

프로젝트 명 : 빅데이터 분석 기반 비트코인 가격 예측 모델 제작

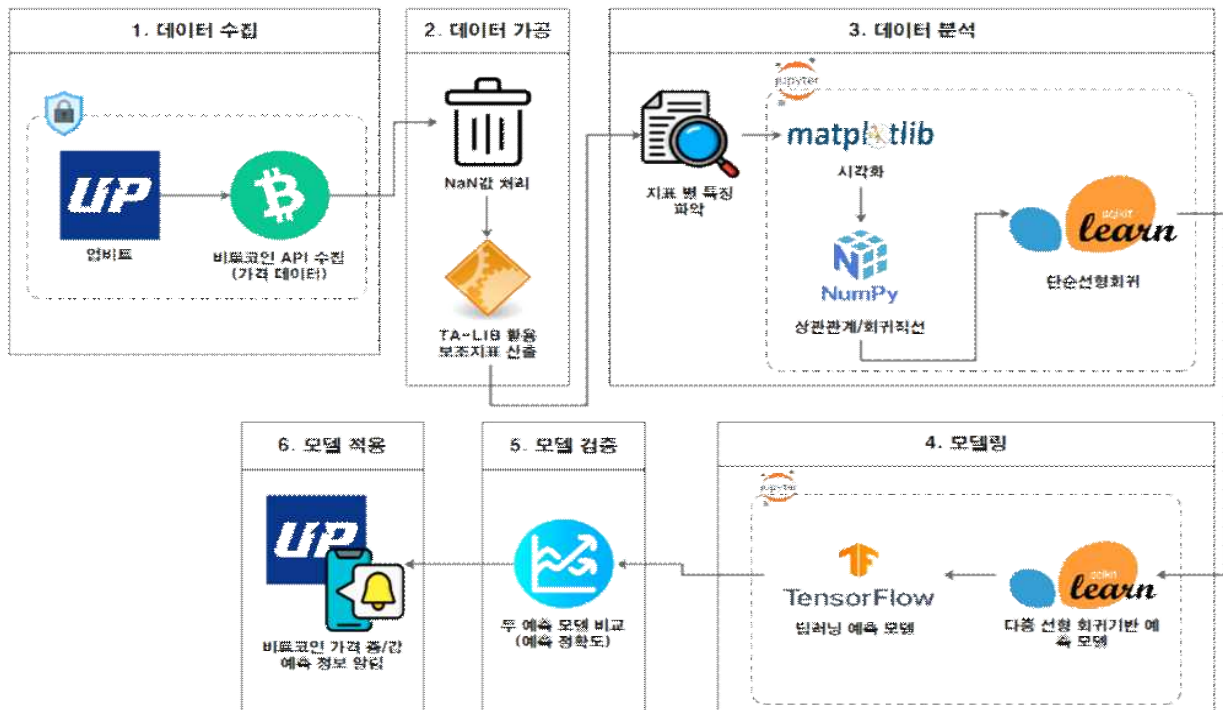
캡스톤 디자인Ⅱ 중간보고서

Version 1.0

개발 팀원 명(팀리더): 김희섭
강현욱

대표 연락처: 010-3338-6292
e-mail: djlucifer@naver.com

1. 요구사항 정의서에 명시된 기능에 대하여 현재까지 진척된 결과 및 그 내용을 기술하시오.



1) 데이터 수집(완료) : upbit 제공 api를 통해 일별 비트코인 가격 데이터 수집 완료 (23.04.21)

2) 데이터 가공(완료) : 불러온 가격데이터 전처리 작업 후, ta_lib이라는 python library를 활용, 보조지표 데이터를 산출 완료 (23.05.10)

3) 데이터 분석(완료) : 지표 별 특징 분석, 시각화, 상관분석, 회귀직선/분석을 통해 종가와의 연관성을 파악 (23.06.11)

4) 모델링(진행 중) : sklearn으로 결과 비교용 다중 선형 회귀 예측 모델 모델링 완료 (23.06.11), tensorflow를 통해 LSTM/GRU 기반의 딥러닝 예측 모델 모델링 진행 중 (10월 말 전까지 완료를 목표)

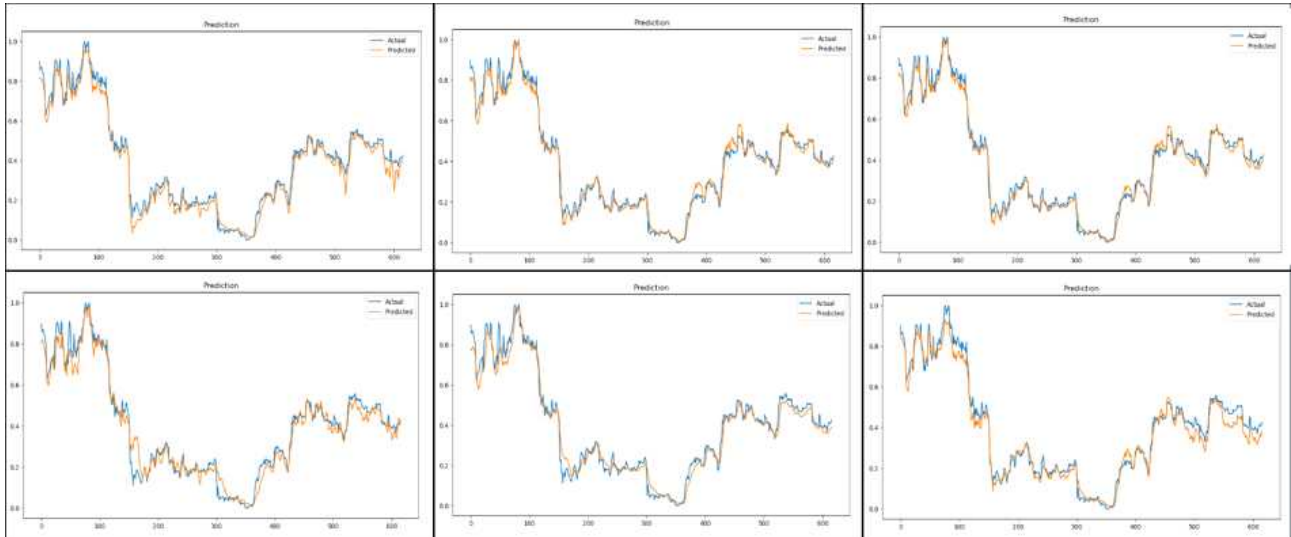
5) 모델 검증(진행 중) : 비교용 모델은 검증 완료 (23.06.11), 딥러닝 모델 진행 중 (딥러닝 모델 완성 직후 바로 진행)

6) 모델 적용 (진행 예정) : 예측 결과를 사용자에게 공지, upbit api의 매수/매도 기능을 통해 선택하도록 함 (11월 초~중순 완료를 목표)

- 현 진행 중인 딥러닝 모델링 진행 상황

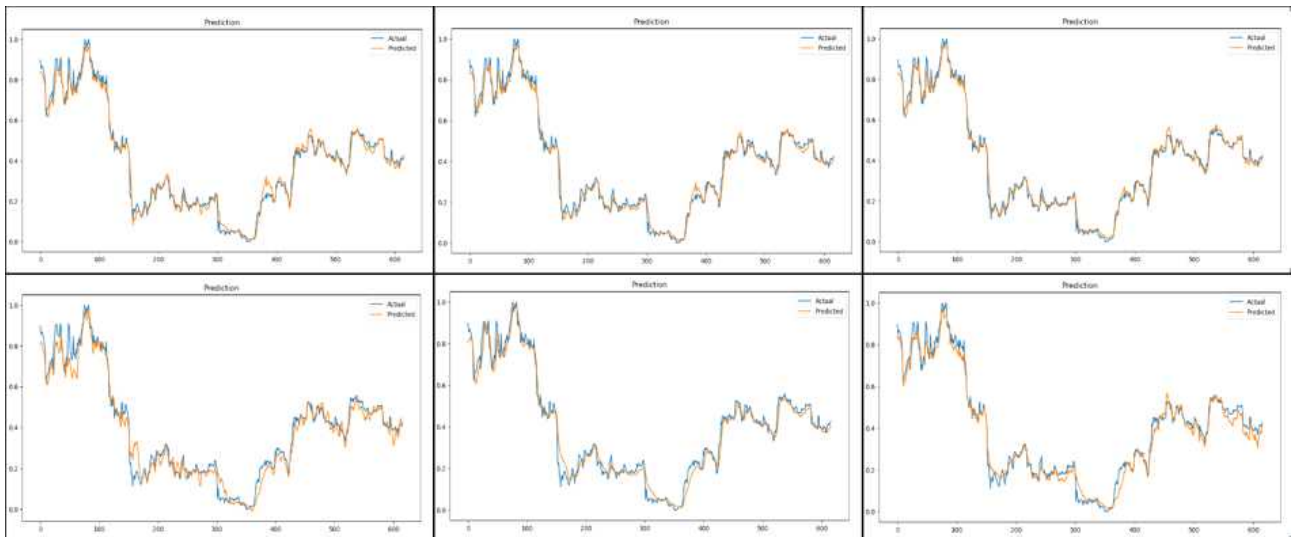
: 앞서 진행되었던 데이터 분석 작업을 토대로 학습 시킬 지표 선정 작업 진행 중, 학습 데이터로 선택되는 지표의 기준은 먼저 모든 지표들을 학습 시킨 후 지표 별 특징에 따라 진행, 지표 별 특징 끼리 조합하여 진행 후 상관분석 및 회귀직선/분석을 통해 양/음의 상관계수를 띄거나 회귀분석의 결과가 종가와 유사한 지표들로 조합하여 진행함

> LSTM 기반 좋은 예측치를 보여주는 지표 시각화 자료



순서대로 전체 지표, 추세 지표, 추세 및 시장 지표, 탄력성 및 시장 지표, 양의 상관관계 지표, 추세 및 시장과 양의 상관관계 지표

> GRU 기반 좋은 예측치를 보여주는 지표 시각화 자료



순서대로 전체 지표, 추세 지표, 추세 및 거래량 지표, 탄력성 및 시장 지표, 양의 상관관계 지표, 추세 및 양의 상관관계 지표

2. 프로젝트 수행을 위해 적용된 추진전략, 수행 방법의 결과를 작성하고, 만일 적용과정에서 문제점이 도출되었다면 그 문제를 분석하고 해결방안을 기술하시오.

1) 지표 별 특징에 따른 예측차이

종가(실제가격) 데이터를 학습 데이터로 사용하지 않는 목표에 부합하기 위해서 여러 조합으로 진행을 해본 결과 당연하게도 이동평균선과 같은 실제 가격 데이터를 기반으로 밀접하게 값이 나오는 지표와 탄력성 지표인 볼린저 밴드에서 높은 정확도를 보여줌. 이는 지표를 계산할 때 종가를 이용하는 지표라는 공통점이 있음. 하지만 같은 종가를 사용하는 RSI, MACD등과 같은 지표들은 예측 시 상당히 다른 결과를 보여주는데, 이는 각 지표 별 도출해 내는 값의 차이에 있다고 보여짐. 따라서 이동평균선과 볼린저 밴드에 학습 데이터가 편향된다고 볼 수 있지만, RSI라도 모든 추세 지표와 같이 학습 될 경우 유의미한 결과를 낸다는 점, 이런 지표들이 밀접한 지표들과 조합을 어떻게 가져가느냐에 따라 다른 결과(밀접한 지표로만 편향되지 않고 다른 결과가 나옴) 나온다는 점을 고려해 볼 때 지속해서 높은 정확도를 지니는 조합을 찾는 방향으로 해결 중.

2) 학습 때마다 다른 예측 결과

이는 같은 학습데이터를 사용한 모델로 학습을 반복해서 수행 시 다른 결과가 나옴. 이를 찾아본 결과 여러 이유가 있지만 선택한 해결 방안으로는 tensorflow의 global 및 operation seed 값을 지정하여 해결하는 방법을 선택함.

3) 그 외 정확도 향상 관련 고찰

종가 제외 지표만 사용한 모델링이 비트코인의 변동성으로 인해 베이스가 될 모델이 많지가 않기도 하며, 애초에 모델링에 있어 정답이 없다보니 정확도를 올리기 위해 여러 고민을 함. 추가적으로 생각한 진행 방향은 시각화 및 상관분석에서 종가와의 연관성이 좋았지만 학습 데이터로는 부족한 OBV, ATR과 같은 지표들을 어떻게 하면 살려서 사용할 수 있을지, 대강 학습데이터의 방향이 잡힌다면 추가적으로 성능 향상을 위한 조정(활성화 함수, epoch, batch_size 등)에 대해 해결할 예정.

프로젝트 명 : 빅데이터 분석기반 비트코인 가격 예측 모델 제작

소프트웨어 요구사항 정의서

Version 1.0

개발 팀원 명(팀리더):김희섭
강현욱

대표 연락처:010-3338-6292
e-mail: djlucifer@naer.com

목차

1. 개요
2. 시스템 장비 구성요구사항
3. 기능 요구사항
4. 성능 요구사항
5. 인터페이스 요구사항
6. 데이터 요구사항
7. 테스트 요구사항
8. 보안 요구사항
9. 품질 요구사항
10. 제약 사항
11. 프로젝트 관리 요구사항

요구사항 정의서에 사용되는 양식 설명

요구사항 고유번호(ID): 제안요청서에 정의된 요구사항에 대해 계약, 사업수행, 사업완료 및 검수까지 변경, 삭제, 수정 여부에 대한 추적관리를 위해 고유의 번호를 부여하도록 한다.

요구사항 구분 및 ID부여 규칙

요구사항 구분		ID 부여 규칙
시스템 장비 구성 요구사항	Equipment Composition Requirement	ECR-000
기능 요구사항	System Function Requirement	SFR-000
성능 요구사항	Performance Requirement	PER-000
인터페이스 요구사항	System Interface Requirement	SIR-000
데이터 요구사항	Data Requirement	DAR-000
테스트 요구사항	Test Requirement	TER-000
보안 요구사항	Security Requirement	SER-000
품질요구사항	Quality Requirement	QUR-000
제약사항	Constraint Requirement	COR-000
프로젝트 관리 요구사항	Project Mgmt. Requirement	PMR-000
프로젝트 지원 요구사항	Project Support Requirement	PSR-000

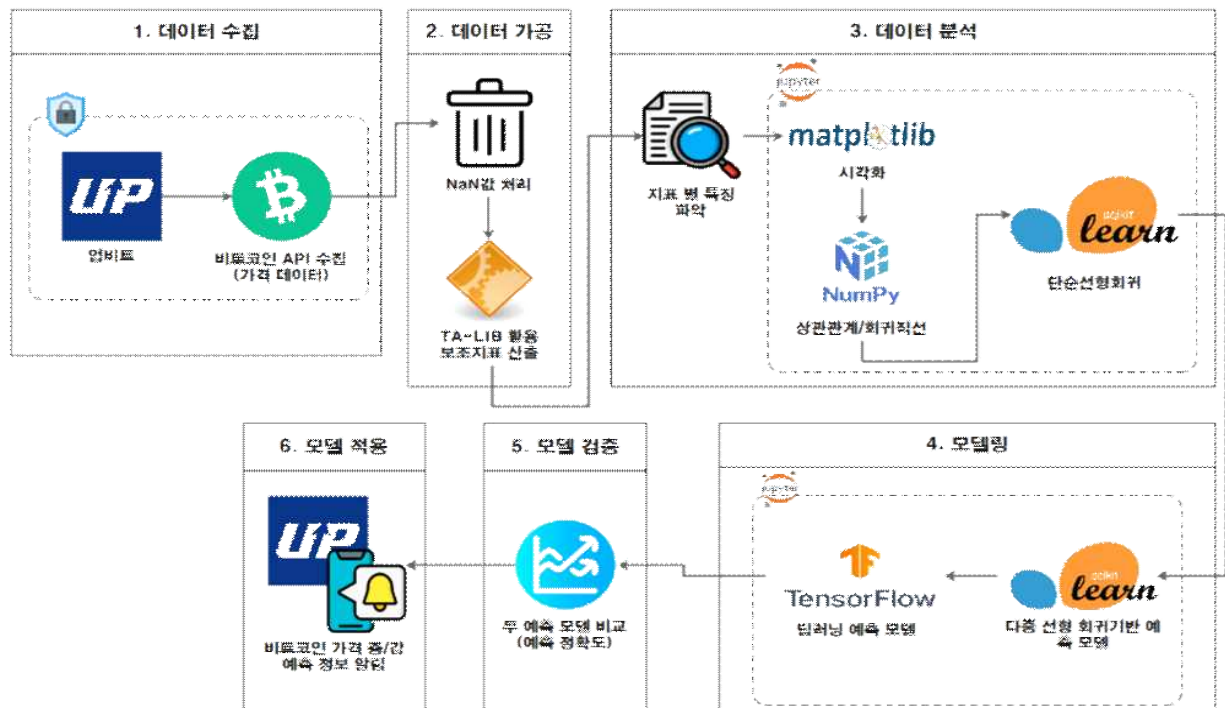
요구사항 세부내용 작성표 양식 및 항목설명

요구사항 고유번호		(설명) 요구사항 추적관리를 위해 독립적인 고유번호(ID) 부여
요구사항 명칭		(설명) 요구사항 명칭을 작성함
요구사항 분류		(설명) 요구사항 분류기준에 따른 분류를 기입
요구사항 상세 설명	정의	(설명) 요구사항 정의
	세부 내용	(설명) 요구사항 구체적인 세부 내용을 설명
산출정보		(설명) 해당기능을 통해 산출되는 결과물 혹은 정보를 표기
관련 요구사항		(설명) 정의된 요구사항과 관련된 요구사항에 대해 기술
요구사항 출처		(설명) 기능 도출내용에 대한 출처(source) 표기

1. 시스템 개요

비트코인은 가격 변동성이 크고 예측하기 어렵기 때문에 본 연구에서는 기존 연구와 비교해 상대적으로 더 높은 예측 정확도와 신뢰성을 가지는 예측모델을 만드는 것을 목표로 한다. 따라서 기존 연구들에 사용하는 증가, 고가, 저가 등의 가격 데이터가 아닌 투자 시 참고로 쓰이는 보조지표들 중에 비트코인에 적합한 지표들을 학습데이터로 하여 현 가격의 상승/하락을 예측한다.

이를 활용하여 투자자가 매수/매도 혹은 그대로 유지할지 선택할 때 해당 모델의 예측 결과를 바탕으로 도움을 주고자 한다.



2. 시스템 장비 구성요구사항

- » 장비품목 : Jupyter notebook, Tensorflow
- » 장비 수량 : 단일
- » 장비 기능 : 데이터 분석, 훈련
- » 장비 성능 및 특징 : 데이터 분석에 사용되는 오픈 소스 웹 어플리케이션 및 머신러닝을 위한 오픈소스 플랫폼

3. 기능 요구사항

요구사항 고유 번호	SFR-001		
요구사항 명칭	비트코인 가격 데이터 호출		
요구사항 분류	기능	응락수준	필수
요구사항 상세 설명	정의	업비트 API를 이용해 비트코인 가격 데이터 수집	
	세부 내용	1. open, high, low, close, volume, value 값을 수집 2. 수집한 데이터를 csv 파일로 저장	

요구사항 고유 번호	SFR-002		
요구사항 명칭	보조지표 계산		
요구사항 분류	기능	응락수준	필수
요구사항 상세 설명	정의	수집한 가격 데이터를 이용해 보조지표 계산	
	세부 내용	ta-lib 패키지를 이용해 각 보조지표들을 계산	

요구사항 고유 번호	SFR-003		
요구사항 명칭	보조지표 데이터 전처리		
요구사항 분류	기능	응락수준	필수
요구사항 상세 설명	정의	보조지표 데이터 가공	
	세부 내용	1. 시각화 2. 실제 가격과의 상관관계수 계산 3. 기술 및 시계열 분석	

요구사항 고유 번호	SFR-004		
요구사항 명칭	LSTM 모델 훈련 및 예측		
요구사항 분류	기능	응락수준	필수
요구사항 상세 설명	정의	시분할 모델 중 LSTM 모델을 이용해 훈련	
	세부 내용	1. 학습용 데이터로 선정된 지표들을 여러 방식으로 조합해 보며 학습 2. 파라미터 최적화	

4. 성능 요구사항

요구사항 고유 번호	PER-001		
요구사항 명칭	예측 정확도 및 소요 시간		
요구사항 분류	성능	응락수준	필수
요구사항 상세 설명	정의	예측 모델의 정확도 및 소요 시간	
	세부 내용	1. 학습 시 학습시킬 보조지표들의 조합을 통해 최적의 정확도 보이도록 함 2. 예측 범위에 따른 소요시간을 조절(분봉 예측 시 예측 소요 시간이 이를 넘어서는 안됨)	

5. 인터페이스 요구사항

요구사항 고유 번호	SIR-001		
요구사항 명칭	예측 결과 시각화		
요구사항 분류	인터페이스	응락수준	필수
요구사항 상세 설명	정의	모델의 예측을 확인하기 위한 결과의 시각화	
	세부 내용	모델의 예측값과 실제 비트코인 가격 디스플레이	

6. 데이터 요구사항

요구사항 고유 번호	DAR-001		
요구사항 명칭	데이터 저장		
요구사항 분류	데이터	응락수준	필수
요구사항 상세 설명	정의	수집한 비트코인 가격 데이터 및 학습용 보조지표 데이터 저장	
	세부 내용	1. 보조지표 별 각각 따로 데이터를 저장 2. 각 데이터는 csv 파일로 저장	

7. 테스트 요구사항

요구사항 고유 번호	TER-001		
요구사항 명칭	예측 모델 테스트		
요구사항 분류	테스트	응락수준	필수
요구사항 상세 설명	정의	예측값의 정확도 테스트	
	세부 내용	1. 실제 비트코인 가격 시각화 2. 예측값 시각화 3. 두 값을 비교하여 파라미터 값을 조정하여 최적화	

8. 보안 요구사항

요구사항 고유 번호	SER-001		
요구사항 명칭	API 발급 키 보안		
요구사항 분류	보안	응락수준	필수
요구사항 상세 설명	정의	API 발급을 위한 개인키 보안	
	세부 내용	예측 모델 학습과 추후 성능 개선에 쓰일 업비트 api 공개키 및 암호키 비공개	

9. 품질 요구사항

요구사항 고유 번호	QUR-001		
요구사항 명칭	예측 모델의 실용성		
요구사항 분류	품질	응락수준	필수
요구사항 상세 설명	정의	실제 투자에서의 적합성	
	세부 내용	1. 예측 모델을 통한 비트코인 모의투자를 통해 수익률 기록 2. 비트코인의 변동성을 고려, 최소한의 손실률을 기록	

10. 제약 사항

요구사항 고유 번호	COR-001		
요구사항 명칭	개발에서의 제약사항		
요구사항 분류	제약사항	응락수준	필수
요구사항 상세 설명	정의	개발 과정에서의 제약사항	
	세부 내용	1. 여러 보조지표들을 학습 데이터로 사용하기 위해 각 지표 값을 통일할 필요성 2. jupyter notebook, tensorflow 이용을 위한 python 사용 3. 지속해서 변동하는 비트코인 가격으로 인한 지속적인 데이터 갱신 필요	

11. 프로젝트 관리 요구사항

요구사항 고유 번호	PMR-001		
요구사항 명칭	품질관리		
요구사항 분류	프로젝트 관리	응락수준	필수
요구사항 상세 설명	정의	프로젝트 관리	
	세부 내용	1. 우선적으로 예측 모델을 만들기 위한 데이터셋 구축 2. 데이터셋을 여러 방식으로 조합해 보며 최적의 예측 모델 결정 3. 추후 예측 정확도를 높이기 위한 추가적인 데이터셋을 구축하거나 자동 매수, 매도 시스템을 추가	