

캡스톤디자인 중간보고서

제 목	국문	LSTM을 이용한 주식 가격 예측		
	영문	Stock Price Prediction Using LSTM		
진 행 상 황	중요마일스톤	1) 뉴스 크롤링 - 소스코드 작성 2) 데이터 전처리 - 소스코드 작성 3) 뉴스 감정분석 - 소스코드 작성 4) LSTM모델 훈련 - 소스코드 작성		
	진행상황	1) 뉴스 크롤링 - 프로토타입 작성 완료 2) 데이터 전처리 - 프로토타입 작성 완료 3) 뉴스 감정분석 - 구현 논의 중 4) LSTM모델 훈련 - 구현 논의 중		
산출물	요구사항 정의서(별첨 1), 중간보고서(별첨 2)			
팀 구성원	학년	학 번	이 름	연락처(전화번호/이메일)
	4	20151970	진용원	010-6806-3566 / dbw0812@naver.com
	4	20157092	김상민	010-4922-3675 / ksm2485@naver.com
<p>컴퓨터공학과의 프로젝트 관리규정에 따라 다음과 같이 요구사항 정의서와 중간보고서를 제출합니다</p> <p style="text-align: center;">2015 년 월 일</p> <p style="text-align: right;">책임자 : (인) 지도교수 : (인)</p>				

[별첨1]

프로젝트명 : LSTM을 이용한 주식 가격 예측

소프트웨어 요구사항 정의서

Version 1.0

개발 팀원 명(팀리더):진용원
김상민

대표 연락처:010-6806-3566
e-mail: dbw0812@naver.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

1. 시스템 개요

목표 : 뉴스 데이터를 기반으로 자연어처리(감성분석)을 활용한 주가 방향 예측 모델 구현

네이버 뉴스 -> PYTHON -> 데이터 전처리 -> 감성분석 -> LSTM모델 훈련

2. 시스템 장비 구성요구사항

- » 장비품목 : Tensorflow
- » 장비 수량 : 단일
- » 장비 기능 : 데이터분석 및 훈련
- » 장비 성능 및 특징 : 머신러닝을 위한 오픈소스 플랫폼

3. 기능 요구사항

요구사항 고유 번호	SFR-001	
요구사항 명칭	뉴스 크롤링	
요구사항 분류	기능	
요구사항 상세 설명	정의	네이버 뉴스에서 기업 관련 뉴스 데이터 수집
	세부 내용	1. 네이버 뉴스에서 기업 관련 뉴스 수집 2. 수집한 데이터 파일 형태로 저장

요구사항 고유 번호	SFR-002	
요구사항 명칭	데이터 전처리	
요구사항 분류	기능	
요구사항 상세 설명	정의	수집한 뉴스 데이터 가공
	세부 내용	1. 형태소 분석 2. 불용어 처리

요구사항 고유 번호	SFR-003	
요구사항 명칭	자연어 처리(감성분석)	
요구사항 분류	기능	
요구사항 상세 설명	정의	감성분석 모델 적용
	세부 내용	1. 학습셋, 테스트셋 분리 2. 모델 학습

요구사항 고유 번호	SFR-004	
요구사항 명칭	LSTM모델 훈련 및 예측	
요구사항 분류	기능	
요구사항 상세 설명	정의	LSTM모델을 사용해 훈련
	세부 내용	1. 시계열 데이터 분석에 특화된 LSTM 모델 사용 2. 파라미터 최적화

--	--	--

4. 성능 요구사항

요구사항 고유 번호	PER-001	
요구사항 명칭	정확도	
요구사항 분류	성능	
요구사항 상세 설명	정의	예측 데이터 모델 정확도
	세부 내용	수집한 데이터와 테스트셋을 통해 정확도를 향상

5. 인터페이스 요구사항

요구사항 고유 번호	SIR-002	
요구사항 명칭	모델 성능 시각화	
요구사항 분류	인터페이스	

요구사항 상세 설명	정의	모델의 성능의 확인을 위한 정확도 시각화
	세부 내용	1. 데이터 전처리에 워드 클라우드 구현 2. 실제 주가와 예측주가 디스플레이

6. 데이터 요구사항

요구사항 고유 번호	DAR-001	
요구사항 명칭	데이터 저장	
요구사항 분류	데이터	
요구사항 상세 설명	정의	수집한 뉴스 데이터 저장
	세부 내용	1. 수집한 뉴스 데이터와 주가변동을 묶어서 저장 2. Csv 파일 형태로 저장

7. 테스트 요구사항

요구사항 고유 번호	TER-001	
요구사항 명칭	예측 모델 테스트	
요구사항 분류	테스트	

요구사항 상세 설명	정의	예측 모델 정확도 테스트
	세부 내용	1. 실제 주가와 예측 주가 시각화 2. 모델의 파라미터 조정

8. 보안 요구사항

요구사항 고유 번호	SER-001	
요구사항 명칭	데이터 암호화	
요구사항 분류	보안	
요구사항 상세 설명	정의	
	세부 내용	예측 결과값 도출식 비공개

9. 품질 요구사항

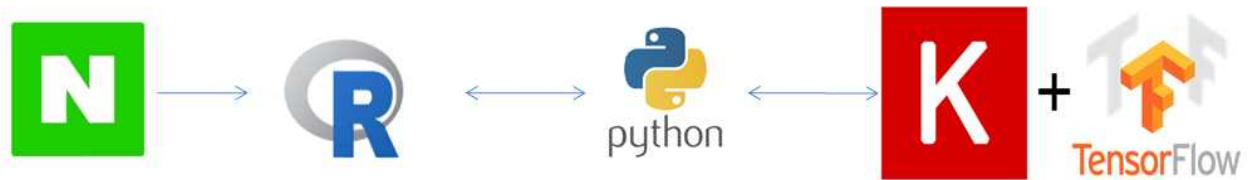
요구사항 고유 번호	QUR-001	
요구사항 명칭	사용자 이해성	
요구사항 분류	품질	

요구사항 상세 설명	정의	사용자 이해성
	세부 내용	1. 모델의 성능을 쉽게 평가할 수 있도록 시각화 2. 모의투자 수익률 기록

10. 제약 사항

요구사항 고유 번호	COR-001	
요구사항 명칭	프로젝트 개발	
요구사항 분류	제약사항	
요구사항 상세 설명	정의	시스템 개발 제약사항
	세부 내용	1. TensorFlow 를 이용하여, 프로그래밍 언어로 Python 사용 2. 뉴스 크롤링시 웹사이트 변화에 맞춰 소스코드 수정 필요

11. 프로젝트 관리 요구사항



	3	4	5	6	7	8	9	10
자료수집 & 현황조사								
환경 구축								
소프트웨어 설계								
테스트 & 수정								
기능 개선 & 완성								
최종보고서 작성 & 발표								

[별첨2]

중간보고서

1. 요구사항 정의서에 명시된 기능에 대하여 현재까지 분석, 설계, 구현(소스코드 작성) 및 테스트한 내용을 기술하시오.

네이버 뉴스 -> PYTHON -> 데이터 전처리 -> 감성분석 -> LSTM모델 훈련

기능별 상세 요구사항

뉴스 크롤링 : 네이버 뉴스 헤드라인, 본문, 일자 데이터 크롤링해서 파일로 저장

데이터 전처리 : 문장부호, 불용어 제거, 텍스트 토큰화, 워드 클라우드로 시각화

뉴스 감성분석 : 학습셋, 테스트셋 분리하고, 모델 학습

LSTM 모델 훈련 : 하이퍼 파라미터 조정하며 학습, 입력 데이터별로 성능 비교

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

데이터를 모델에 학습시켰을 때 주가 예측이 전날과 다음날의 예측 결과 값이 동일한 문제점이 발생

주식가격을 sequence 데이터로 LSTM에 집어넣는게 맞는지 고민을 하여 분석 후 수정

//수익률(가격) 말고 변수(뉴스, 코스닥, 나스닥 지수 등)가 많기 때문에 주식가격을 배제 후 테스트//

주식가격 데이터는 시계열 데이터이므로 평균, 분산이 일정한 데이터를 확인하여

[오늘가격-전날가격] 이러한 방법으로 검증

캡스톤 디자인 | 중간보고서 채점표

평가도구	평 가 항 목	평 가 점 수				
		1	2	3	4	5
중간 보고서 및 실행 결과	1. 요구사항 정의서(기능, 성능, 인터페이스 등)가 구체적으로 작성되었는가?					
	2. 요구분석, 설계 산출물(모델, 프로토타입 등)의 내용이 충실한가?					
	3. 설계 및 구현 문제를 위해 적용한 이론, 문제해결 방법이 제시되었으며 그 적용이 적합한가?					
	4. 구현된 소프트웨어(또는 이와 동등한 하드웨어 시스템)가 버그 없이 실행되었는가?					
	5. 구현된 소프트웨어(또는 이와 동등한 하드웨어 시스템)의 성능 요구사항은 충족되었는가?					
도구활용	6. 설계 및 구현을 위해 도구가 적절히 활용되었는가?					
	7. 도구의 활용수준(능숙도)은 프로젝트 수행에 적합한가?					
팀원의 업무 및 역할	8. 팀원의 업무분담에 따른 역할 및 협력이 충실히 이루어졌는가? (평가자에 의한 질의)					
	9. 프로젝트 중간 진척상황에 대해 팀원이 충분히 인지하고 있는가?(평가자에 의한 질의)					
합계						
*검토 의견(최종완료 때까지 보완해야할 점에 대해 작성 요망)						
심사위원(소속):		(이름)			(인)	

