

산학연계SW프로젝트 중간보고서

프로젝트 제목 : SNS 정보를 활용한 개인신용평가 서비스

프로젝트 수행기간 : 20xx.xx.01 ~ 20xx.xx.xx.

프로젝트 팀명 : int

지도 교수 : 정보융합학부 김현경 교수님


참여업체 명 : 에잇퍼센트

2020. 12. 23



광운대학교
KwangWoon University

산학연계SW프로젝트 중간보고서

과 제 명	SNS 정보를 활용한 개인신용평가 서비스			
팀 명	int			
수행기간	20XX년 XX월 01일 ~ 20XX년 XX월 XX일			
과제비	총 원			
지도교수	성 명	김현경	학 부	정보융합학부
참여학생	성 명	학 부	학 번	email
	한정수	정보융합학부	2018204035	jungsu_@naver.com
	권나현	정보융합학부	2018204021	knh990313@naver.com
	윤서안	정보융합학부	2018204036	yoonsa5126@naver.com
	조민경	정보융합학부	2018204046	cmkbooty21@naver.com
참여업체	회사명	에잇퍼센트	담당자	이호성
	연락처	010-3006-2533	email	hslee@8percent.kr
산학연계SW프로젝트의 결과물에 대한 동의서				동의여부
프로젝트에 대한 지원	본 프로젝트는 과학기술정보통신부 및 정보통신기술진흥센터의 SW중심대학사업의 지원으로 수행된 것임 (과제번호: 2017-0-00096)			<input checked="" type="checkbox"/>
프로젝트 결과물의 활용	본 프로젝트의 결과물은 프로젝트에 참여하고 지원한 광운대학교 (학생들과 지도교수), 참여업체 그리고 정보통신기술진흥센터의 소유물이며 향후 활용과 소유에 관해서는 참여자 간의 협의에 의해 결정할 수 있음			<input checked="" type="checkbox"/>
『산학연계SW프로젝트』지원계획에 따라 중간보고서를 제출합니다.				
2020년 12 월 23 일				
	팀 장	한정수		
	팀 원	권나현		
	팀 원	윤서안		
	팀 원	조민경		
지도교수		김현경	(인) 김현경	
광운대학교 소프트웨어융합대학 귀하				

목 차

1. 과제의 개요	
가. 배경 및 필요성	
나. 과제의 목표 및 내용	
다. 시스템 개요	
2. 과제 수행 결과	
가. 관련 기술	
나. 중간 결과	
다. 향후 계획	
3. 과제의 평가	
가. 개선 방안	
나. 기타 보고사항	
4. 별첨	

1. 과제의 개요

가. 배경 및 필요성

최근 성별, 나이, 국적을 불문하고 수많은 사람들이 SNS를 활발하게 이용하고 있습니다. 부분적으로는 신용과 무관해 보이는 데이터들을 수집하고 분석하여 신용 평가를 할 수 있다는 점에서 흥미가 생겼습니다.

이러한 서비스가 활성화된다면 다양한 빅데이터를 기반으로 신용 평가 근거의 폭을 넓혀 부족한 금융 거래 실적으로 인해 절대적으로 불리했던 특정 계층까지 포용함으로써 보다 많은 사람들이 금융 서비스를 이용할 수 있는 기회를 제공할 것입니다. 특히 대학생, 사회초년생 또는 소상공인처럼 금융 거래 정보가 부족한 개인에게 도움이 될 것이며, 금융 정보뿐만 아니라 비금융 정보를 활용하기 때문에 더욱 합리적인 개인 신용 평가가 가능해질 것입니다. 또한 코로나19로 인해 어려움을 겪는 소상공인들을 지원함에 따라 정부의 정책 설정에도 기여할 것으로 기대됩니다.

나. 과제의 목표 및 내용

SNS 계정 연동을 통해 얻은 정보를 활용하여 신용카드 사용 이력 또는 금융 거래 정보가 부족한 대학생, 소상공인, 사회 초년생 등 불리한 신용등급 책정을 받은 금융이력부족자를 비금융 정보를 활용하여 합리적으로 평가하는 서비스를 개발합니다.

모두가 공정한 개인 신용 평가를 받을 수 있게 돕는 것이 과제의 목표입니다.

다. 시스템 개요

- 1) 인스타그램에서 해시태그를 통한 크롤링으로 랜덤 데이터 수집
- 2) 수집된 데이터를 통해 개인 신용 평가를 위한 적절한 기준 마련
- 3) 기준을 이용하여 인스타그램 정보로 간단한 신용 평가가 가능한 프로그램 제작
- 4) 사용자가 프로그램에 인스타그램 아이디와 비밀번호를 입력하여 로그인
- 5) 프로그램이 신용 등급 또는 신용 평가에 활용 가능한 스코어를 측정

2. 과제 수행 결과

가. 관련기술

파이썬과 selenium을 이용하여 데이터 수집

오토셋(AutoSet) 프로그램의 phpMyAdmin을 이용하여 데이터베이스 관리

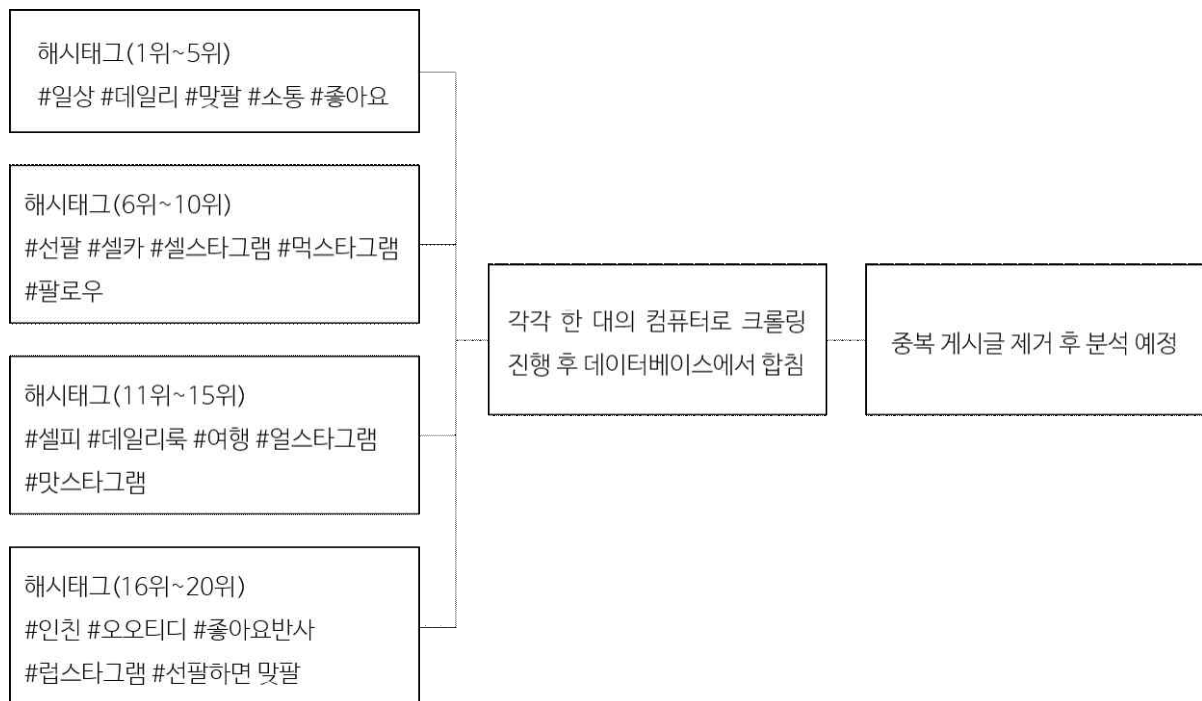
파이썬과 R을 이용하여 데이터 분석(예정)

나. 중간 결과

랜덤 데이터 수집을 위해 인스타그램에서 한국인들이 많이 사용하는 해시태그 검색으로 최근 게시글을 크롤링(해시태그 랭킹 참고 사이트: <http://cocotag.kr/>)

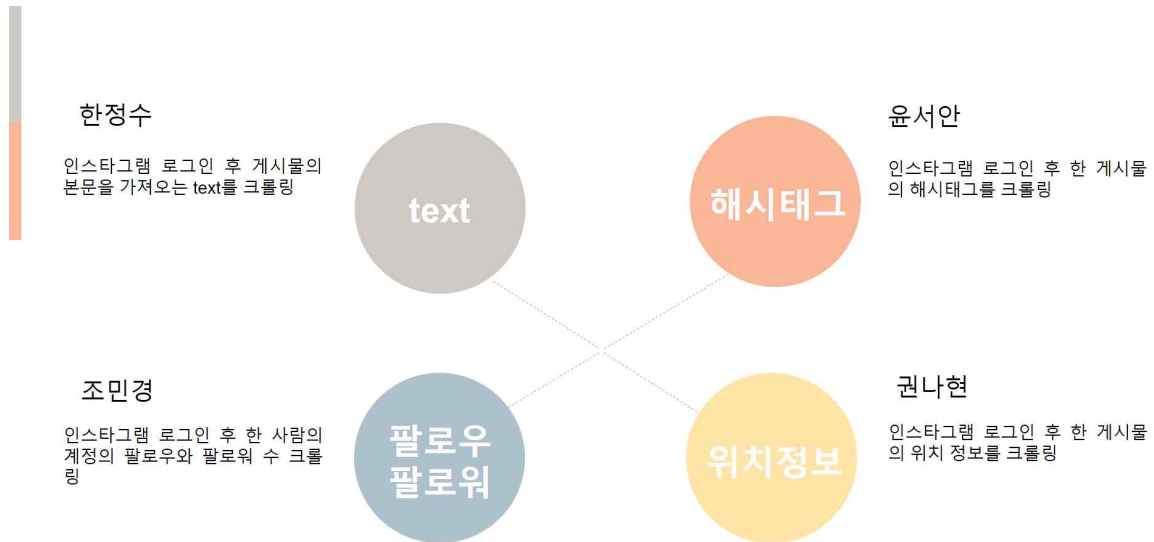
1위부터 20위까지의 해시태그를 검색하여 각각 약 500개의 게시글의 데이터를 수집(총 약 1만개)

순위가 높은 해시태그 수집을 선택한 이유: 순위가 높은 해시태그는 게시글의 내용과 관계없이 자신의 게시글이 많은 사람들에게 노출되기를 바라고 사용하는 경우가 대부분이다. 따라서 분석 시에 의미 없는 해시태그들은 따로 처리할 예정이다.



수집 내용: 게시글의 내용, 포함된 해시태그, 위치 정보, 게시글 작성자의 아이디, 작성 게시글 수, 팔로워(작성자를 팔로우하는 사람) 수, 팔로잉(작성자 본인이 팔로우하는 사람) 수

코딩 역할 분담(Jupyter notebook 이용)



데이터베이스 저장 예시

	no	url	id	postnum	followernum	followingnum	location	hashtag	text
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	379	https://www.instagram.com/p/CJGeCEIgbPy/	hyeokj	192	509	98			.
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	380	https://www.instagram.com/p/CJGaiMpFRaj/	sehee_j	1	876	198			너어무 예쁜 숙소
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	381	https://www.instagram.com/p/CJGelyPhk8n/	hyeok0_0e	4	149	145			습다
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	382	https://www.instagram.com/p/CJGjQIhX1i/	hap_pypolly	102	156	127	Seoul, Korea	#아름다#하지민#유일한#복지#자통역후#아근요...	최근전까지 내가 먹고싶은 음식은 #아름다 고
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	383	https://www.instagram.com/p/CJGeGQ3H8B8/	jwook_the_fit	48	770	231	도구마리 공촌	#탈로우#ulzzang#02년생#고딩#맞팔#선팔#선...	#탈로우 #ulzzang #02년생 #고딩 #맞팔 #선팔...
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	384	https://www.instagram.com/p/CJD8Tz6nqkA/	ye_bin_18	137	2,028	128			Like back
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	385	https://www.instagram.com/p/CJGiv29j_hn/	ring_s2	12	384	196	꿈나라ZZ		미치겠다 너무 귀엽다...
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	386	https://www.instagram.com/p/CJGjxr_inFW/	h_bin_19	129	282	459		#일상#일상스타그램#ootd#데일리#dailylook#오...	고독
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	387	https://www.instagram.com/p/CJGbmJon8Ow/	199.3.29	258	634	161	함곡동	#코로나이닝	외출 오래간만 ! #코로나이닝
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	388	https://www.instagram.com/p/CJGjbmGjUB5/	95.0.1.2.8	45	1,081	100			+10kg
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	389	https://www.instagram.com/p/CJFht7nTWq/	oo_ub	34	1,199	160			나 뭐 달라졌게 ?
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	390	https://www.instagram.com/p/CJGjvChBaSq/	this_twor	79	815	330		#첫출반시#첫출출반#감성#유산소#03년생#줄이...	Running
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	391	https://www.instagram.com/p/CJGd1eXhPel/	hyeok0_0e	4	149	145			쓰-쓰
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	392	https://www.instagram.com/p/CJGdWer7k2/	raon_s_	188	610	120			.
<input type="checkbox"/> 수정 <input type="checkbox"/> 삭제	393	https://www.instagram.com/p/CJGjyhLn14q/	ssumer.rrr	344	205	137		#오늘#일상#강아지#강아지그림#인스타그램#명일이#명팔#...	아 이렇게 하는거 아니야 . #오늘#일상#강아지#강아지그림#인스타그램 #명팔 이 #명...

다. 향후 계획

역할 분담

성명	역할
한정수	데이터 전처리, 분석, UI 만들기, 보고서 작성
권나현	데이터 전처리, 분석, UI 만들기, 보고서 작성
윤서안	데이터 전처리, 분석, UI 만들기, 보고서 작성
조민경	데이터 전처리, 분석, UI 만들기, 보고서 작성

수행 일정

일정	내용	세부 내용	비고
1	기존 크롤링 코드 최종 수정	예외 처리가 필요한 부분들을 추가 작업	
		오류 없이 빨리 처리하도록 적절한 시간 설정	
		데이터베이스에 누락 없이 저장되는지 확인	
2	데이터 전처리와 분석	수집한 데이터 전처리	설문조사 진행 예정
		다양한 분석 기법들로 데이터 분석	
		적절한 평가 기준 마련	
3	UI와 추가 작업	사용자가 사용할 프로그램의 UI 디자인	
		부가적인 기능들 추가	
		최종 UI 만들기	
4	최종 검토하기	전체적인 재검토	
		신용 평가가 적절하게 이루어지는지 확인	

3. 과제의 평가

가. 개선 방안

현재까지의 과제 수행에 대한 개선 사항으로는 학교 일정으로 인해 예정되어있던 계획대로 실행하기가 힘들었다는 점이 있습니다. 현실을 고려해서 실질적인 수행이 가능하도록 일정을 정했어야 했습니다.

또 다른 개선 사항으로는 현재 사용하고 있는 php 데이터베이스는 미흡한 점이 많다는 점입니다. 저희에게 적합한 DB를 찾는 일이 필요합니다.

데이터를 분석하기 위해 엑셀 파일로 변환하면 글자가 깨지는 현상이 발생합니다. 따라서 깨짐 현상이 없도록 하거나 데이터베이스를 직접 가져와서 분석하는 등 적절한 해결방안을 찾고 있는 중입니다.

나. 기타 보고사항

과제의 기대 효과로는 앞서 말했듯 금융 이력이 부족해서 불합리한 신용 평가를 받은 사람들에게 도움을 줄 수 있을 것으로 기대됩니다. 정확한 신용 등급을 측정하는 것은 무리지만 신용 평가에 사용 가능한 정보를 제공하는 것이 목표입니다.

기타 보고사항으로는 코드를 완성하고 제대로 동작하는 것까지 확인했는데도 불구하고 몇 달 뒤에는 해당 코드에서 오류가 생기는 일이 발생합니다. 저희가 활용한 SNS인 인스타그램에서 자체적으로 무언가 변경이 있었다고 생각했고 코드의 수정이 계속 필요하다고 결론 지었습니다.

4. 별첨