

[붙임4]

과제 유형	일반 캡스톤디자인					
<b>캡스톤디자인(종합설계) 과제 결과 보고서</b>						
개설학과	컴퓨터공학			학기	<input type="checkbox"/> 1학기 <input checked="" type="checkbox"/> 1·2학기 연계 <input type="checkbox"/> 2학기 <input type="checkbox"/> 하계 <input type="checkbox"/> 동계	
교과목명	캡스톤디자인			학점	6	
과제명	캡스톤디자인					
지도교수	소속(학과)	컴퓨터공학			성명	김영찬
구분	소속(학과)	학번	학년	성명	연락처	E-mail
참여학생 (총명)	컴퓨터공학	20157092	4	김상민	010-4922-3675	20157092@edu.hanbat.ac.kr
과제요약	모델에 데이터를 학습시켜 스팸 SMS와 정상 SMS를 분류하고 모델의 성능을 분석하여 SMS 분류에 있어 가장 적합한 모델이 무엇인지 알아본다. 사용자의 SMS 활용 목적에 따라 적합한 모델이 무엇인지 분석해본다.					
위와 같이 캡스톤디자인(종합설계) 과제를 수행하였으며, 결과보고서를 제출합니다. <div style="text-align: center;">2022. 12 . 2 .</div> <div style="text-align: right;">           지도교수 :        김영찬        (인)            대표학생 :        김상민        (인)         </div>						
<b>한밭대학교 산학협력교육원장 귀하</b>						

# 캡스톤디자인(종합설계) 과제 수행 보고서

## 1. 과제(주제) 선정 배경 및 필요성

수많은 정보가 오가는 정보화시대에 불필요한 정보를 담은 SMS, 피싱 SMS 등으로 인하여 금전적인 피해 및 개인정보 유출등의 피해를 보는 사람들이 늘어나고 있다. 이동통신사와 휴대폰 제조사와 협력해 제공하는 안드로이드 스마트폰의 휴대전화 스팸 간편 신고 기능과 스팸 모니터링 시스템(스팸트랩시스템)을 통해 수집된 스팸 차단 조치 등으로 휴대전화 스팸 신고는 최근 크게 늘어난 추세다. 지난 2017년 5308만 건에 달했지만 2018년 1434만 건까지 크게 줄어들었고, 2019년 1828만 건, 2626만 건, 3051만 건으로 다시 증가했다. 2020년 1월부터 5월까지 신고된 스팸 수는 1323만 건이다. 따라서 이러한 피해를 입는 것을 방지하기 위해 이메일을 필터링하여 스팸 SMS와 정상 SMS으로 분류할 필요가 있다.

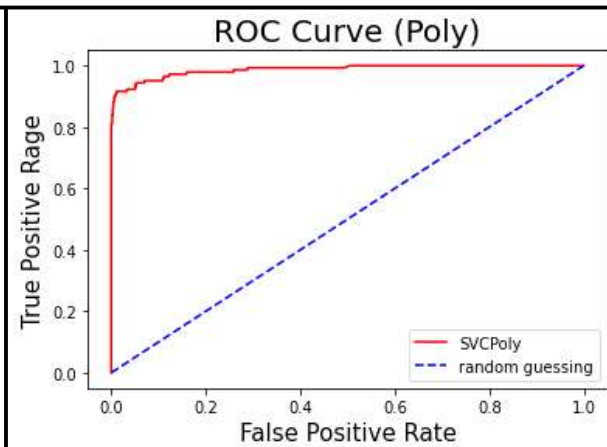
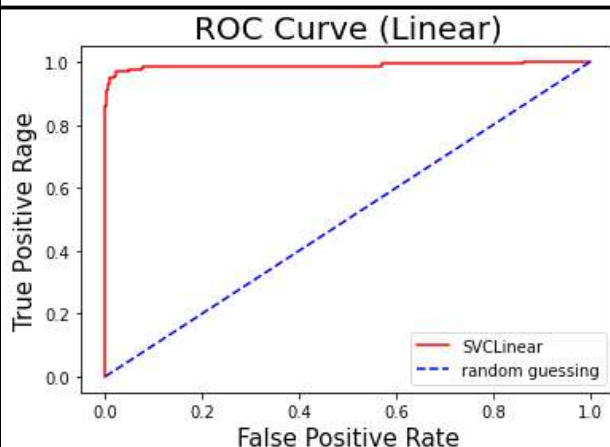
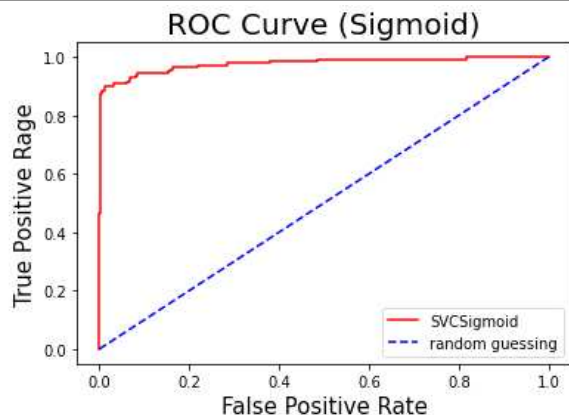
## 2. 과제(주제) 설계 내용 및 추진 방법

Kaggle의 SMS 데이터셋을 활용하여 SMS 데이터를 수집

공백, 숫자, 문장기호, 불용어(stop words), 단어 원형화, 소문자로 변환과 같은 데이터 전처리  
스팸, 정상 SMS 별로 자주 등장하는 단어들을 워드클라우드로 시각화

SVM(서포트 벡터 머신) 모델을 SVM모델의 경우 K-fold Cross Validation(k-fold 교차검증)을 활용하여 4가지 kernel로 정확도 측정

## 3. 과제(주제) 활동 결과 및 결과물(시제품)



## 5. 결과물의 활용방안(사업화 연계방안)

학습시킨 모델의 스팸 SMS 분류를 통한 피해 발생을 방지할 수 있다.

이러한 모델들을 활용하여 Kaggle 데이터 셋에 국한된 SMS 뿐만 아니라, 실제 현실에서 수신하는 SMS 및 이메일, 메신저들도 분류를 할 수 있도록 어플리케이션 제작

[붙임5]

# 캡스톤 디자인 II 최종결과 보고서

프로젝트 제목(국문): 기계학습 문자필터

프로젝트 제목(영문): MACHINE SMS FILTERING

프로젝트 팀(원): 학번: 20157092      이름: 김상민

프로젝트 팀(원): 학번:                      이름:

프로젝트 팀(원): 학번:                      이름:

지도교수: 김영찬

1. 중간보고서의 검토결과 심사위원의 '수정 및 개선 의견'과 그러한 검토의견을 반영하여 개선한 부분을 명시하시오.

없음

2. 기능, 성능 및 품질 요구사항을 충족하기 위해 본 개발 프로젝트에서 적용한 주요 알고리즘, 설계방법 등을 기술하시오.

Kernel(4가지) Linear, Poly, Sigmoid로 측정하여서 어떤 분류선이 분류하는데 있어 가장 좋은지 알아보았다.

Confusion Matrix를 활용하여 True Positive(TP), True Negative(TN), False Positive(FP), False Negative(FN)의 빈도를 알아보았다.

3. 요구사항 정의서에 명시된 기능 및 품질 요구사항에 대하여 최종 완료된 결과를 기술하시오.

기능

요구사항 고유번호		SFR-FA-001
요구사항 명칭		SMS 데이터 수집
요구사항 분류		기능
요구사항 상세설명	정의	Kaggle의 SMS 데이터셋을 활용하여 SMS 데이터 수집
	세부 내용	1. 스팸 또는 합법적인 것으로 태그가 지정된 SMS 메시지 수집 2. 수집한 데이터 파일 형태로 저장

요구사항 고유번호		SFR-FA-002
요구사항 명칭		데이터 전처리
요구사항 분류		기능
요구사항 상세설명	정의	데이터 분석을 위해 불필요한 칼럼 제거
	세부 내용	1. 형태소 분석 2. 불용어 처리

성능

요구사항 고유번호		PER-001		
요구사항 명칭		정확도 향상		
요구사항 분류		성능	응답 수준	필수
요구사항 상세설명	정의	데이터 모델 정확도 향상		
	세부 내용	모델의 정확도를 향상시켜 SMS 필터링		

## 인터페이스

요구사항 고유번호		SIR-001		
요구사항 명칭		시각화		
요구사항 분류		인터페이스	응답 수준	필수
요구사항 상세설명	정의	데이터 수집을 기반으로 시각화		
	세부내용	수집한 데이터를 기반으로 전처리 후 시각화		
주석		시각화를 통해 가독성 향상		
요구사항 출처				

4. 구현하지 못한 기능 요구사항이 있다면 그 이유와 해결방안을 기술하시오,

최초 요구사항	구현 여부(미구현, 수정, 삭제 등)	이유(일정부족, 프로젝트 관리미비, 팀원변동, 기술적 문제 등)
3가지 모델 구현	2가지 모델 미구현	늦게 시작으로 인한 일정부족

5. 요구사항을 충족시키지 못한 성능, 품질 요구사항이 있다면 그 이유와 해결방안을 기술하시오.

분류(성능, 속도 등) 및 최초 요구사항	충족 여부(현재 측정결과 제시)	이유(일정부족, 프로젝트 관리미비, 팀원변동, 기술적 문제 등)
3가지 모델 구현	한가지 모델 구현	일정부족

6. 최종 완성된 프로젝트 결과물(소프트웨어, 하드웨어 등)을 설치하여 사용하기 위한 사용자 매뉴얼을 작성하시오.

분석 실행

7. 캡스톤디자인 결과의 활용방안

학습시킨 모델의 스팸 SMS 분류를 통한 피해 발생을 방지할 수 있다.