

BDBD - Best Day of BigData



▶ 프로젝트 개요

▲ 활용 데이터

🖺 이해관계자 리스트

② 예측 모델

ACCOUNT PAGES

Profile

Hanyang Woman's Report

Capstone Project Report

사회 이슈 발생 시 이해관계자 개입에 따른 소비량 예측 및 시각화



프로젝트 배경 및 필요성 / 기대효과

Home

🕦 프로젝트 개요

세 활용 데이터

🖺 이해관계자 리스트

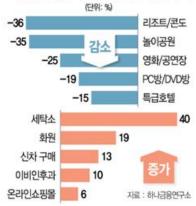
이 예측 모델

ACCOUNT PAGES

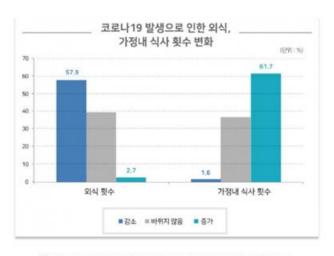
Profile

현대 사회는 빠르게 변화하고 있으며, 이로 인해 다양한 사회 이슈가 지속적으로 발생하고 있다. 이러한 사건과 사고는 소비자의 구매 패턴과 소비 양상에 큰 영향을 미치며, 기업과 자영업자들의 매출에도 직접적인 영향을 끼친다. 예를 들어, 특정 사건이 발생했을 때 언론에서 이를 보도하는 방식과 그로 인한 여론의 형성은 소비자의 행동에 큰 변화를 유도하며, 외식업계나 특정 산업의 매출에 즉각적인 영향을 미칠 수 있다. 이태원 참사와 같은 사고 이후 외식업계의 예약 취소가 급증한 사례는 이러한 현상을 잘 보여준다.





서은영, "미세먼지 뉴스 많은 날...세탁소 웃고 마트 울고", 2019.04.17



이은용, "코로나19 여파 국민들 '식품소비행태' 변화 가져와", 2020.12.21

하지만 우리는 언론의 보도 방향을 통제할 수 없으며, 여론의 형성 역시 외부적인 변수로 작용하여 기업과 자영업자들에게 예측하기 어려운 매출 변동을 초래한다. 이러한 상황에서 기업과 자영업자들이 언론과 여론의 영향을 예측하고 대비할 수 있는 해결책을 마련하는 것이 중요하며 소비 패턴의 변화를 정확히 분석하고 예측함으로써, 매출 변동에 효과적으로 대응할 수 있도록 지원하는 것은 필수적이다. 특히 이러한 매출 변동 예측은 자영업자들의 생존과 직결된 중요한 문제이며, 이에 대한 해결 방안을 제시하는 것이 본 프로젝트의 핵심 목표이다.



프로젝트 배경 및 필요성 / 기대효과

- # Home
- 🕦 프로젝트 개요
- 세 활용 데이터
- 이해관계자 리스트
- 이 예측 모델

ACCOUNT PAGES

Profile

▋ 사회 이슈에 대한 소비 패턴 예측

개발 모델은 사회 이슈가 발생함에 따라, 이해관계자들의 개입이 소비 패턴에 영향을 미쳐 소비량 변화를 예측한다. 이를 통해 기업, 기관, 소상공인 등은 사건/사고에 신속하게 대응이 가능하고, 소비전략을 조정할 수 있는 기반을 얻을 수 있다.

▋ 데이터 기반의 인사이트 제공

뉴스 데이터와 소비 패턴 간의 연계성을 분석하는데 사용되는 데이터 기반의 인사이트가 생성될 것이다. 이를 통해 기업, 기관, 소상공인들은 합리적인 의사결정을 내릴 수 있으며, 다양한 측면에서 데이터 주도의 전략을 수립할 수 있다.

경제 안정성 향상

사회 이슈에 따른 소비 패턴 변화를 더 정확하게 예측할 수 있다면 전반적인 경제 안정성을 향상시킬 수 있다. 더불어 예측 가능한 소비흐름은 경제의 예측 가능성을 향상시켜 투자와 성장을 촉진할 것이다.



Home

■ 프로젝트 개요

🚻 활용 데이터

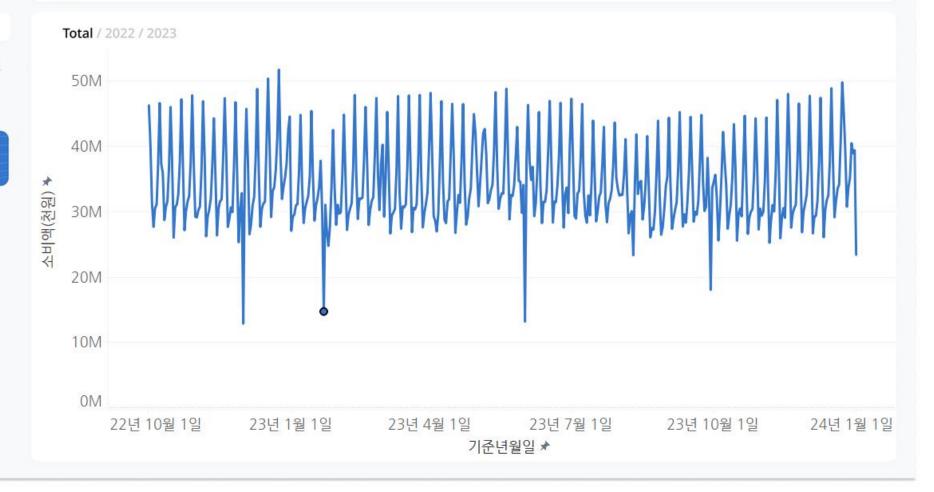
☑ 이해관계자 리스트

② 예측 모델

ACCOUNT PAGES

Profile





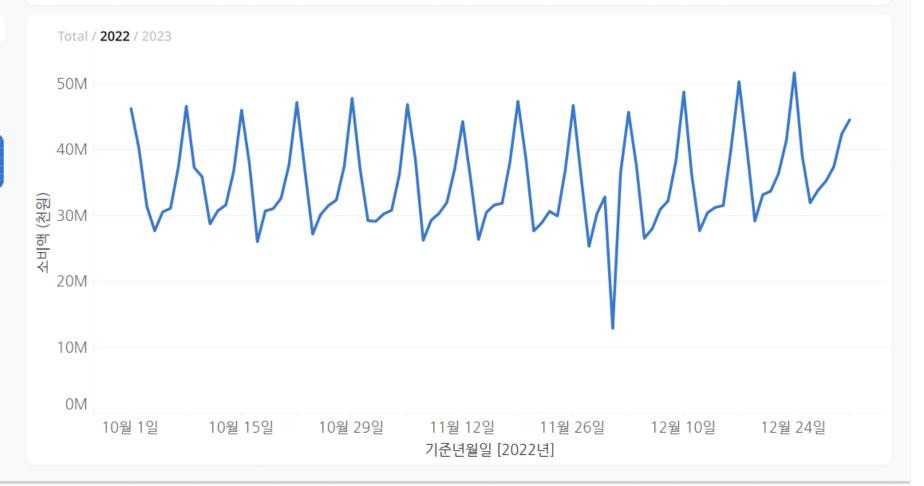


- # Home
- 프로젝트 개요
- 🚻 활용 데이터
- 🖺 이해관계자 리스트
- ② 예측 모델

ACCOUNT PAGES

2 Profile





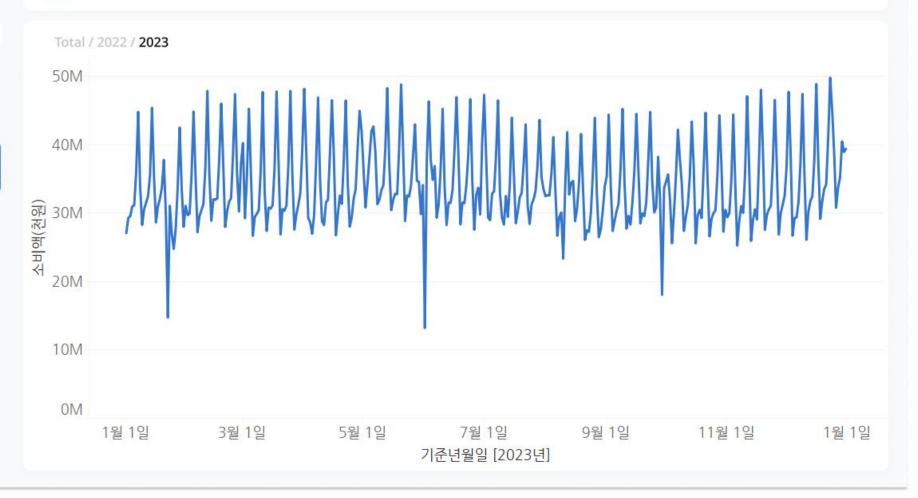


- # Home
- 프로젝트 개요
- 🚻 활용 데이터
- 이해관계자 리스트
- ② 예측 모델

ACCOUNT PAGES

Profile







이해관계자의 역할 / 이해관계자 리스트

- # Home
- 프로젝트 개요
- 세 활용 데이터
- 📴 이해관계자 리스트
- ② 예측 모델

ACCOUNT PAGES

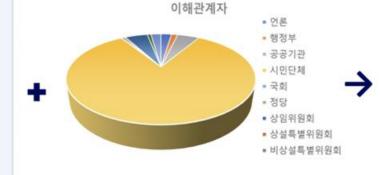
Profile

🚨 이해관계자의 역할

기존 사회 문제를 이해하기 위한 단계로 사회 이슈에 관련된 다양한 이해관계자 참여율을 파악하여 분석하고, 해당 이슈 소비량 변화에 얼마나 관여하고 있는지 판단합니다.

기존 연구

- 사회 이슈 파악
- 데이터 수집 & 가설 설정
- 목적에 맞는 모델 생성 & 학습
- 분석 후 가설 채택



진행할 연구

- 사회 이슈 파악
- 데이터 수집
- 이해관계자 추출 및 참여율 파악
- 목적에 맞는 모델 생성 & 학습
- 훈련 사례 구축 후 예측

@ Capstone Design Project Team 8D8D



이해관계자의 역할 / 이해관계자 리스트

- # Home
- 프로젝트 개요
- ₩ 활용 데이터
- 📵 이해관계자 리스트
- ② 예측 모델

ACCOUNT PAGES

Profile

🚨 이해관계자의 리스트

뉴스 데이터 수집

- 빅카인즈를 활용한 뉴스 데이터 수집 이해관계자 리스트 구축

- 이해관계자 추정 키워드 추출 - 이해관계자 분류 설정 - 리스트 카테고리화 이해관계자 참여율 구축

- 뉴스 기사에 대한 이해관계자의 일자별 빈도를 계산

@ Capstone Design Project Team 8D8D



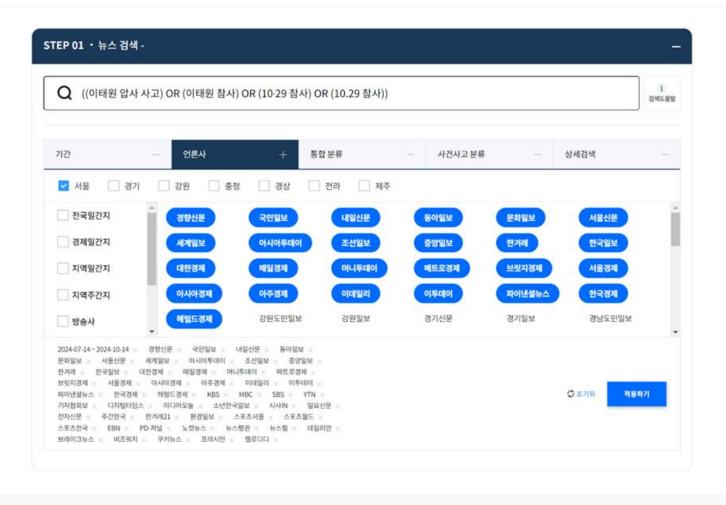
< 뉴스 데이터 수집

- # Home
- ▶ 프로젝트 개요
- 세 활용 데이터
- 📵 이해관계자 리스트
- 이 예측 모델

ACCOUNT PAGES

Profile

빅카인즈 데이터 수집 / 수집 데이터



@ Capstone Design Project Team 8D8D



< 뉴스 데이터 수집

- # Home
- 프로젝트 개요
- ₩ 활용 데이터
- 📴 이해관계자 리스트
- ② 예측 모델

ACCOUNT PAGES

Profile

빅카인즈 데이터 수집 / **수집 데이터**

뉴스 식별자	일자	언론사	기고자	제목	통합 분류1	통합 분류2	통합 분류3	사건/사고 분류1	사건/사고 분류2
01100101	20221	경향신문	박하얀	이태원 '	사회>사회	NULL	NULL	사고>스포츠	NULL
01100101	20221	경향신문	이성희	"이태원	사회>사회	NULL	NULL	사고>스포츠	NULL
01100101	20221	경향신문	윤기은	소방당	지역>강원	지역>경남	지역>광주	사고>스포츠	NULL
01100101	20221	경향신문	유경선	'이태원	사회>사회	NULL	NULL	사고>스포츠	NULL
01100101	20221	경향신문	유경선	용산소	사회>사회	NULL	NULL	사고>산업사	NULL
01100101	20221	경향신문	유경선	'이태원	사회>사건	NULL	NULL	사고>스포츠	NULL
01100101	20221	경향신문	이성희	서울시,	사회>사회	NULL	NULL	사고>교통사	NULL
01100101	20221	경향신문	유경선	'이태원	사회>사건	NULL	NULL	사고>스포츠	NULL
01100101	20221	경향신문	최인진	경기도,	지역>경기	지역>강원	NULL	사고>스포츠	NULL
01100101	20221	경향신문	김보미	'이태원	국제>아시	지역>경기	국제>중	사고>산업사	사고>산업사
01100101	20221	경향신문	조미덥	국민의	정치>국회	정치>선거	정치>청	사고>스포츠	NULL
01100101	20221	경향신문	민서영	복지부	사회>의료	NULL	NULL	사고>스포츠	NULL
01100101	20221	경향신문	민서영	사망자	국제>아시	국제>일본	사회>사	사고>스포츠	NULL
01100101	20221	경향신문	김원진	'주최'	문화>요리	지역>경남	지역>강원	사고>산업사	사고>산업사
01100101	20221	경향신문	김혜리	인도네	국제>아시	국제>중	국제>일본	사고>산업사	사회>사회갈
01100101	20221	경향신문	최인진	김동연,	지역>경기	지역>충남	지역>강원	사고>스포츠	NULL
01100101	20221	경향신문	박은경	한 총리	국제>일본	국제>아	국제>유	사회>사회갈	NULL
01100101	20221	경향신문	최인진	경기도	지역>경기	지역>강원	사회>사	사고>스포츠	NULL
01100101	20221	경향신문	고귀한	광주 전	지역>전남	지역>광주	문화>전	사고>교통사	NULL
01100101	20221	경향신문	최인진	경기도	지역>경남	지역>경기	지역>제주	사고>스포츠	NULL
01100101	20221	경향신문	김윤나	의사 출	지역>경기	지역>경남	사회>사	사고>스포츠	NULL
01100101	20221	경향신문	박준철	인천시,	사회>사회	NULL	NULL	사고>스포츠	NULL

- # Home
- 프로젝트 개요
- ₩ 활용 데이터
- 🍱 이해관계자 리스트
- ② 예측 모델

ACCOUNT PAGES

2 Profile

리스트 구축 진행 과정

키워드 열에서 이해관계자로 추정 키워드 추출 → 이해관계자 분류 설정 → Chat GPT 4.0 활용 리스트 카테고리화

이해관계자 리스트	대분류	
10.29이태원참사경 10 · 29이태원참사	정치 및	
10 · 29이태원참사	정치 및	
112상황실장	정부	
119구급대	정부	
119구조대	정부	
119대응국장	정부	
119상황실	정부	
21사단	정부	
6.25	정부	
가정법원	정부	
감찰조사팀	경제 및 정치 및	
강기갑	정치 및	
강남경찰서	정부	
강남구청	정부	
강남구청장	정부	
강남아파트	정부	
강동	정부	
강동구의회	정치 및	
강동구청	정부	
강동구청장	정부	
강릉아산병원	의료	
강북경찰서	정부	
718471771	74 H	

대분류 코	대분류 종류	중분류 코드	중분류 종류
		C101	행정
		C102	지자체
		C108	위원회 구성원
C1	정부	C107	사회복지
CI	3T	C106	사법
		C105	군대
		C103	공공기관
		C104	경찰
		C201	정치
C2	정치 및 사회 조직	C204	시민사회
CZ		C202	단체
		C203	노동 조합
		C302	언론
C3	문화 및 언론	C303	방송
		C301	문화

^{*} 이해관계자 대분류, 중분류 구분 코드

이해관계자에 해당하는 1683개의 키워드 추출 후, 대분류 코드 7종과 중분류 코드 25종으로 카테고리화

^{*} 이해관계자 리스트



< 이해관계자 리스트 구축

- # Home
- 프로젝트 개요
- ₩ 활용 데이터
- 📴 이해관계자 리스트
- ② 예측 모델

ACCOUNT PAGES

2 Profile

리스트 구축 진행 과정

Id	일자	언론사	이해관계자
1	2023	브레이	대통령, 여당, 이재명, 행정, 정부, 조직, 정치, 시민들, 민생 국민의힘, 국회, 국회의장실, 대통령, 부대 대통령, 동아일보, 보건복지부, 정신건강혁신위원회, 정부, 경제, 병원, 의료, 민간, 외 정부, TV
2	2023	중앙일보 동아일보	국민의힘, 국회, 국회의장실, 대통령, 부대
3	2023	동아일보	<u>대통령, 동아일보, 보건복지부, 정신건강혁신위원회, 정부, 경제, 병원, 의료, 민간, 웨</u>
4	2023	경향신문	정부, TV
5	2023	아시아	소식
6	2023	경향신문	국제인권기구, 국회, 대법원장, 대통령, 여당, 유족들, 사법부, 법원, 정부, 노동조합, 교수, 조직, 경기 대통령, 윤석열정부, 이개명, 검찰, 국무총리, 정부, 정치, 뉴시스 재난안전상황실, 경남도청, 정부, 행정안전부, 학교
7	2023	세계일보	대통령, 윤석열정부, 이재명, 검찰, 국무총리, 정부, 정치, 뉴시스
8	2023	노컷뉴스 디지털 파이낸 머니투	재난안전상황실, 경남도청, 정부, 행정안전부, 학교
9	2023	디지털	
10	2023	파이낸	국회, 대통령, 여당, 정부, 정치, 민생
11	2023	머니투	국회, 대통령, 여당, 정부, 정치, 민생 국회, 여당, 법안들, 정부, 정치, 민생 경향신문, 국가인권위원회, 서울경찰청, 소위원장, 언론, 정부, 조직, 시민사회, 병 국가인권위원회, 유족들, 정부, 병원
12	2023	경향신문 경향신문 경향신문 경향신문	경향신문, 국가인권위원회, 서울경찰청, 소위원장, 언론, 정부, 조직, 시민사회, 병
13	2023	경향신문	국가인권위원회, 유족들, 정부, 병원
14	2023	경향신문	경향신문, 국가인권위원회, 서울경찰청, 소위원장, 언론, 정부, 조직, 시민사회, 병웬
15	2023	YTN	경향신문, 국가인권위원회, 서울경찰청, 소위원장, 언론, 정부, 조직, 시민사회, 병원 국무위원, 국민의힘, 대통령, 동두천시의원, 법무부, 언론, 이재명, 채널A, 혁신위, 충남도, 정부, 고 국회, 대통령, 언론, 행정, 정부, 행정안전부, 민간
16	2023	미디어	국회, 대통령, 언론, 행정, 정부, 행정안전부, 민간
17	2023	헤럴드	국회, 여당, 법사위원장, 검찰, 아동학대, 정부, 정치, 박주민, 경제, 민생
18	2023	헤럴드 헤럴드	국회, 여당, 법사위원장, 검찰, 아동학대, 정부, 정치, 박주민, 경제, 민생 국회, 여당, 법사위원장, 검찰, 아동학대, 정부, 정치, 박주민, 경제, 민생 법인세, 연구소장, 경제, TV
19	2023	아시아	법인세, 연구소장, 경제, TV
20	2023	일요신문	경찰들, 경찰청, 광주경찰청, 서울경찰청, 검찰, 광주, 경찰관, 조직

* 뉴스 기사 별 추출 이해관계자

추출한 키워드를 기준으로 8071개의 뉴스기사에 대해 키워드열 필터링을 진행



- # Home
- 프로젝트 개요
- ₩ 활용 데이터
- 🔳 이해관계자 리스트
- ② 예측 모델

ACCOUNT PAGES

2 Profile

리스트 구축 진행 과정

대분류 코	대분류 종류	중분류 코드	중
		C101	
		C102	
		C108	위원
C1	정부	C107	,
CI	0T	C106	
		C105	
		C103	-
		C104	
		C201	
C2	정치 및 사회	C204	/
CZ	조직	C202	
		C203	L
		C302	
C3	문화 및 언론	C303	
		C301	
C4	의료	C401	
	70 11 617	C503	
C5	교육 및 연구	C502	Ç
	기관	C501	

대분류 코드	언론사 형태	언론사 코드	언론,
		M101	K
N / 1	거그비스	M102	N
M1	전국방송	M103	S
		M104	Y
	기여비스 미	M201	E
M2	지역방송 및	M202	PD
	전문방송	M203	헬로
		M301	헬로 경향 국민
		M302	국민
		M303	내일
		M304	노키
		M305	동이
		M306	매일
		M307	문회
		M308	서울
		M309	서울
		M310	세기
MO	조하 이기기	M311	아시
M3	종합 일간지	M312	아침

^{*} 언론사 대분류, 중분류 구분 코드

필터링한 뉴스기사의 언론사를 추출하여 **언론사 코드 대분류 7종**과, **중분류 코드 49종**으로 **카테고리화**를 진행해 최종 이해관계자 리스트를 구축



< 이해관계자 참여율 구축

- # Home
- 프로젝트 개요
- ₩ 활용 데이터
- 📴 이해관계자 리스트
- ② 예측 모델

ACCOUNT PAGES

Profile

참여율 구축 진행 과정

일자일	C1	C2	СЗ	C7	C5	C6	M1	M2	МЗ
2022년 10월 29일	8	1	0	0	0	0	0	0	
2022년 10월 30일	1,289	326	152	3	64	61	226	3	1,56
2022년 10월 31일	1,380	516	273	6	215	120	486	12	1,71
2022년 11월 1일	1,037	359	174	5	124	101	321	7	1,10
2022년 11월 2일	1,077	319	189	19	109	59	301	0	1,17
2022년 11월 3일	1,186	338	172	13	82	83	317	5	1,21
2022년 11월 4일	1,056	324	180	16	48	101	474	0	99
2022년 11월 5일	413	151	77	2	30	48	135	0	46
2022년 11월 6일	363	124	53	2	32	51	44	0	42
2022년 11월 7일	1,091	510	168	7	73	126	350	1	1,27
2022년 11월 8일	960	369	138	6	40	52	257	2	1,00
2022년 11월 9일	509	217	80	7	49	63	134	0	63
2022년 11월 10일	397	178	90	3	40	56	72	0	6(
2022년 11월 11일	293	132	61	3	19	35	125	0	29
2022년 11원 12인	107	QΛ	22	S	1/	17	5	0	26

^{*} 이해관계자별 출현 빈도

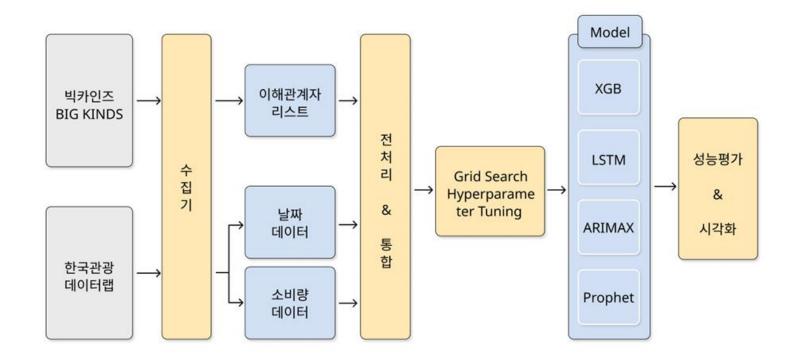
날짜별로 전체 뉴스 기사에 대한 이해관계자의 출현 빈도를 구해 사회 이슈에 대한 이해관계자의 참여율을 구축



- # Home
- 프로젝트 개요
- ₩ 활용 데이터
- 이해관계자 리스트
- 🕜 예측 모델

ACCOUNT PAGES

2 Profile



- '빅카인즈'와 '한국관광 데이터랩'에서 언론데이터와 소비데이터를 수집한 후 각 데이터에서 이해관계자 리스트와 날짜데이터, 소비량데이터를 구축.
- 이후 전처리 및 통합과정을 거치고 독립변수를 설정해 하이퍼파라미터 튜닝을 통하여 모델링 시행.
- 각 모델의 오차율을 구하고 시각화 과정을 거쳐 성능이 가장 좋은 외생변수를 더한 Prophet모델을 채택.



Home

■ 프로젝트 개요

세 활용 데이터

■ 이해관계자 리스트

🕜 예측 모델

ACCOUNT PAGES

2 Profile

XGB / LSTM / Arimax / Prophet / Prophet+ev

모델별 사용 변수 정보

- 날짜 변수 : 이태원 참사 발생일로부터 이후 1년치 일간 날짜 활용 - 외생변수 : 빅카인즈 활용 수집 데이터에서 추출한 이해관계자 리스트

- 소비 변수 : 관광 총 소비 일간 데이터 활용

x (독립변수)	y (종속변수)	
외생	소비	XGB
외생+날짜	소비	LSTM / ARIMAX / Prophet+ev
날짜	소비	Prophet

* 모델별 사용 변수



- # Home
- 프로젝트 개요
- ₩ 활용 데이터
- 이해관계자 리스트
- 🧑 예측 모델

ACCOUNT PAGES

Profile

XGB / LSTM / Arimax / Prophet / Prophet+ev

X (독립변수) : 외생 / y (종속변수) : 소비

- XGBoost Regressor
- 앙상블 학습 기법 중 하나인 그래디언트 부스팅을 기반으로 한 회귀 모델
- 다양한 유형의 데이터에 적용 가능하며, 예측 정확도가 높아 자주 사용



오차율: 14.93 %



Home

■ 프로젝트 개요

₩ 활용 데이터

■ 이해관계자 리스트

🕜 예측 모델

ACCOUNT PAGES

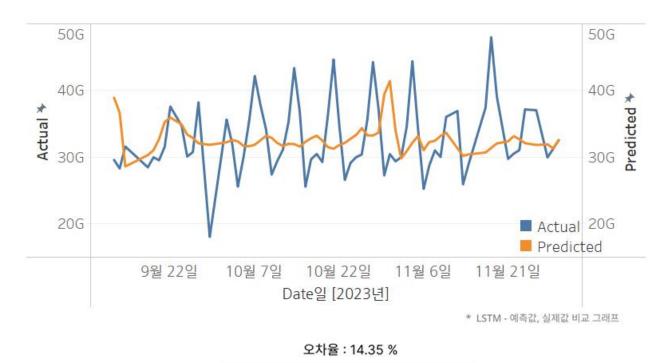
Profile

XGB / LSTM / Arimax / Prophet / Prophet+ev

X (독립변수): 외생+날짜 / y (종속변수): 소비

LSTM

- 순환 신경망(RNN)의 한 종류로, 장기 의존성을 학습하는 데 특화된 모델
- 시계열 데이터, 자연어 처리 등 순서 정보가 중요한 문제에서 성능 우수
- LSTM은 셀 상태를 이용하여 과거 정보를 효과적으로 기억하고, 현재 예측에 반영



@ Capatone Design Project Team 8D8D.



Home

■ 프로젝트 개요

₩ 활용 데이터

■ 이해관계자 리스트

🕜 예측 모델

ACCOUNT PAGES

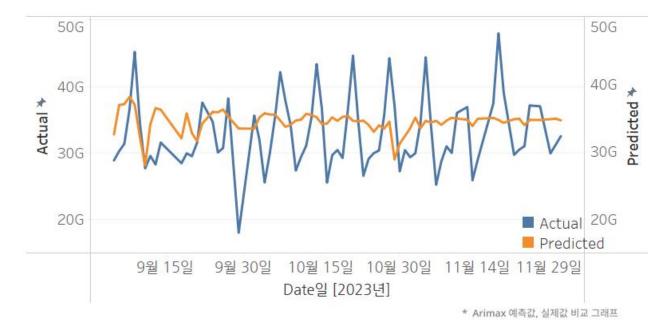
Profile

XGB / LSTM / Arimax / Prophet / Prophet+ev

X (독립변수): 외생+날짜 / y (종속변수): 소비

ARIMAX

- 시계열 데이터 분석에 주로 사용되는 모델로, 자기회귀(AR), 이동평균(MA), 차분(I) 그리고 외생 변수(X)를 결합한 모델
- 시계열 데이터의 추세, 계절성, 그리고 외생 변수의 영향을 동시에 고려하여 예측



오차율: 14.84 %



Home

■ 프로젝트 개요

₩ 활용 데이터

■ 이해관계자 리스트

② 예측 모델

ACCOUNT PAGES

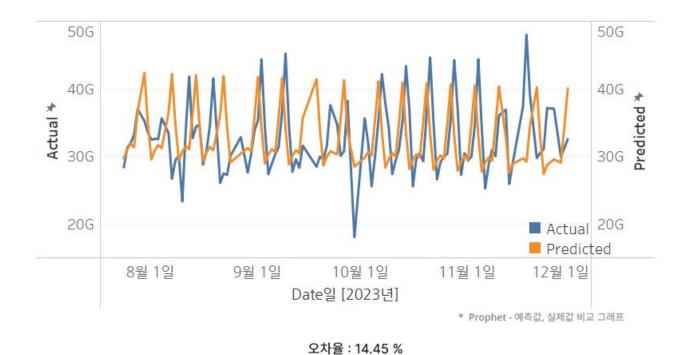
Profile

XGB / LSTM / Arimax / Prophet / Prophet+ev

X (독립변수): 날짜 / y (종속변수): 소비

Prophet

- 페이스북에서 개발한 시계열 예측 모델로, 트렌드, 계절성, 휴일 효과를 자동으로 학습
- 비교적 적은 데이터로도 높은 예측 성능을 보이며, 특히 불규칙한 패턴을 가진 시계열 데이터에 적합



@ Capatone Design Project Team 8D8D.



Prophet+ev

Home

■ 프로젝트 개요

₩ 활용 데이터

🖺 이해관계자 리스트

② 예측 모델

ACCOUNT PAGES

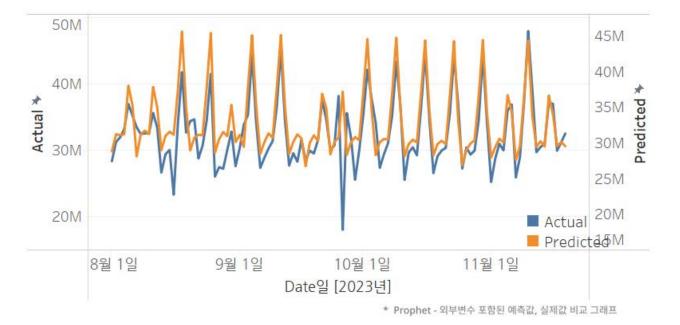
Profile

XGB / LSTM / Arimax / Prophet / Prophet+ev

- 페이스북에서 개발한 시계열 예측 모델로, 트렌드, 계절성, 휴일 효과를 자동으로 학습

- 비교적 적은 데이터로도 높은 예측 성능을 보이며, 특히 불규칙한 패턴을 가진 시계열 데이터에 적합
- Prophet 모델은 시계열 예측에 독립 변수를 포함시켜 예측 정확도를 높이는 방법 중 하나인 '추가 예측 변수(regressor)'를 설정

X (독립변수): 외생+날짜 / y (종속변수): 소비



오차율: 6.30 %



Home

▶ 프로젝트 개요

₩ 활용 데이터

🖺 이해관계자 리스트

🕜 예측 모델

ACCOUNT PAGES

2 Profile

모델 생성시 사용 데이터 정보 / 생성 모델

▌ 사용 모델 : Prophet Model

Facebook에서 개발한 시계열 데이터 예측 라이브러리로, 특히 불규칙하거나 계절성이 있는 데이터에 대해 효과적인 예측을 제공 Prophet은 추세(trend), 계절성(seasonality), 휴일 효과(holidays)를 모델링하는데 중점을 두며 아래와 같은 특징 탑재

• 추세(trend): 시계열 데이터의 전반적인 상승 또는 하강 경향을 모델링

• 계정성(seasonality): 연간, 주간, 일간 등의 주기적 패턴을 모델링

• 휴일 효과(holidays): 특정 공휴일이나 이벤트의 영향을 모델링하며 공휴일이나 특별한 날에 소비 패턴이 달라지는 경우 유용

Prophet은 정적인 **시계열 데이터 예측인 이슈 트래킹, 시간 상관 관계 분석, 트렌 드 예측, 계절성 예측 분양에서 훌륭한 성능**을 보여준다는 점은 큰 강점이다.

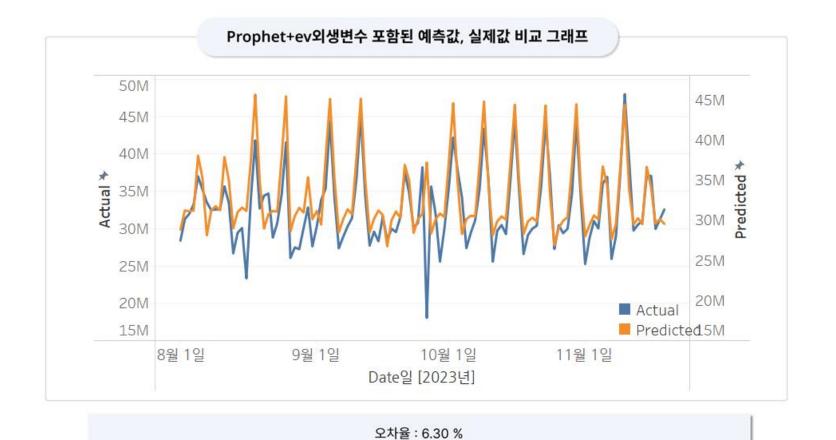


- # Home
- 프로젝트 개요
- 세 활용 데이터
- 이해관계자 리스트
- 🕜 예측 모델

ACCOUNT PAGES

Profile

모델 생성시 사용 데이터 정보 / 생성 모델





- # Home
- ▶ 프로젝트 개요
- 세 활용 데이터
- 이해관계자 리스트
- ② 예측모델

ACCOUNT PAGES

Profile



Team BDBD's GitHub

https://bestdayofbigdata22.framer.website



권유정 2201664 프로젝트 관리 및 문서 작성 대시보드 개발



2201677 시각화 계획 수립 대시보드 개발 및 관리

김세린



박민희 2202685 데이터 전처리 분석 및 모델링



장 민 지 2201769 데이터 전처리 분석 및 모델링



2201783 시각화 계획 수립 대시보드 개발 및 관리

정정현



진민경 2201792 데이터 전처리 분석 및 모델링