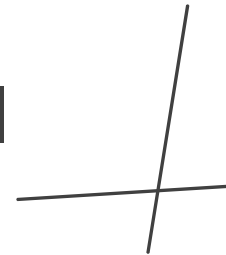


Portfolio

Be the DataScientist

G y u J i n A h n





안규진
Gyu-Jin Ahn

Ability

Software

R, Python, MySQL

관심분야

데이터마이닝, 딥러닝,
NLP, 머신러닝

datascientist.jr@gmail.com

<https://www.facebook.com/profile.php?id=100003357662264>

<https://www.instagram.com/9ugenie/>

학부시절 빅데이터 전문가가 되고 싶었습니다.

그러나, 학부에서 배우는 통계학으로는 부족했습니다.

그래서, 데이터 세미나들을 다니며 빅데이터 전문가가 되기 위한 역량을 질문하였습니다.

현직자들은 저에게 공통된 조언을 해주었습니다.

“다양한 실제 데이터를 만져봐라”

저는 제한된 환경에서 도전하기 위해 3가지 목표를 세웠습니다.

- 1) 다양한 플랫폼의 데이터 분석 공모전에 도전하자
- 2) 빅데이터 기업에 인턴쉽을 하며 실제 사회의 데이터를 다루어보자
- 3) 여러 분야의 데이터 세미나를 참석하자

다음은 제가 세운 3가지 목표를 달성하기 위한
경험들입니다.

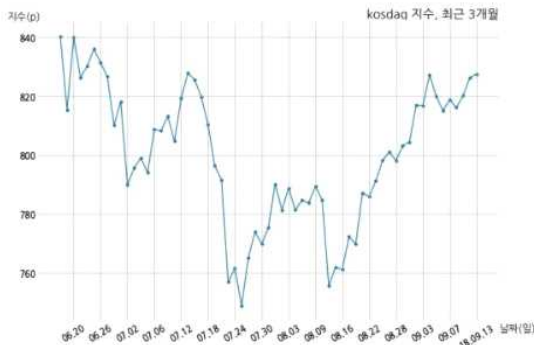
M-robo 로봇기자 서비스 (스톡봇) 개발

역할 데이터 분석 및 시스템 개발

시사 > 경제기사

[13:00] 코스닥 시장 보합세(828p, +1.23p)

입력 : 2018-09-13 13:04



M-robo 공시봇 서비스 개발

역할 데이터 분석 및 시스템 개발

OVERVIEW

기업공시나 재무재표는 일반투자자가 이해하기 쉽지 않으며 중요하지 않은 정보 또한 많음. 로봇저널리즘을 기반으로 공시 내용 중 중요한 정보만을 뽑아 요약하고 시각적으로 한눈에 볼 수 있는 실시간 서비스 개발

PM 1명 / 기획 1명 / 분석 및 개발 2명

PROCESS

- 전자공시시스템 API 활용 실시간 데이터 수집하여 DB 저장
- 실시간으로 공시를 기사로 재구성하는 알고리즘 개발
- 주가 차트 및 기업 재무 차트 시각화
- 실시간 메일링 서비스 구현

성과

데이터 오픈마켓인 데이터스토어에 API로 입점 계약 진행 중
리포트형식(pdf)으로 실시간으로 전달하는 로직 구현



중앙오션(054180)

사업보고서 (2018.06)

중앙오션 올해 매출 139 억원, 영업이익 8 억원, 당기순이익 17 억원

중앙오션은 올해까지 매출 139 억원, 영업이익 8 억원, 당기순이익 17 억원을 각각 기록했다고 09월 28일 발표했다. 전년 대비 매출액은 20 억원 감소, 영업이익은 6 억원 증가, 당기순이익은 111 억원 감소했다. 한편, 영업이익률은 6.2%로 나타났다며, 올해 부채비율은 183.5%로 작년 부채비율 189.2%보다 5.7%p 감소했다.



【주가정보】

2018-09-28 17:24:28
현재주가 3,400 원
(전날 대비 170 원 0.05% 거래량 399,990 주)

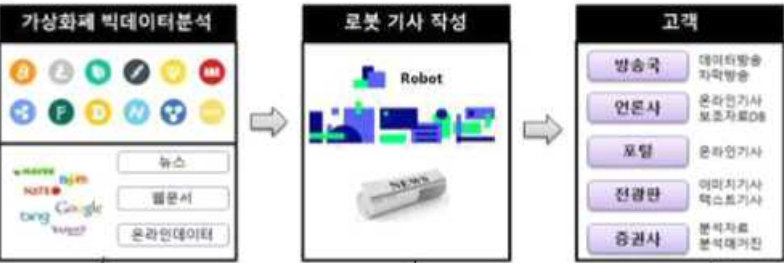
【투자정보】

시가총액 821 억원
시가총액순위 30스타 742 위
상장주식수 24,141,378
액면가1배단위 500 원 1주
주총일1권자부표 2018.09.28 30일
외국인한모투자수(A) 24,141,378
외국인보유주식수(B) 0
외국인소유율(B/A) 0.00%
투자이익1차표주가 N/A/N/A
52주최고1표가 7,150원 695
PERIEPS(WISEfn) -42.50 배 1.80 원
PERIEPS(KRX) N/A/N/A
주정 PERIEPS N/A/N/A
PBRIBPS(WISEfn) 4.20 배 1809 원
배당수익률 N/A
동일일종 PER 9.95 배
동일일종 등비율 +0.15%

공시봇 알고리즘이 실시간 작성한 리포트 sample

M-robo 가상화폐 지수 모델링

역할 데이터 분석 및 시스템 개발



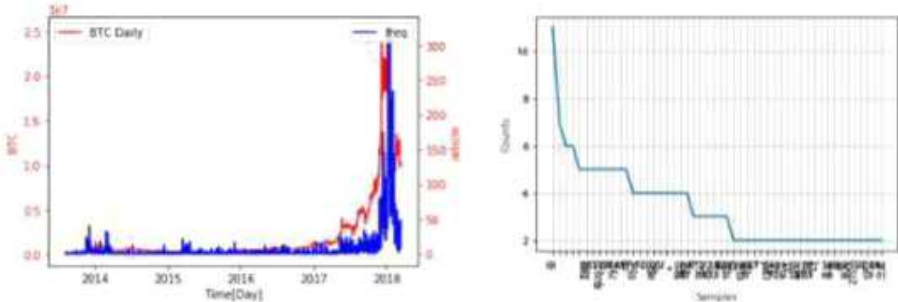
OVERVIEW

봇(Bot)과 인공지능(A.I) 기술을 이용하여 암호 화폐의 가격 변동성을 설명하는 지표를 정확하게 도출한 분석 뉴스 콘텐츠를 제공하는 로봇 저널리즘을 이용한 암호 화폐의 분석 뉴스 콘텐츠 서비스 시스템

PM 1명 / 기획 1명 / 분석 및 개발 2명

PROCESS

- 한국언론진흥재단 API이용 뉴스 데이터 확보
- 뉴스와 가상화폐 가격과의 연관성을 분석하여 가상화폐 지수를 만듦
- 뉴스 발생 빈도분석, 텍스트마이닝 이용 단어 출현 빈도 분석 등으로 가상화폐 가격과의 연관성 도출



성과

본인을 특허 발명자로 하여 특허출원 완료 (2018년 4월)

M-robo 특허정보 크로스체킹 및 가치평가 시스템

역할 데이터 분석 및 시스템 개발



OVERVIEW

기업이 가진 특허에 대한 가치평가 기술로 특허정보의 시장가치를 객관적으로 평가할 수 있는 특허정보의 확인 및 가치평가 모듈

PM 1명 / 기획 1명 / 분석 및 개발 2명

PROCESS

- 금융감독원의 기업특허공시정보와 특허청(Kipris) 특허정보를 크롤링
- 키프리스 특허정보와 기업공시 특허정보를 토픽 모델링(LDA)을 이용하여 전문에 대한 유사도 도출과 특허 세부항목 카테고리를 비교하여 유사도를 추출
- 기업특허공시 발생시점 전후로 주가의 변동성을 이용해 특허의 시장가치를 평가

성과

2018 지식재산 정보활용 대회 우수상
본인을 발명자로 하여 특허 가치평가 시스템 특허 출원(2018.12)



2018 ASIA OPEN DATA HACKATHON

폭염 피해 예측 모델링



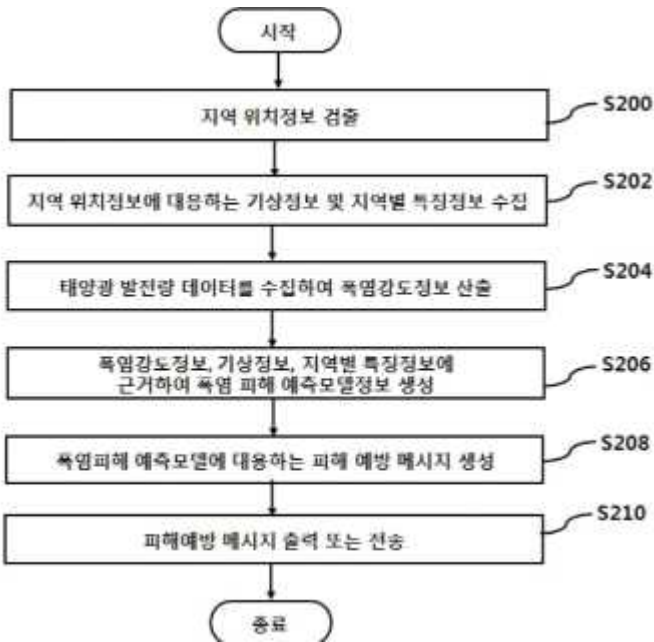
OVERVIEW

행정안전부에서 주관하는 한국, 대만, 태국, 일본이 참가하는 오픈 데이터 활용 대회로 공공 데이터를 활용하여 사회적 현안을 해결하는 아이디어를 공유하기 위함

팀장 1명 / 팀원 4명

PROCESS

- 위치별 기상정보를 활용해 태양광 발전량 예측 모델을 구축 폭염을 예측하는 서비스 제안
- Here Api를 이용한 위치 데이터를 통해 현 위치에서 실시간으로 폭염예측지수와 폭염 대피 경로를 챗봇을 통해 서비스

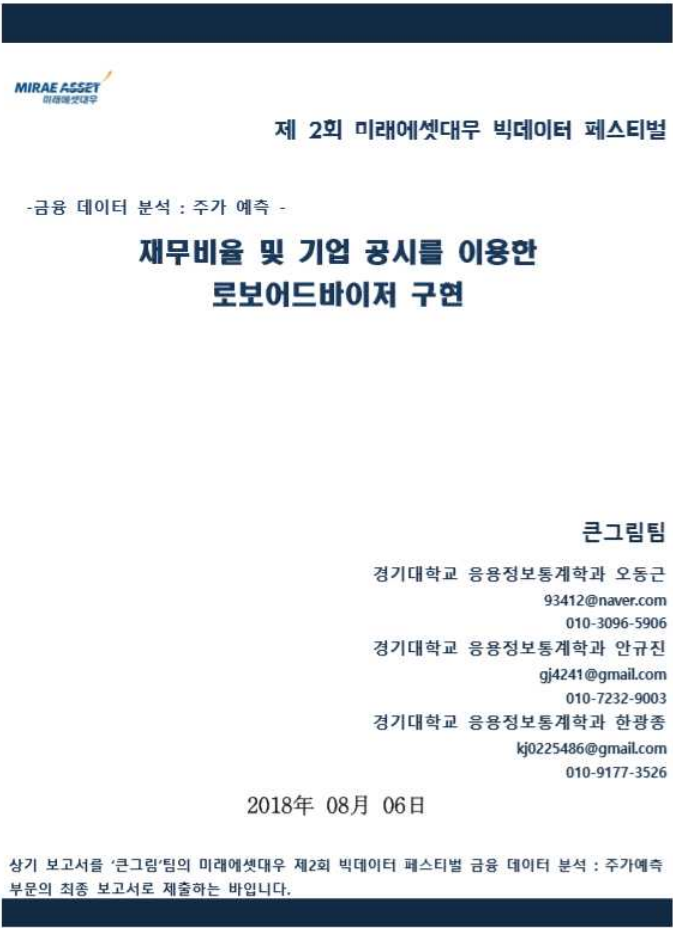


성과

예측모델링에서 예측된 값을 위치데이터와 결합해 챗봇 서비스로 접근용이성을 높임

제2회 미래에셋대우 빅데이터 페스티벌

재무비율 및 기업 공시를 이용한 로보어드바이저 구현



OVERVIEW

재무비율 및 기업 공시를 변수로 이용해 주가의 상승 or 하락을 예측, 백테스팅을 통해 이를 검증하고 실제 투자에 이용할 수 있는 로보어드바이저 구현

팀장 1명 / 팀원 2명

PROCESS

- 2000년부터 2016년까지 장기간 백테스팅
- 198개의 변수 사용
- 분석기간 중 상장된 2942개 기업의 일별 주가 데이터 수집
- 알고리즘에 의한 백테스팅 시 15년간 평균 수익률 14.16%, 누적수익률 212.36%

성과

제2회 미래에셋대우 빅데이터 페스티벌 장려상 수상

POSCO AI Challenge

철강 원료 가격 변동 추이 예측



순위	팀명/참가자
1위	BigPicture
2위	AI PP
3위	신사4
4위	HY-AI LAB
5위	포코포코
6위	서*일

OVERVIEW

창의적인 AI 솔루션 발굴을 통하여 중국 철강 수입가의 가격을 예측하는 프로젝트

팀장 1명 / 팀원 2명

PROCESS

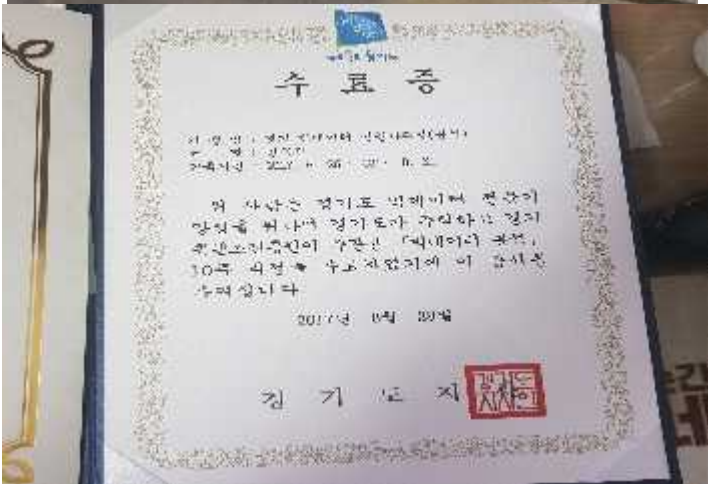
- 선박운임지수 데이터, 다양한 원자재가격 등의 historial 데이터를 사용
- T+1 (다음주 가격)을 예측하기 위해 데이터를 정제하고 코사인유사도, 시점별 피어슨 상관계수를 도출하여 선행변수를 추출
- 여러 시점별 데이터셋을 정렬하여 LSTM, XGb등 최종 알고리즘 개발

성과

예선 1차 리더보드에서 예측 1등
T+2, T+3 등 그이상의 시점을 예측하지 못함
시계열 데이터에 대한 해석을 알게된 계기

2017 경기 빅데이터 분석 전문가 양성과정

2017.06~2017.08



OVERVIEW

경기도청과 경기콘텐츠진흥원이 주관하는 빅데이터 전문인력 양성과정. 빅데이터 관련 기술과 원리를 R과 Pyhton의 실습과정을 통해 진행. Python을 통한 머신러닝 알고리즘 학습을 통한 빅데이터 전문인력 양성 교육

PROCESS

- 통계학과 재학중 학부로서의 데이터 분석 역량 개발의 한계를 느껴 실제 빅데이터 처리와 분석과정을 공부하기위해 지원하였음
- 3주간의 실습 프로젝트에서 '경기도 태양광 발전량 예측모델 및 최적 발전소 입지 선정 시각화' 프로젝트를 진행하였음

성과

머신러닝의 기초적원리와 딥러닝에 대한 실습을 통해 데이터 분석가로서의 목표가 확고해졌음

L.POINT Big Data Competition

고객의 니즈와 취향에 맞는 맞춤형 콘텐츠 제안



OVERVIEW

롯데에서 주최한 제4회 빅데이터 컴페티션에 참여. 고객 구매 데이터를 이용해 스마트 라이프 큐레이터를 주제로 고객의 니즈와 취향에 맞는 맞춤형 콘텐츠 제안 프로젝트

팀장 1명 / 팀원 1명

PROCESS

- 단순 협업필터링 알고리즘 적용에서 벗어나 새로운 알고리즘을 만들고자 노력
- Youtube 추천알고리즘 논문, Africa TV 추천알고리즘을 참고해 연령별 추천점수를 도출
- 고객의 각 변수별 score을 파생변수로 도출하여 설명가능한 모델을 위한 노력

성과

기존의 협업필터링만 쓰는 알고리즘보다 추천시스템의 성능을 훨씬 향상

BIG CONTEST 퓨처스 리그

개봉 예정 영화 관객 수 예측



순위	영화명	개봉일	매출액
1	독전	2018-05-22	1,777,534,300
2	데드풀 2	2018-05-16	496,743,500
3	인벤션스: 인피니티 워	2018-04-25	100,945,600
4	트루스 오어 데이	2018-05-22	90,089,600
5	데자뷰	2018-05-30	90,004,900
6	한 솔로: 스타워즈 스토리	2018-05-24	97,589,700
7	배필	2018-05-17	58,286,300
8	스탠바이, 웬디	2018-05-30	37,856,300
9	피터 래빗	2018-05-16	19,626,100
10	겨울왕국의 무인	2018-02-08	11,141,000

클래식

The Classic, 2003

관람객 ? ★★★★★ 9.79

네티즌 ? ★★★★★ 9.36 내 평점 ★★★★★ 등록 >

개요 멜로/로맨스, 드라마 | 한국 | 132분 | 2003.01.30 개봉

감독 박재용

출연 손예진(지혜/주희), 조승우(준하), 조인성(상민) 더보기 >

등급 [국내] 12세 관람가

다운로드 4,050

OVERVIEW

한국정보화진흥원, 한국빅데이터연합회에서 주최한 빅데이터 경진대회인 제 5회 빅콘테스트에 참여. 개봉 예정 영화 관객 수 예측을 주제로 데이터 모델링 및 예측 알고리즘 개발

팀장 1명 / 팀원 1명

PROCESS

- 영화관입장권 통합전산망에서 웹페이지 크롤링으로 2002년부터의 데이터 확보
- 감독, 배우, 배급사 등을 점수화하여 이용
- 네이버 평점 크롤링하여 변수로 이용
- 여러 알고리즘을 ensemble하여 최종 알고리즘 개발

성과

관심 있는 분야여서 다방면의 시도가 가능했으며 도메인 지식의 중요성을 알게된 계기

활동 사진



YTN science
4차 산업혁명 JOB아라 방송영상



한국직업방송
빅데이터 로봇기자
방송영상



월드 IT쇼
스톡봇 서비스 홍보부스



싱가포르 World Fintech Festival



데이터 활용 공모전 수상

감사합니다.