 'tf_value']
counter_tf_uni = pd.merge(word_data, tf_frame_uni, how = 'inner', on = 'word')
vocabs_uni = list(counter_tf_uni['word'])
word_vectors_list_uni = [model_ft.wv[v] for v in vocabs_uni]

#bigram
c = Counter(data_set_index['bigram'].loc[TCK])
word_common_list = c.most_common(30)
word_data = pd.DataFrame(word_common_list, columns = ['word', 'count'])
count_data = pd.DataFrame(word_data['count'])
scaler = MinMaxScaler().fit(count_data)
count_scaled = scaler.transform(count_data)
word_data['count_scaled'] = count_scaled

tf_frame_bi = pd.DataFrame(data = dtm_frame_bi, columns = feature_names_bi).loc[TCK].sort_values(ascending=False)
tf_frame_bi = pd.DataFrame(tf_frame_bi).reset_index()
tf_frame_bi.columns = ['word', 'tf_value']

counter_tf_bi = pd.merge(word_data, tf_frame_bi, how = 'inner', on = 'word')

vocabs_bi = list(counter_tf_bi['word'])
word_vectors_list_bi = [model_ft.bigram.wv[v] for v in vocabs_bi]

#trigram
c = Counter(data_set_index['trigram'].loc[TCK])
word_common_list = c.most_common(30)
word_data = pd.DataFrame(word_common_list, columns = ['word', 'count'])
word_data = word_data[word_data['count']>=2]
count_data = pd.DataFrame(word_data['count'])
```

- 모닝스타 애널리스트 리서치 리포트 PDF 파일

제목명

애널리스트 투자 의견

텍스트 데이터 추출 - 사내 EDW 적재

	영문 테이블명	한글 테이블명
1	MOS_RPT_TIL_NM	모닝스타보고제목명
2	MOS_RPT_CTET_CTS	모닝스타보고컨텐츠내용
3	MOS_RPT_BUL_ONN_CTS	모닝스타보고강세의견내용
4	MOS_RPT_BHM_ONN_CTS	모닝스타보고약세의견내용
5	MOS_ANA_MMO_NM	모닝스타애널리스트메모명

BSE_DT	MOS_JEM_ID	KRL_ENG_NM_CD	MOS_RPT_TIL_NM	MOS_RPT_PBS_DTM	MOS_RPT_CTET_CT
0	20170320	0P0000029Z	프랑스 탈레를 회사 Orange, liad, Bouygues에 대한 견 해 수정 ...	20170522003100	Orange는 여러 국가에서 서 사업을 운영하고 있 지만 입지 국가와 관련 한 지역은 프랑스
1	20170320	0P000002AB	Franklin Should Continue to See Benefits From ...	20170722232300	A confluence of several issues—poor relative
2	20170320	0P000002AB	Franklin Legg Mason 인수로 자습 되는 혜택	20170623004800	열악한 상대적 실적 자 성장, 자습을 지수 반 상품의 성장, 소액 주자들의

텍스트 데이터 (예시 - NVIDIA)

엔비디아, 급성장하는 데이터 센터 비즈니스... GPU 전문성이 넓은 경제적 해자 보장 엔비디아 Nvidia는 컴퓨팅 플랫폼의 시카 경험을 향상하는 그래픽처리장치 GPU의 선도적 설계기업이다. 엔비디아의 칩은 게이밍용 하이엔드 PC와 데이터센터를 포함한 다양한 최종 시장에서 활용되고 있다. 카드 CAD, 영상...

관련 테이블

	영문 테이블명	한글 테이블명
6	MOS_ISS_CO_DSC	모닝스타발급회사설명

제목
+
투자 의견
+
Bull & Bear
+
발급 회사 설명

2. 활용 데이터

Amazon.com Inc AMZN (XNAS)

모닝스타 등급	★★★★	주가/총가	3,327.59 USD	주당 이익	4200.00 USD
30 7월 2021	30 7월 2021	30 7월 2021	30 4월 2021	30 4월 2021	30 4월 2021
21:25, UTC	02:09, UTC				

모닝스타 핵심 지표	매일 리스	평가
경영적 성과	높음	높음
가치평가	★★★★	적정평가
분배성	높음	매우 높음
재무 건전성	—	강력

출처: 모닝스타 주식 분석

정량적 가치평가

	AMZN	
과소평가	적정평가	과대평가

주가/총가	3,327.59	주가/이익	4200.00
PER	57.9	PER	16.2
예상 PER	55.8	—	12.3
주가/현금흐름	28.8	30.8	10.2
주가/장여현금흐름	241.8	63.4	17.7
매출수익률 %	—	—	2.48

출처: Morningstar

상승론자의 견해

— 아마존은 분명 전자상거래 분야를 이끄는 리더 기업으로, 타의 추종을 불허하는 규모를 누리며 성장 기회와 최상의 고객 경험 향상을 위한 투자를 지속하고 있다.

— 마진율이 높은 광고 및 아마존 웹 서비스의 사업 부문은 아마존 전체 평균보다 더 빠른 속도로 성장하고 있으며, 이는 향후 수년간 지속적으로 수익성을 확대하는 역할을 할 것으로 예상된다.

— 아마존 프라임(Amazon Prime) 멤버십은 아마존에서 많은 소비를 하는 고객들을 끌어들이고 유지하는 것에 도움을 주며, 반복적이고 높은 마진의 매출액을 창출하는 한편 강력한 네트워크 효과를 강화하는 작용을 한다.

하락론자의 견해

— 아마존을 포함한 대형 기술 기업 대상으로 한 규제 관련 우려가 부상하고 있다. 이와 함께 아마존은 전 세계로 확대되면서 규제 및 컴플라이언스 문제가 증가하는 상황을 겪게 될 수도 있다.

— 신규 투자, 특히 물류먼트, 배송, 아마존 웹 서비스 관련 신규 투자로 인해 잉여 현금 흐름 성립에 압박을 받을 것으로 예상된다. 또한, 물류먼트 네트워크로 인해 일부 국가에서의 아마존 진출은 미국보다 어렵게 이뤄질 수 있다.

— 고객 선호와 더 큰 규모의 소매 기업들이 더 나은 전자상거래 경험을 제공하기 때문에, 아마존이 사치품과 같은 새로운 소매 상품 카테고리에서 성공적으로 침투하지 못할 수도 있다.

주가/총가	3,327.59	주가/이익	4200.00
0.79	0.00	1,685.23	
30 7월 2021	30 7월 2021	30 7월 2021	

평가	★★★★	적정평가
분배성	높음	매우 높음
재무 건전성	—	강력

ESG 리스크 등급 평가	ESG 리스크 등급 평가
07 7월 2021	05:00, UTC

중요 공시: 모닝스타 애널리스트의 행동에는 윤리 정책/행동 강령, 개인 증권 매매거래 정책(또는 그에 상응하는 정책) 및 투자 리서치 정책이 적용됩니다. 이해상충에 대한 자세한 정보는 <https://global.morningstar.com/equitydisclosures>에서 확인하실 수 있습니다. ESG 리스크 등급 평가(ESG Risk Rating Assessment)는 서스테인리틱스(Sustainalytics)의 ESG 리스크 등급(ESG Risk Rating)을 나타냅니다.

아마존, MGM 스튜디오 인수해 비디오 라이브러리 확장... 추정 적정가치는 그대로 4,200불

투자 의견

Dan Romanoff, 주석 애널리스트, 29 4월 2021

아마존(Amazon)은 활동 중인 시장에서 지배적인 입지를 보유하고 있으며, 특히 전자상거래와 클라우드 서비스에서 이러한 경향이 두드러진다. 아마존은 수많은 경쟁 우위로 혜택을 누리며 있으며 사이즈와 스케일 측면에서 확연하게 전자상거래 리더 기업으로 부상했다. 이러한 측면을 통해 아마존은 타의 추종을 불허하는 소비자용 저가 상품을 제공한다. 전자상거래로의 장기적인 전환은 계속되고 있으며 아마존은 그 규모에도 불구하고 계속해서 시장 점유율을 확대하고 있다. 프라임(Prime) 제도는 아마존의 전자상거래 관련 노력을 하나로 묶는 역할을 하며, 아마존 소유의 업체에서 자주 구매하는 고객들을 통해 높은 마진의 반복 발생하는 매출액이 꾸준히 창출될 수 있도록 한다. 대신 고객들은 수백만 가지의 제품을 1일 배송으로 받아볼 수 있으며, 독점적으로 제공되는 영상 콘텐츠 및 기타 서비스를 누릴 수 있다. 그 결과 고객과 판매자가 서로를 끌어들이는 강력한 선순환 구조가 형성된다. 킨들(Kindle)과 다른 기기들은 새로운 고객을 유치하는 한편 기존 고객을 유지하는 측면에서 거절하기 힘든 수준의 가치제안을 형성해 이러한 생태계를 한층 더 강화한다. 또한, 아마존은 아마존 웹 서비스(Amazon Web Services, AWS)로 퍼블릭 클라우드 서비스 분야를 선도하는 기업이기도 하다.

다들여 아마존의 광고 사업은 이미 큰 규모로 확보하고 있으며 지속적으로 확대되고 있다. 이에 따라 대규모의 청중에게 달기를 원하는 마케터를 대상으로 그 특정한 소비자들과 관련된 독점적인 데이터포인트와 함께 매력적인 선택지를 제공한다. 아마존 웹 서비스, 광고, 구독료는 2020년 이전적인 수준의 온라인 쇼핑 증가와 함께 전자상거래보다 빠른 속도로 성장하고 있다. 당사는 이 세 분야가 향후 수년간 주로 성장을 견인할 것으로 예상된다. 이는 이 세 분야가 아마존 기업 평균보다 더 높은 마진율을 보여주기 때문에 매우 중요하며, 이에 따라 마진이 지속적으로 확대되면서 영업이익과

EPS가 매출액 증가 속도를 추월할 수 있을 것으로 예상된다.

당사는 소매 측면에서는 지속적인 혁신이 시장 점유율 확대에 도움을 줄 것으로 예상된다. 또한, 당사는 식료품 및 사치품과 같은 카테고리로의 지속적인 진출이 이뤄질 것으로 기대한다. 기존에는 이러한 카테고리에서는 다른 소매 부문에서 거둔 수준의 성과를 이루지 못했었다. 또한, 당사는 아마존 웹 서비스의 기술적 발전과 엔터프라이즈급 고객 서비스 제공 추진이 해당 분야에서 아마존 웹 서비스가 확보한 선도적인 지위를 유지하는 데 도움을 줄 것으로 본다. 종합적으로, 당사는 향후 수년간 강력한 매출액 및 잉여 현금 흐름 성장이 이뤄질 것으로 기대한다.

애널리스트 노트

Dan Romanoff, 주석 애널리스트, 28 5월 2021

5월 26일, 넓은 경제적 해자를 보유한 아마존(Amazon)이 비상장 기업 MGM 스튜디오(MGM Studios)를 84억 5천만불에 인수한다고 공식 발표했다. 며칠간 해당 계약에 관한 소문이 돌았다. 시기 언급은 없었지만 규제 관련 지연은 없을 것으로 보인다. 이 계약은 1조 6천억불 규모의 아마존 시가 총액에 비하면 그리 중요하지 않고 당사 추정 적정가치 4,200불도 조정하지 않을 것이나, 아마존 프라임 비디오(Amazon Prime Video) 라이브러리에 콘텐츠가 더 많이 추가되므로 이는 전략상 중요한 조치라고 본다. MGM이 보유한 콘텐츠는 영화, TV(대본이 있는 프로그램과 대본이 없는 프로그램 모두), 비디오 게임, 팟캐스트 등이다. 특히 MGM은 제임스 본드(James Bond), 록키(Rocky), 핑크팬더(Pink Panther), 시녀이야기(The Handmaid's Tale), 리얼 하우스 와이브스(Real Housewives) 등 다양한 프랜차이즈와 영화 4천여 편, TV 에피소드 1만 7천여 편을 소유하고 있다. 당사는 아마존이 유의미한 스트리밍 회사 중 하나라는 지위를 유지하려면 콘텐츠 라이브러리를 계속 확장해야 한다고 생각한다. 아마존이 MGM 인수를 근거로 내년 프라임 멤버십 가격 인상을 정당화하는 전략을 그려볼 수

있다. 아마존은 단기적으로는 COVID-19 팬데믹 기간(식료품과 필수재 부문이 특히 강세를 보임)에 전자상거래로의 전환이라는 추세로, 장기적으로도 장기적인 전자상거래 전환 추세로 변할 수 있는 좋은 입지에 있다고 생각한다. 또한, Prime 회원에게는 콘텐츠가 차별화 요소라고 본다. 프라임 회원에게는 일부 콘텐츠가 무료 제공될 것으로 예상된다. 아마존은 아마존닷컴에서 더 자주 쇼핑하고, 더 많은 소매 카테고리에서 구매하고, 전반적으로 아마존에서 지출을 더 많이 하는 이 중요한 소비자 그룹의 혜택을 계속 확대해나가고 있기 때문이다. 아마존의 구독 서비스는 2020년에 252억불 매출액을 올렸고, 이는 총매출의 7%를 차지하며 2019년 대비 31% 증가한 수치다. 대부분이 프라임 멤버십이겠지만, 킨들 언리미티드(Kindle Unlimited), 프라임 비디오, 아마존 뮤직 언리미티드(Amazon Music Unlimited)를 포함한 다른 유료 구독도 믹스 내에서 빠르게 성장하고 있다.

아마존 종목 예시

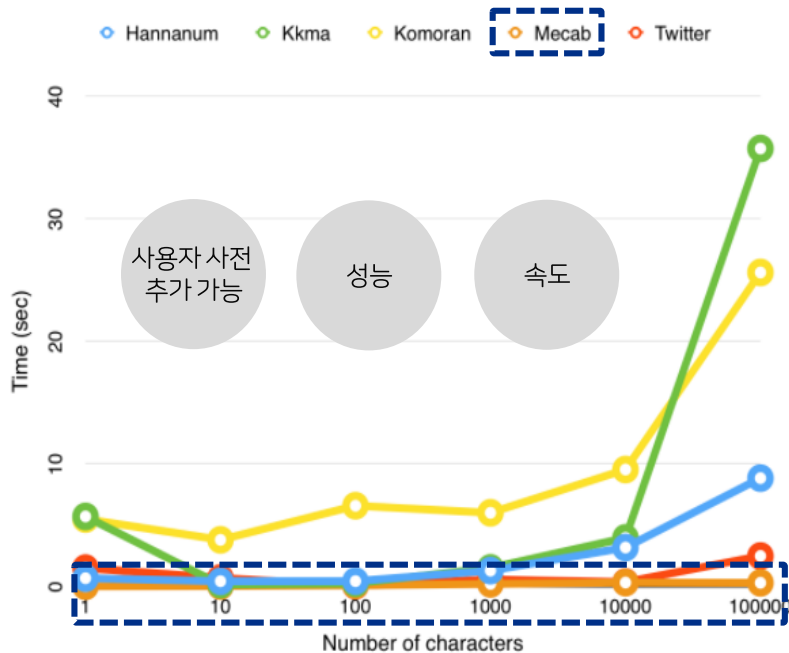
3. 분석(개발) 방법론

✓ 텍스트 데이터를 학습시키기 위해서는 “토큰화” 과정이 선행됨

* 토큰화 : 말뭉치(corpus)에서 토큰(token)이라 불리는 단위로 나누는 작업을 의미, 토큰의 단위는 보통 의미를 가지는 단위로 선정

✓ 텍스트 데이터는 Mecab 형태소 분석기를 사용하였으며, 명사만 추출

✓ "Mecab"의 경우 성능 및 속도의 우수성, 사용자사전 추가 가능 등의 장점으로 최종 선택하게 됨



결과

아마존(Amazon)은 활동 중인 시장에서 지배적인 입지를 보유하고 있으며, 특히 전자상거래와 클라우드 서비스에서 이러한 경향이 두드러진다.

아마존, '(', Amazon, ')', 은, 활동, 중, 인, 시장, 에서, 지배, 적, 인, 입지, 를, 보유, 하, 고, 있, 으며, ', ', 특히, 전자, 상거래, 와, 클라우드, 서비스, 에서, 이러, 한, 경향, 이, 두드러진다, ' . '

아마존, 활동 중, 시장, 지배, 입지, 보유, 전자, 상거래, 클라우드, 서비스, 경향

한 단어 → 제거

불용어 → 제거

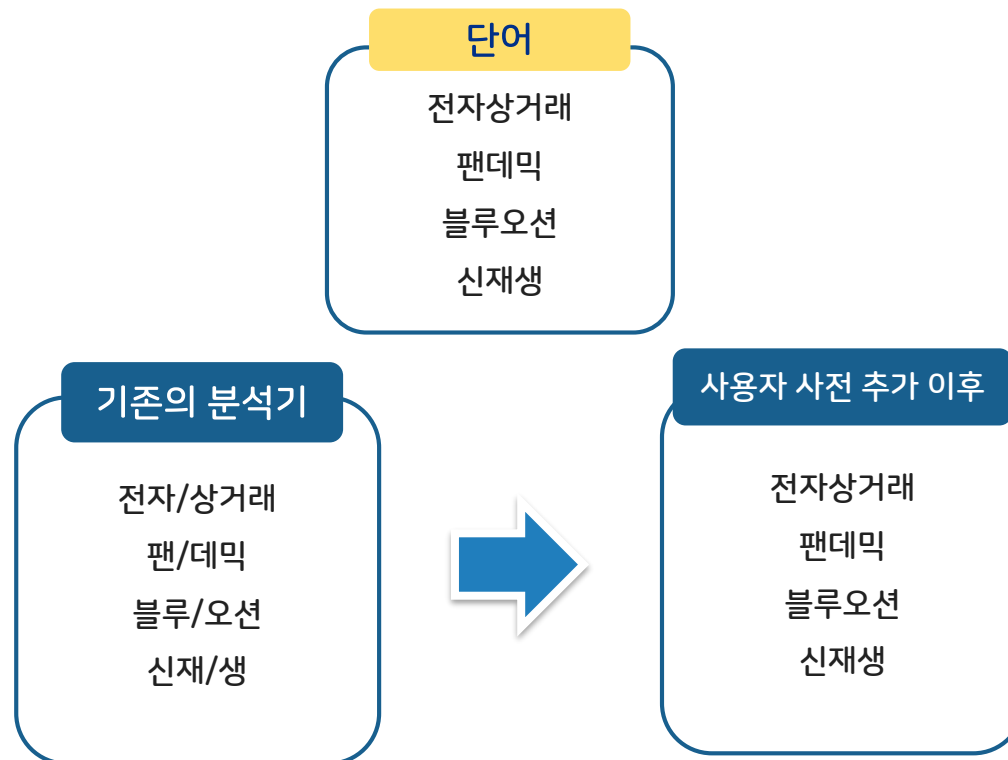
- 1 사용자 사전 추가
- 2 단어 합치기

특수문자
숫자
엔터
다중 공백
“제거” 후, 명사 추출

3. 분석(개발) 방법론

1 사용자 사전 추가 : 원하는 단어에 대해서 사용자가 직접 단어를 사전에 추가 가능

- ✓ 한국어의 형태소는 띄어쓰기가 조금만 지켜지지 않아도 기계가 인식할 때 굉장히 난항을 겪음
- ✓ 아무리 우수한 형태소 분석기라도 100%의 정확도를 보여주기 어려움
- ✓ 형태소 분석이 잘못된 단어들이거나 신조어, 외래어와 같은 단어들의 예외처리를 할 때 유용



3. 분석(개발) 방법론

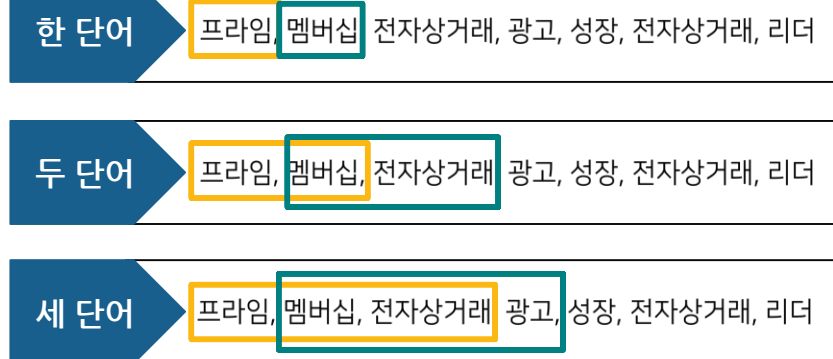
2 단어 합치기 : 한 단어(unigram), 두 단어(bigram), 세 단어(trigram)로 보기

- ✓ 전자상거래/ 전자상거래 전환/ 전자상거래 전환 추세 → 키워드 상세 효과
- ✓ 팬,데믹/팬데믹 → 사용자 사전 추가하여 단어 새로 정의한 것과 같은 효과

* '전자상거래' 단어 사용자 사전에 추가하여 한 단어로 추출됨

한 단어 <unigram>	두 단어 <bigram>	세 단어 <trigram>
전자상거래 소매 프라임 비디오 멤버십 콘텐츠 성장 확대	전자상거래 전환 전환 추세 전자상거래 리더 소매 전자상거래 프라임 멤버십 비디오 라이브러리 라이브러리 확장 점유 확대 흐름 성장	전자상거래 전환 추세 현금 흐름 성장

명사 단어 추출



3. 분석(개발) 방법론

카운트 기반의 단어 표현

✓ Counter

: 해당 문서의 단어 빈도수를 세는 방법

✓ TF-IDF

: 해당 문서와 타 문서를 비교하여, 문서 내에서 등장하는 각 단어의 "중요도" 를 계산하는 방법론

문서의 핵심어 추출, 검색 엔진에서의 검색 결과의 순위를 결정, 문서들 사이의 비슷한 정도를 구하는 등의 용도로 사용



"다른" 관점

* TF-IDF

TF



[해당 문서 내에서 특정단어 i의 등장 횟수] ↑

×

IDF



$\log\left(\frac{\text{리포트의 총 개수}}{\text{특정 단어 i가 등장한 리포트 수}}\right)$ ↓

단어 i의 중요도 =

1) 해당 문서에서 나타난 횟수가 많을수록 (TF)

2) 다른 문서에서 나타난 횟수가 적을수록 (IDF)

"해당 문서를 대표하는 키워드"

4. 분석(개발) 결과

Amazon.com Inc AMZN (XNAS)

모닝스타 등급	★★★★	주가/총가	3,327.59 USD	주당 이익가치	4200.00 USD
30 7월 2021	30 7월 2021	30 7월 2021	30 7월 2021	30 7월 2021	30 7월 2021
21:25, UTC	02:00, UTC				

모닝스타 핵심 지표	매출액	이익
경제적 성과	매출액	이익
가치평가	★★★★	적정평가
분할실적	매출	이익
재무 건전성	—	강력

출처: 모닝스타 주식 분석

정량적 가치평가

모닝스타	AMZN
가치평가	적정평가
가치평가	가치평가

주가/총가	3,327.59	주가/이익	0.99	주가/매출	0.80	주가/자산	0.83
PER	57.9	PER	185.2	PER	16.2	PER	20.1
매출 PER	55.8	매출 PER	—	매출 PER	12.3	매출 PER	13.9
주가/현금흐름	28.8	주가/현금흐름	30.8	주가/현금흐름	10.2	주가/현금흐름	13.1
주가/영업이익	241.8	주가/영업이익	63.4	주가/영업이익	17.7	주가/영업이익	19.5
매출 수익률 %	—	매출 수익률 %	—	매출 수익률 %	2.48	매출 수익률 %	2.35

출처: Morningstar

상승론자의 견해

아마존은 본연 전자상거래 분야를 이끄는 리더 기업으로, 타의 추종을 불허하는 규모를 누리며 성장 기회와 최상의 고객 경험 향상을 위한 투자를 지속하고 있다.

마진율이 높은 광고 및 아마존 웹 서비스 사업 부문은 아마존 전체 평균보다 더 빠른 속도로 성장하고 있으며, 이는 향후 수년간 지속적으로 수익성을 확대하는 역할을 할 것으로 예상된다.

아마존 프라임(Amazon Prime) 멤버십은 아마존에서 많은 소비를 하는 고객들을 끌어들이고 유지하는 데 도움을 주며, 반복적이고 높은 마진의 매출액을 창출하는 한편 강력한 네트워크 효과를 강화하는 작용을 한다.

하락론자의 견해

아마존을 포함한 대형 기술 기업 대상으로 한 규제 관련 우려가 부상하고 있다. 이와 함께 아마존은 전 세계로 확대되면서 규제 및 컴플라이언스 문제가 증가하는 상황을 겪게 될 수도 있다.

신규 투자, 특히 플랫폼, 배송, 아마존 웹 서비스 관련 신규 투자로 인해 영업 현금 흐름 성장이 압박을 받을 것으로 예상된다. 또한, 물가 상승과 네트워크에 의해 일부 국가에서의 아마존 진출은 미국보다 어렵게 이뤄질 수 있다.

고객 선호와 더 큰 규모의 소매 기업들이 더 나은 전자상거래 경험을 제공하기 때문에, 아마존이 사치품과 같은 새로운 소매 상품 카테고리에서 성공적으로 침투하지 못할 수도 있다.

주가/이익가치	0.79	시가총액(십억)	1,685.23
30 7월 2021	30 7월 2021	30 7월 2021	30 7월 2021

인더넷 소매

가치평가	모범	ESG 리스크 등급 평가	07 7월 2021
05:00, UTC			

중요 공시: 모닝스타 애널리스트의 행동에는 윤리 정책/행동 강령, 개인 증권 매매거래 정책(또는 그에 상응하는 정책) 및 투자 리서치 정책이 적용됩니다. 이해상충에 대한 자세한 정보는 <http://global.morningstar.com/equity/disclosures>에서 확인하실 수 있습니다. ESG 리스크 등급 평가(ESG Risk Rating Assessment)는 서스테인리틱스(Sustainalytics)의 ESG 리스크 등급(ESG Risk Rating)을 나타냅니다.

아마존, MGM 스튜디오 인수해 비디오 라이브러리 확장... 추정 적정가치는 그대로 4,200불

투자 의견

Dan Romanoff, 주식 애널리스트, 29 4월 2021

아마존(Amazon)은 활동 중인 시장에서 지배적인 입지를 보유하고 있으며, 특히 전자상거래와 클라우드 서비스에서 이러한 경향이 두드러진다. 아마존은 수많은 경쟁 우위로 혜택을 누리며 있으며 사이즈와 스케일 측면에서 확연하게 전자상거래 리더 기업으로 부상했다. 이러한 측면을 통해 아마존은 타의 추종을 불허하는 소비자용 저가 상품을 제공한다. 전자상거래로의 장기적인 전환은 계속되고 있으며 아마존은 그 규모에도 불구하고 계속해서 시장 점유율을 확대하고 있다. 프라임(Prime) 제도는 아마존의 전자상거래 관련 노력을 하나로 묶는 역할을 하며, 아마존 소유의 업체에서 자주 구매하는 고객들을 통해 높은 마진의 반복 발생하는 매출액이 꾸준히 창출될 수 있도록 한다. 대신 고객들은 수백만 가지의 제품을 1일 배송으로 받아볼 수 있으며, 독점적으로 제공되는 영상 콘텐츠 및 기타 서비스를 누릴 수 있다. 그 결과 고객과 판매자가 서로를 끌어들이는 강력한 선순환 구조가 형성된다. 킨들(Kindle)과 다른 기기들은 새로운 고객을 유치하는 한편 기존 고객을 유지하는 측면에서 거절하기 힘든 수준의 가치제안을 형성해 이러한 생태계를 한층 더 강화한다. 또한, 아마존은 아마존 웹 서비스(Amazon Web Services, AWS)로 퍼블릭 클라우드 서비스 분야를 선도하는 기업이기도 하다.

더불어 아마존의 광고 사업은 이미 큰 규모로 확보하고 있으며 지속적으로 확대되고 있다. 이에 따라 대규모의 청중에게 달기를 원하는 마케터를 대상으로 그 특정한 소비자들과 관련된 독점적인 데이터포인트와 함께 매력적인 선택지를 제공한다. 아마존 웹 서비스, 광고, 구독료는 2020년 이전적인 수준의 온라인 쇼핑 증가와 함께 전자상거래보다 빠른 속도로 성장하고 있다. 당사는 이 세 분야가 향후 수년간 주로 성장을 견인할 것으로 예상된다. 이는 이 세 분야가 아마존 기업 평균보다 더 높은 마진율을 보여주기 때문에 매우 중요하며, 이에 따라 마진이 지속적으로 확대되면서 영업이익과

EPS가 매출액 증가 속도를 추월할 수 있을 것으로 예상된다.

당사는 소매 측면에서는 지속적인 혁신이 시장 점유율 확대에 도움을 줄 것으로 예상된다. 또한, 당사는 식료품 및 사치품과 같은 카테고리로의 지속적인 진출이 이뤄질 것으로 기대한다. 기존에는 이러한 카테고리에서는 다른 소매 부문에서 거둔 수준의 성과를 이루지 못했다. 또한, 당사는 아마존 웹 서비스의 기술적 발전과 엔터프라이즈급 고객 서비스 제공 추진이 해당 분야에서 아마존 웹 서비스가 확보한 선도적인 지위를 유지하는 데 도움을 줄 것으로 본다. 종합적으로, 당사는 향후 수년간 강력한 매출액 및 영업 현금 흐름 성장이 이뤄질 것으로 기대한다.

애널리스트 노트

Dan Romanoff, 주식 애널리스트, 28 5월 2021

5월 26일, 넓은 경제적 해자를 보유한 아마존(Amazon)이 비상장 기업 MGM 스튜디오(MGM Studios)를 84억 5천만 달러에 인수한다고 공식 발표했다. 며칠간 해당 결정에 관한 소문이 돌았다. 시기 언급을 받았지만 규제 관련 지엽은 없을 것으로 보인다. 이 계약은 1조 6천억 달러 규모의 아마존 시가 총액에 비하면 그리 중요하지 않고 당사 추정 적정가치 4,200불도 조정하지 않을 것이나, 아마존 프라임 비디오(Amazon Prime Video) 라이브러리에 콘텐츠가 더 많이 추가되므로 이는 전략상 중요한 조치라고 본다. MGM이 보유한 콘텐츠는 영화, TV(대본이 있는 프로그램과 대본이 없는 프로그램 모두), 비디오 게임, 팟캐스트 등이다. 특히 MGM은 제임스 본드(James Bond), 록키(Rocky), 핑크팬더(Pink Panther), 시녀이야기(The Handmaid's Tale), 리얼 하우스 와이브스(Real Housewives) 등 다양한 프랜차이즈와 영화 4천여 편, TV 에피소드 1만 7천여 편을 소유하고 있다. 당사는 아마존이 유의미한 스트리밍 회사 중 하나라는 지위를 유지하려면 라이브러리를 계속 확장해야 한다고 생각한다. 아마존이 MGM 인수를 근거로 내년에 프라임 멤버십 가격 인상을 정당화하는 전략을 그려볼 수

있다. 아마존은 단기적으로는 COVID-19 팬데믹 기간(식료품과 필수재 부문이 특히 강세를 보임)에 전자상거래로의 전환이라는 추세로, 장기적으로도 장기적인 전자상거래 전환 추세로 변할 수 있는 좋은 입지에 있다고 생각한다. 또한, Prime 회원에게는 콘텐츠가 차별화 요소라고 본다. 프라임 회원에게는 일부 콘텐츠가 무료 제공될 것으로 예상된다. 아마존은 아마존닷컴에서 더 자주 쇼핑하고, 더 많은 소매 카테고리에서 구매하고, 전반적으로 아마존에서 지출을 더 많이 하는 이 중요한 소비자 그룹의 혜택을 계속 확대해나가고 있기 때문이다. 아마존의 구독 서비스는 2020년에 252억 달러 매출을 올렸고, 이는 총매출의 7%를 차지하며 2019년 대비 31% 증가할 수 있다. 대부분이 프라임 멤버십이겠지만, 킨들 어마타이드(Kindle Unlimited), 프라임 비디오, 아마존 뮤직 언리미티드(Amazon Music Unlimited)를 포함한 다른 유료 구독도 믹스 내에서 빠르게 성장하고 있다.

아마존의 키워드는
아래와 같은 단어들이 추출될 수 있음

키워드

비디오 라이브러리

전자상거래

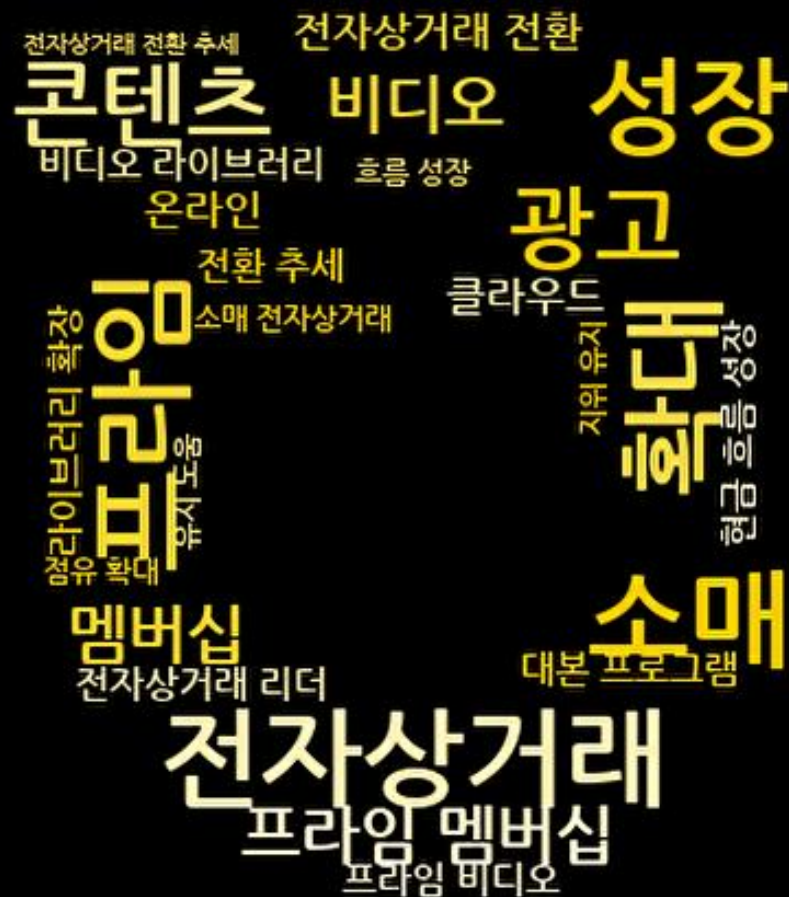
클라우드 서비스

콘텐츠 라이브러리

전자상거래 전환

프라임 멤버십

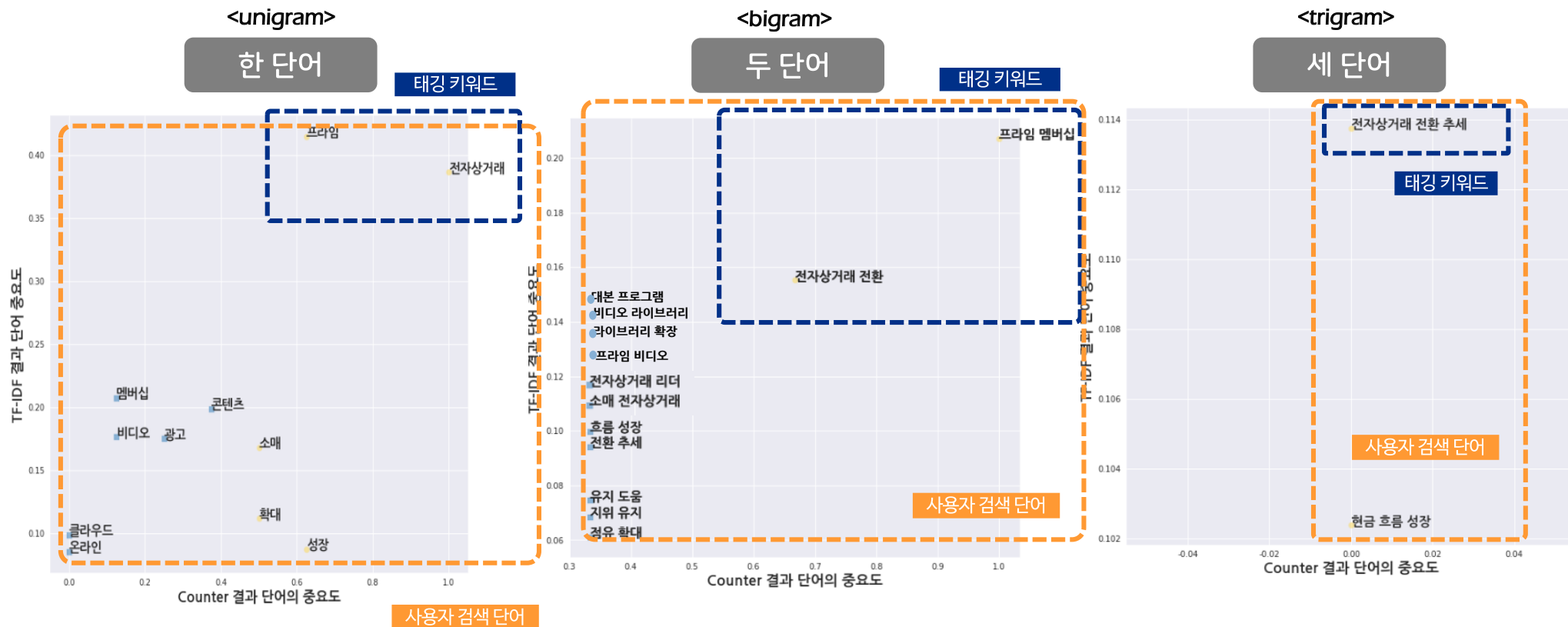
4. 분석(개발) 결과 – 키워드 추출 모델의 키워드 시각화



4. 분석(개발) 결과

- ✓ Tf-idf 값과 counter의 두가지 방법론 모두 수치가 높게 나온 단어들에 한해 태깅 키워드에 추가
- ✓ 키워드로 추출되었지만 tf-idf, count값 모두 높게 나오지 않은 단어들은 사용자 검색에 활용될 수 있도록 할 예정
- ✓ 가로축(단어 빈도수), 세로축(tf-idf) 값이 "높을수록" 좋은 키워드

가로축 : Counter 결과 단어의 중요도
세로축 : TF-IDF 결과 단어 중요도



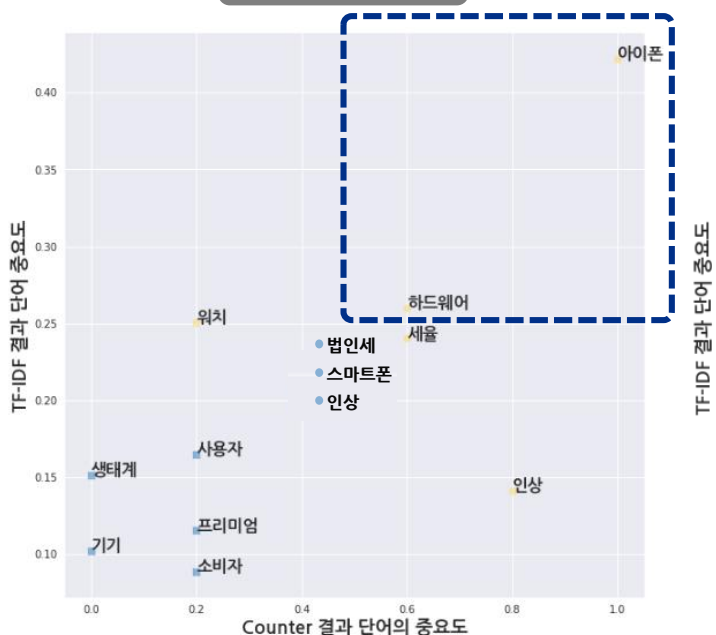
#전자상거래 #프라이م 멤버십 #전자상거래 전환 추세

4. 분석(개발) 결과 – 애플 예시

- ✓ Tf-idf 값과 counter의 두가지 방법론 모두 수치가 높게 나온 단어들이 [아이폰, 하드웨어, 법인세 인상]이 태깅 키워드로 추출됨
- ✓ 두가지 방법론에서 모두 높은 수치를 얻지 못했지만 키워드로 추출된 [인상,하드웨어,세울,스마트폰,법인세,프리미엄,소비자,워치,기기,하드웨어 프리미엄 부여 ...] 단어들은 사용자의 검색 단어에 활용될 예정

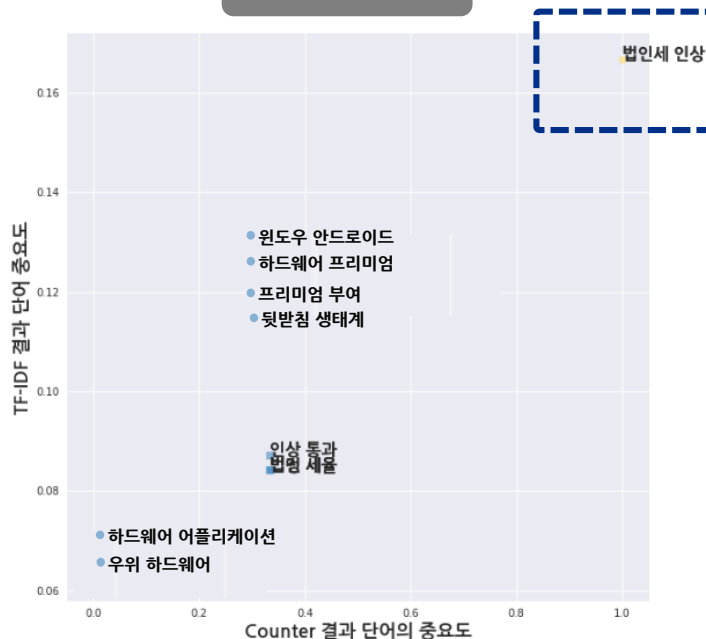
<unigram>

한 단어



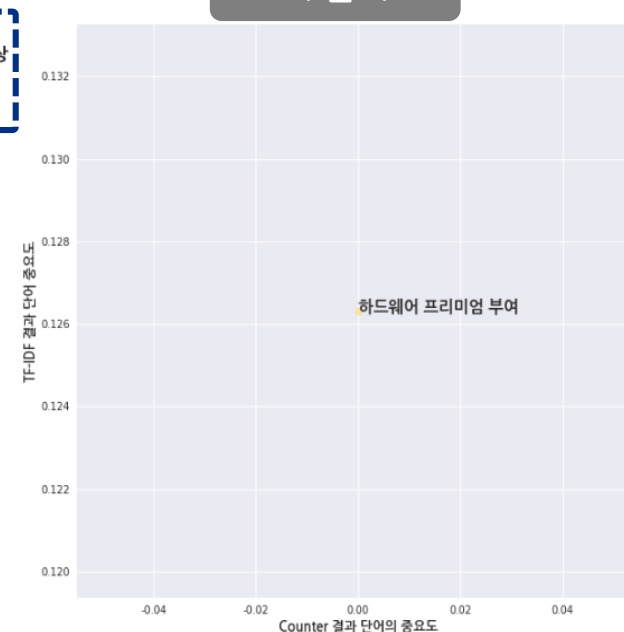
<bigram>

두 단어



<trigram>

세 단어



애플 키워드

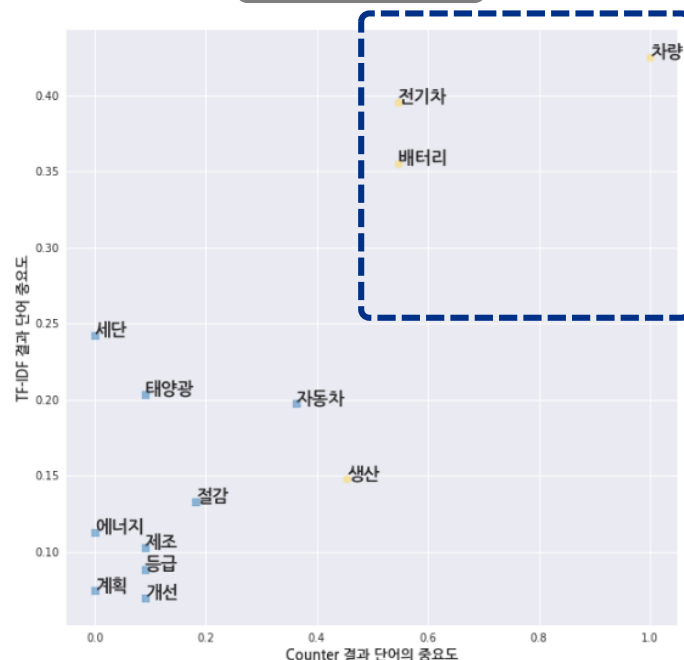
#아이폰 #하드웨어 #법인세 인상

4. 분석(개발) 결과 – 테슬라 예시

- ✓ Tf-idf 값과 counter의 두가지 방법론 모두 수치가 높게 나온 단어들이 [전기차, 배터리, 법인세 차량 ...]이 태깅 키워드로 추출됨
- ✓ 두가지 방법론에서 모두 높은 수치를 얻지 못했지만 키워드로 추출된 [태양광, 자동차, 기관 엔진 ...] 단어들은 사용자의 검색 단어에 활용될 예정

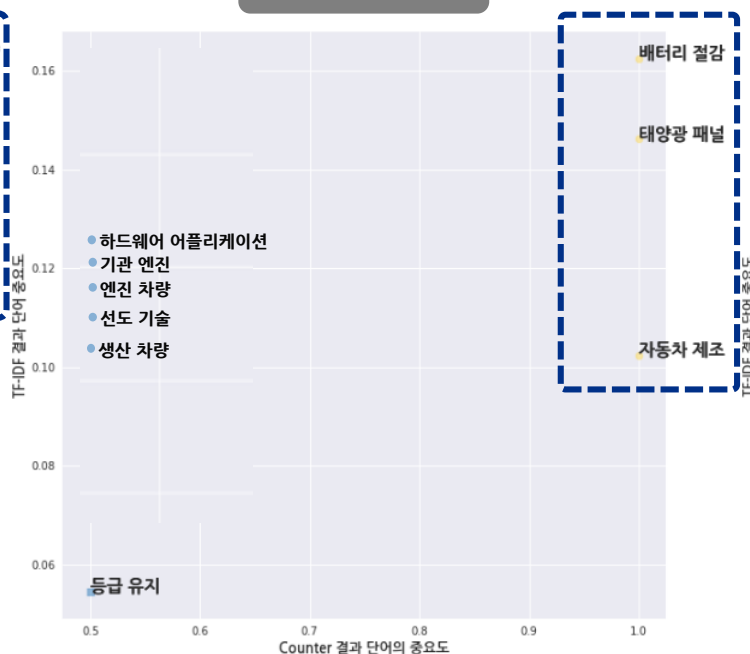
<unigram>

한 단어



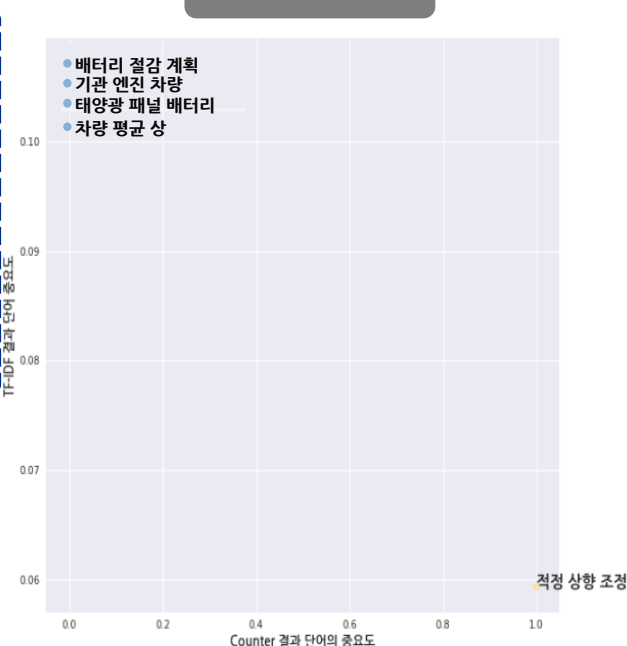
<bigram>

두 단어



<trigram>

세 단어



테슬라 키워드

#전기차 #배터리 #차량 #자동차 제조 #태양광 패널 #배터리 절감

4. 분석(개발) 결과

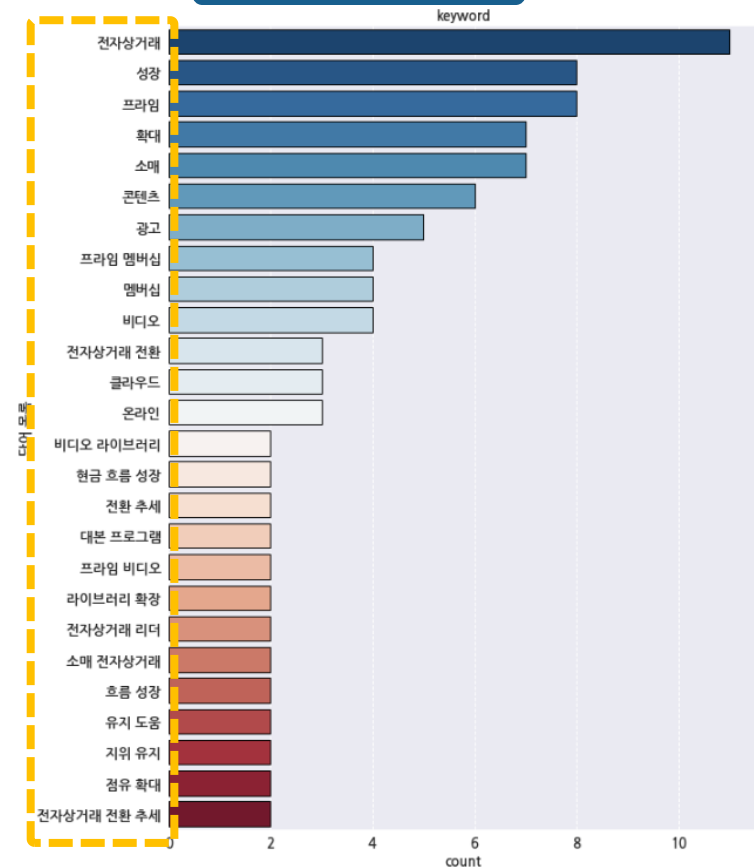
키워드 추출 라이브러리(soykeyword : 두 집합의 단어 출현 확률의 비율을 바탕으로 키워드 추출)와 비교

- ✓ 모델결과를 타 모델과 비교해보기위해 soykeyword와 키워드 결과를 비교해 보았는데, counter+tf-idf 모델 결과가 단어의 양상이 더 자세하게 나옴을 확인

soykeyword

```
[KeywordScore(word='전자상거래', frequency=180, score=0.9821317941575944),  
KeywordScore(word='콘텐츠', frequency=219, score=0.9544889604176743),  
KeywordScore(word='배송', frequency=107, score=0.934130488374377),  
KeywordScore(word='역할', frequency=108, score=0.9335448489145806),  
KeywordScore(word='구독', frequency=109, score=0.9329599433109073),  
KeywordScore(word='광고', frequency=273, score=0.9328430501315375),  
KeywordScore(word='형성', frequency=110, score=0.9323757701848443),  
KeywordScore(word='진출', frequency=120, score=0.9265740243732729),  
KeywordScore(word='독점', frequency=121, score=0.9259978184982606),  
KeywordScore(word='리더', frequency=132, score=0.9197065312195372),  
KeywordScore(word='반복', frequency=150, score=0.9095940671360606),  
KeywordScore(word='잉여', frequency=161, score=0.9035229711368668),  
KeywordScore(word='소매', frequency=592, score=0.8890807282523806),  
KeywordScore(word='경험', frequency=189, score=0.8884288619862106),  
KeywordScore(word='물류', frequency=102, score=0.8805483448012665)]
```

Counter + Tf-idf



5. 결론 및 활용

- ✓ (as-is)현재 양질의 국내외 리서치 주식 비정형 데이터가 적재 및 활용이 안되고 있음
- ✓ 모닝스타 해외 주식 데이터를 활용하여 데이터 분석 후 Keyword 에 따른 해외주식 검색, 비정형 주식 리포트 데이터 화 후 가독성 있는 투자의견 제시를 목표로 하고 있음
- ✓ 향후 리서치본부와의 협업을 통해 국내 리서치 자료를 데이터화 하여 분석 모델 개발 예정
- ✓ 또한 펀드와 ETF 에도 분석 데이터를 이용하여 Keyword 검색이 가능한 모델 개발 예정

모닝스타 데이터를 이용한 HTS 화면 예시

MTS 화면 예시

5. 결론 및 활용 - 한계점 및 개선방향

1) 불용어 사전 추가

- 종목을 설명하지 못하는 단어(확대, 하향, 시장, 단계 등) 제거하지 않은 상태인데, 문맥을 보고 제거하는 방향
- 한 단어, 두 단어, 세 단어를 볼 때, 불용어 사전을 각각 다르게 하여 각 단어들에 맞는 키워드 추출
- 번역 오류로 기업명이 제대로 번역이 안된 경우가 있어, 불용어가 잘 제거가 안된 종목들이 있음
추후 보완하여 기업명 올바르게 제거

2) 사용자 사전 추가

3) Mecab 형태소 분석기에서 한글만(명사) 추출하였는데, 영어 단어도 키워드로 추출되게 하기

- * 현재 GPU, esg 같은 주요 영어 키워드가 추출되지 않음

4) Tf-idf, count 방법 이외에 단어의 의미, 문맥을 고려하는 다른 방법론 적용

- * 단어의 빈도수 기반 알고리즘으로, 단어의 의미를 전혀 고려하지 못한다는 단점 존재