#1 안녕하세요 팀C의 발표를 맡은 이지혜 입니다. 저희는 하이 고슴도치라는 금융챗봇을 제작하였습니다.

#2 프로젝트의 목차는 이렇게 되어있구요

#3 저희 하이고슴도치라는 금융챗봇은 주식투자 정보를 AI모델의 추세와 투자종목을 예측하고 실시간 정보 등을 간단한 대화로 쉽게 접근 할 수 있도록 제공하는 챗봇을 만들었습니다. (금융서비스)

#4 자연어처리기술의 발전으로 챗봇이 다양한 분야에서 각광받고 있습니다. 특히 사람이 제공하는 여러가지 서비스를 대체할수있어 많은 금융회사들이 금융서비스로 챗봇을 선보이고 있습니다.

때문에 우리는 우리가 분석한 정보를 사용자에게 제공하는 수단으로 챗봇을 선택하게 되었음

#5 프로젝트의 목표는 데이터 베이스를 구축하여 정보를 제공 할수 있게하고 재무데이터 뉴스 데이터 등을 통해 모델을 구축해서 종목, 업종의 상승을 예측합니다. 예측결과와 실시간 정보등을 챗봇으로 구현합니다.

#7 주가, 재무, 뉴스, 공시를 이용해서 수익률을 예측하고 챗봇을 통해 일반금융대화와 실시간 정보, 상승종목 추천과 시장추세의 서비스를 제공하게 되었습니다.

#8 저희가 적용한 기술들은 이렇게 되어있구요

#9 일정은 약 한달간 진행하였습니다.

#10 저희 팀원은 저 이지혜, 김원중, 유일심 총 3명이 참여했습니다.

프로세스 진행에 대하여 설명하겠습니다.

#12 수집한 데이터를 가지고 예측모델을 제작하고 퀀트전략을 이용하여 수익률을 예측하고 DB구축과 실시간 정보를 출력 할 수 있게 하였습니다.

#13 데이터 수집에는 DART와 KRX를 이용해 재무데이터와 공시데이터를 수집했고 네이버에서 실시간 데이터를 수집하여 각 데이터를 전처리 했습니다.

#14 먼저 환경구축 부분에서 AWS와 docker를 사용했는데요 AWS는 클라우드 컴퓨팅을 제공하는 플랫폼으로 저렴한 비용과 속도, 즉각적 융통성으로 많은 회사 및 다양한 사용자들이 사용하고 있습니다.

#15 AWS에서 데이터베이스의 인스턴스를 생성해서 주가재무데이터의 DB를 구축하였습니다.

#16 docker란 컨테이너 기반의 오픈소스 가상화 플랫폼입니다. 기존의 주로 OS를 가상화는 방식과 달리 docker는 프로세스를 격리 하는 방식으로 가볍고 빠르게 동작할 수 있게 합니다.

#17 docker에서 ubuntu컨테이너를 설치하고 jupyter notebook의 환경을 구축했습니다.

#18 주가데이터로 아리마 모델을 구현하여 시리얼데이터에 대해 이해를 했습니다. 아리마 모델은 ar모형과 ma모형을 합친 시계열데이터 모델로 정상성을 나타내지 않는 논스테셔녀리한 데이터를 차분을 이용해 스테셔너리한 데이터로 바꾸면서 p,d,q의 하이퍼파라미터를 구할 수 있도록 합니다.

#19 차분을 통해 논스테셔너리한 데이터를 스테셔너리한 데이터로 변환합니다.

#20 오토아리마를 이용하여 베스트 파라미터로 종목별로 주가를 예측해보았습니다.

#21 오토인코더를 통해 종목별로 주가예측을 했는데요 오토인코더란 차원축소 통해 이상치를 탐지하고 중요한 특징을 학습하므로써 노이즈가 아닌 주요 특징에 대해서만 학습하는 방식입니다.

#22 컨볼루션레이어로 모델을 구현하고 인코더와 디코더부분이 있습니다. 결과는 어큐러시와 로스그래프로 어느정도 워킹이 되는 회사들도 있었고 잘 되지 않는 회사들도 있었습니다.

#23 LSTM을 오토인코더 방식으로 구현하였고 바이너리와 리니어방식으로 두가지방식을 다 트라이얼 시도했을때

#24 오토인코더 cnn방식이 워킹하는 회사가 더 많이 있었습니다.

#25 머신러닝을 이용해서 상승 종목을 예측하였는데요. 1,3,6개월로 수익기간을 정하고 5,10,15 퍼센트의 수익률을 기준으로 타겟을 지정하여 랜덤포레스트, 로지스틱레그레이션, 서포트벡터머신, xg부스트,lbgm모델을 구현해서 학습시켰을 때 데이터에 맞는 모델을 선정해서 종목별로 모델을 만들었습니다.

#26 수익률 예측 결과는 이러합니다. 생각보다 워킹이 잘되는 회사들이 많이 있었습니다.

#27 뉴스데이터를 이용해서도 상승 종목을 예측했는데요 5년치 뉴스제목데이터를 카운트벡터라이즈 시켜서 모델에 사용했습니다.

#28 토큰화된 데이터 테이블을 만들었고

#29 500종목 중에 프리시즌기준 0.5이상 종목을 선출했고, 이렇게 종목이 선출 되었습니다.

#30 공시데이터도 머신러닝에 이용하려고 했지만 공시데이터 같은경우는 파싱이 되는 것과 안되는 문서가 공존하였고, 보고서 형식이 너무 다양해서 유의미한 텍스트를 추축하기 힘들었습니다.

#31 코스피지수를 예측에는 fb-prophet이라는 페이스북 회사에서 만든 시계열 예측 라이브러리를 이용했는데요 종속적이지 않고 커브핏팅으로 문제를 해결하는 방식으로 학습속도가 빠르고 직관적으로 이해 할 수 있는 파라미터를 통해 모형을 쉽게 조정이 가능한 장점이 있습니다.

#32 종가기준 주가를 가지고 prophet을 학습하면 테이블이 만들어지고 여기서 y-hat을 시각화를 하면 이렇게 분포 되는 것을 볼 수 있습니다.

#33 업종추천에는 예측한 코스피지수를 가지고 강세장과 약세장의 시점을 분리하여 코스피와 업종들의 코렐레이션을 구하게 됩니다. 강세장일 때에는 상관계수가 높은 업종을 추천하고 약세장에서는 상관계수가 낮은 업종을 추천하게 됩니다.

#34 퀀트 투자 전략을 비교하여 종목을 추천 했는데요 퀀트에서는 크게 기술지표와 가치투자전략을 나누어 기술지표에서는 볼린저밴드, 절대모멘텀을 사용했고, 가치투자전략에서는 마법의 공식, famaLSV를 사용했습니다. 각 전략들은 3, 6, 12개월의 리밸런싱을 통해 벤치마크 대비 오버퍼품이 가장 많이 나온 전략을 추천하였습니다.

#35 볼린저밴드는 20일 이동평균선의 플러스마이너스 곱하기 20일 이동 표준편차를 사용하는데 밴드의 상황이탈시전에 매수 시그널을 잡았습니다. (표에)종목의 코드가 잡히고 바이라고 매수시그널을 나타냅니다.

#36 절대모멘텀은 pct-체인지에 금융적 1년 기준 252을 넣었을 때 양수면 매수시그널이 되었습니다.

#37 마법의 공식은 시가총액순으로 나열한 다음 rank n개를 뽑아 roa, rank기준 총 상위 10개의 종목을 추출했습니다.

#38 famalsv는 저 PSR + 저 PBR + 저 시총으로 순위를 매겨 상위 5개의 종목을 추출했습니다.

#39 버트란 프리트레인방식으로 마스크드방식과 다음문장 예측 방식이 있는데 마스크드 방식은 문장에서 한단어를 빼고 맞추는 식으로 대명사를 맞히기에 좋고, 다음문장 예측방식은 문장 두개를 넣어서 맞으면 1, 틀리면 0을 나타내어 문장의 의미를 찾는데 좋습니다.

#40 클루버트라는 프리트레이딩 모델을 파인튜닝해서 쓰게 되었습니다. 클루버트는 전문지식을 갖춘 연구자들이 수동으로 감정 라벨링을 하여 만든 모델입니다.

#41

#42 분류결과는 긍부정으로 대부분 잘 나누어 졌습니다.

#43 지피티는 트랜스포머 구조에서 디코더 부분을 사용한 일방향모델입니다 문장생성에 강점을 가지는 장점을 가지고 있습니다. 버트와 크게 다른점은 트랜스포머에서 구조에서 버트는 인코더부분을 지피티는 디코더부분을 사용했다고 나눌수 있습니다.

#44 저희는 코지피티2라는 모델을 이용하여 챗봇을 구현하였습니다. 코지피티를 큐앤에이로 파인튜닝하여 이런식의 대화를 나오게 학습시켰습니다.

#45 텔레그램 봇이라는 유아이를 사용하여 예측모델과 DB정보 실시간정보들을 함수화하여 구동될수 있도록 했습니다.

#47 예측 모델과 퀀트전략 예측의 평균수익률을 벤치마크대비로 이런 평가지표를 보여주고 있고 #48 미리 녹화해둔 시연영상입니다.