


# 캡스톤 디자인 I

## 종합설계 프로젝트

프로젝트 명	<i>Security Monitoring Platform Using Event Logs (SeMo)</i>
팀 명	<i>Do Mo! (Do Monitoring)</i>
문서 제목	수행 계획서

Version	1.6
Date	2020-MAR-26

팀원	전 하훈 (조장)
	김 성은
	최 운호
	최 현인
	허 윤서

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

**CONFIDENTIALITY/SECURITY WARNING**

이 문서에 포함되어 있는 정보는 국민대학교 소프트웨어융합대학 소프트웨어학부 개설 교과목 캡스톤 디자인I 수강 학생 중 프로젝트 “SeMo”를 수행하는 팀 “Do Mo(Do Monitoring)”의 팀원들의 자산입니다. 국민대학교 소프트웨어학부 및 팀 “Do Mo(Do Monitoring)”의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

## 문서 정보 / 수정 내역

<b>Filename</b>	7조_수행계획서-SeMo.doc
<b>원안작성자</b>	전하훈, 김성은, 최운호, 최현인, 허윤서
<b>수정작업자</b>	전하훈, 김성은, 최운호, 최현인, 허윤서

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2020-02-28	전하훈	1.0	최초 작성	프로젝트 명, 팀 명 작성 역할 분담 작성
2020-03-02	김성은	1.1	내용 추가	프로젝트 개요 작성 개발 목표 초안 작성
2020-03-04	최현인	1.2	내용 추가	배경 기술 초안 작성
2020-03-06	허윤서	1.3	내용 추가	개발 내용 작성
2020-03-09	최운호	1.4	내용 수정	개발 목표 및 내용 수정
2020-03-13	최현인	1.5	내용 추가	개발 일정 초안 작성
2020-03-18	전원	1.5.1	내용 수정	연구/개발 수정 개발 일정 수정
2020-03-26	전하훈	1.6	최종	최종 수정 및 작성




계획서		
프로젝트 명	SeMo	
팀 명	Do Mo! (Do Monitoring)	
Confidential Restricted	Version 1.6	2020-MAR-26

## 목 차

<b>1</b>	<b>개요</b>	<b>5</b>
1.1	프로젝트 개요	5
1.2	추진 배경 및 필요성	5
1.2.1	추진 배경	5
1.2.2	현재 기술 시장 현황	8
1.2.3	현재 기술 시장 문제점	10
<b>2</b>	<b>개발 목표 및 내용</b>	<b>11</b>
2.1	목표	11
2.2	연구/개발 내용	11
2.2.1	개발 시나리오	11
2.2.2	연구/개발 방법	12
2.2.2.1	기계학습	12
2.2.2.2	ELK Stack 구축	13
2.2.2.3	웹	13
2.3	개발 결과	14
2.3.1	시스템 기능 요구사항	14
2.3.2	시스템 비기능(품질) 요구사항	15
2.3.3	시스템 구조	16
2.3.4	결과물 목록 및 상세 사양	17
2.4	기대효과 및 활용 방안	18
2.4.1	기대효과	18
2.4.2	활용방안	18
<b>3</b>	<b>배경 기술</b>	<b>19</b>
3.1	기술적 요구사항	19
3.2	현실적 제한요소 및 그 해결 방안	19
3.2.1	하드웨어	19
3.2.2	소프트웨어	19
<b>4</b>	<b>프로젝트 팀 구성 및 역할 분담</b>	<b>20</b>
<b>5</b>	<b>프로젝트 비용</b>	<b>21</b>

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

<b>6</b>	<b>개발 일정 및 자원 관리 .....</b>	<b>22</b>
6.1	개발 일정 .....	22
6.2	일정 별 주요 산출물 .....	23
6.3	인력자원 투입계획 .....	24
6.4	비 인적자원 투입계획 .....	25
<b>7</b>	<b>참고 문헌 .....</b>	<b>26</b>

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

# 1 개요

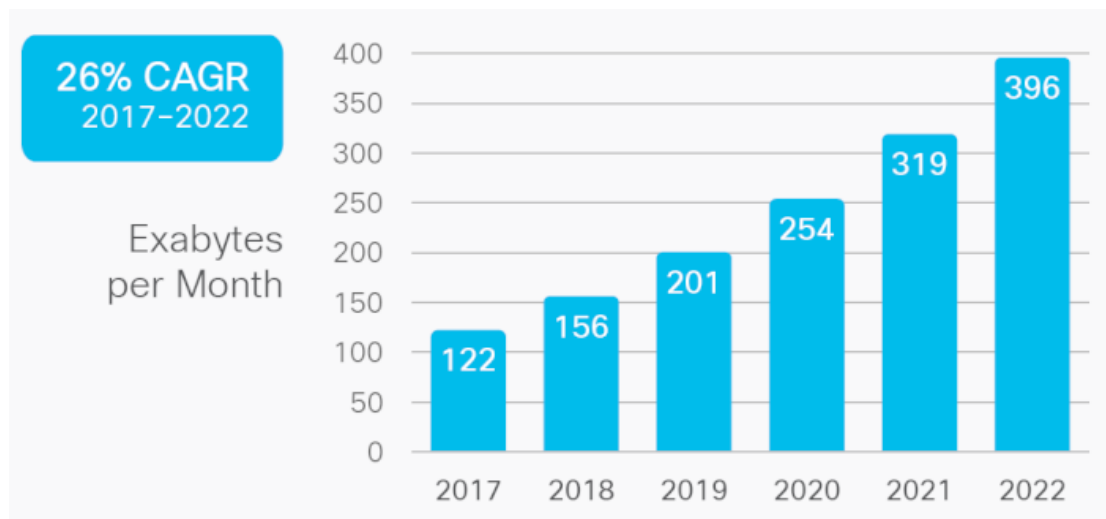
## 1.1 프로젝트 개요

정보화 시대를 맞아 네트워크 트래픽의 양이 방대해지면서 보안분야에서의 보안 관제의 역할이 더욱 중요해지고 있다. SeMo는 이러한 문제를 해결하고자 보안관제사들이 자동 처리 규칙들을 업데이트하는 데 도움을 주기 위해서 고안된 프로젝트이다. 수집된 로그데이터를 받아서 딥러닝을 통해 정탐과 오탐 여부를 판별한 후 분석하여 사용자에게 분석결과를 시각화 하여 웹을 통해 보여준다.

## 1.2 추진 배경 및 필요성

### 1.2.1 추진 배경

네트워크 장비회사 CISCO에 네트워크 트래픽 전망 조사에 따르면 앞으로 네트워크 트래픽 양이 방대해질 것이라고 전망했다[그림 1].



[그림 1] 전세계 월별 IP 트래픽 전망  
(출처 : Cisco VNI Global IP Traffic Forecast, 2017-2022)



이처럼 증가한 네트워크 트래픽 양으로 인해 보안의 중요성이 커지면서 보안관제에 대한 중요성도 커지고 있다. 과학기술정보통신부와 ‘한국 정보보호 산업 협회’가 발표한 ‘2019 국내 정보보호산업 실태 조사’에 따르면, 국내 정보보안 서비스 부문에서 보안관제 서비스 매출규모는 2018년대비 4.7%의 매출 규모 상승률을 보이고 있으며, 보안관제 서비스의 수출 성장세는 40.8%(220억원)를 기록했다.

[표 3-10] 정보보안산업 중분류 매출 현황

(단위 : 백만원, %)

구분		2018년	2019년(E)	증감율(%)
정보보안 시스템 개발 및 공급	네트워크보안 시스템 개발	729,393	771,656	5.8
	시스템보안 솔루션 개발	488,402	523,115	7.1
	정보유출방지 시스템 개발	426,128	456,251	7.1
	암호/인증 시스템 개발	151,879	161,760	6.5
	보안관리 시스템 개발	297,920	327,790	10.0
	소계	2,093,723	2,240,572	7.0
정보보안 관련 서비스	보안컨설팅 서비스	302,099	321,478	6.4
	보안시스템 유지관리/ 보안성 지속 서비스	351,942	359,645	2.2
	보안관제 서비스	273,927	286,880	4.7
	보안교육 및 훈련 서비스	1,740	2,990	71.8
	공인/사설 인증서	59,496	66,122	11.1
	소계	989,203	1,037,115	4.8
합계		3,082,926	3,277,687	6.3

[그림 2] 2019 국내 정보보호산업 실태조사 보고서

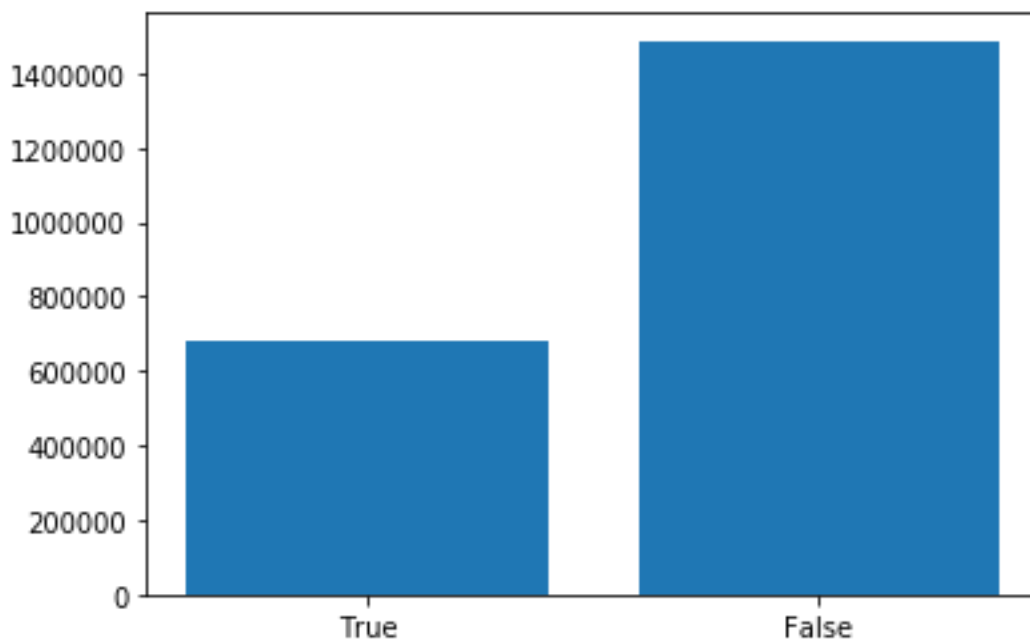
(출처: <https://www.dailyseclu.com/news/articleView.html?idxno=106428>)

현재, 대부분의 보안관제센터에는 오랜 기간 관제요원들의 업무 경험을 축약시켜서 만든 침입탐지 이벤트 자동 처리 규칙들(Signatures)이 침입 방지 시스템(Intrusion Prevention System) 장비에 집합되어 실행된다.


그러나 보안관제 서비스가 이와 같은 과정을 거침에도 불구하고 오탐(False Positive)이 지속적으로 발생하고 있다는 문제점이 존재한다. 이러한 오탐은 빅데이터 기반 보안관제에 큰 부담을 주고 있으며, 대량으로 발생하는 경보를 처리할 관제 인력의 부족과 경보 내 다수의 오탐은 보안관제사의 업무 효율성을 떨어뜨리는 요소가 되었다.



계획서		
프로젝트 명	SeMo	
팀 명	Do Mo! (Do Monitoring)	
Confidential Restricted	Version 1.6	2020-MAR-26



[그림 3]KISTI IPS Event Log 데이터 정탐/오탐 비율

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

## 1.2.2 현재 기술 시장 현황

### 1) SK 인포섹



[그림 4] SK infosec 로고

(출처 = <http://www.skinfosec.com/>)

SK 인포섹은 보유하고 있는 위협정보, 소속 보안전문가의 분석논리, 글로벌 보안기업과 공유하는 위협인텔리전스 등 양질의 정보를 학습하는 머신러닝 분석 알고리즘을 개발하여 보안관제 시스템 ‘시큐디움’에 적용시켰고, 국내외 2,000여 곳에 8,000대 이상의 보안 시스템에서 탐지한 이상징후를 수집한 후 다양한 분석규칙을 통해 공격 여부를 판별하고 있다. 또한 머신러닝을 활용해서 탐지 결과에 대한 효과 검증이 가능한 분야인 정탐과 오탐을 자동판정하는 모델을 우선 적용하였다. 그 결과, 머신러닝 적용 전보다 리소스를 70%나 감소시켰고 그만큼 줄어든 리소스를 위협 가능성이 높은 탐지이벤트 분석에 집중시켜 품질을 향상시켰으며, 공격 여부를 판별하는 사람의 노력과 시간을 매우 단축시켰다.

### 2) 이글루 시큐리티(IGLOO SECURITY)




[그림 5] IGLOO SECURITY 로고

(출처 = <http://www.igloosec.co.kr/index.do>)

이글루시큐리티는 한국 정보보안 업체 중 하나로, 종합보안관제 서비스를 제공하고 있다. 공공, 금융, 교육, 기업, 해외를 대상으로 솔루션사업과 보안관제 서비스를 제공하고 있으며, 보안컨설팅 사업도 하고 있다. 이글루 시큐리티의 SPIDER TM AI Edition은 ELK Stack을 사용하여 빅데이터를 기반으로 위협 상황의 실시간 식별 및 분석을 수행하고, 머신러닝 기반의 지도 학습과 비지도 학습을 통해 관제의 효율성을 극대화했다. 또한 보안진단 자동화 솔루션과의 연동으로 사이버 위협 요인을 사전에 해소하고 사이버 위협 정보 공유 시스템을 통해 글로벌 위협 인텔리전스 정보를 제공한다.



 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

### 3) 로그프레스(LOGPRESSO)



[그림 6] LOGPRESSO 로고

(출처 = <https://ko.logpresso.com/>)

로그프레스는 빅데이터 기반의 통합보안관제와 AI기반의 이상거래차단, 빅데이터기반의 통합 로그 관리 등을 제공하는 업체이다. 로그프레스의 통합보안관제는 실시간 이벤트 연관분석(CEP)와 배치연관분석을 지원하고 있고 비지도 학습 머신러닝 모델을 통해 이상징후를 탐지한다. 또한 드래그&드롭을 통해 분석을 간편화했으며 IP가시성과 CTI연계를 지원한다.

### 4) 시만텍(Symantec)

시만텍은 미국의 보안소프트웨어 회사로, 시만텍의 MSS-ATP는 네트워크 및 엔드포인트에서 탐지된 내용을 자동으로 비교분석하여 고객사의 환경을 위협하는 요소들을 탐지해 우선적으로 처리하고, 자동 역추적을 통해 공격받은 실제 엔드포인트를 파악한다. 또한 엔드포인트 보안기술과 네트워크 기반의 지능적 보안 위협 탐지 기술을 양방향으로 통합해 신속한 대응이 가능하며 지속적으로 정보를 수집하는 시만텍의 GIN(Global Intelligence Network)과 평판 기반 보안기술인 Insight를 사용해 잠재적 악성파일을 평가하면서 보안 위협 조사의 효율을 높였다.

보안 관제 서비스: 지능적 보안 위협 차단



[그림 7] 시만텍 보안관제 서비스 특징

(출처 = [http://143.127.10.117/content/ko/kr/enterprise/fact\\_sheets/b-overview-solution-mss-advanced-threat-protection-21332713\\_KR.pdf](http://143.127.10.117/content/ko/kr/enterprise/fact_sheets/b-overview-solution-mss-advanced-threat-protection-21332713_KR.pdf))



 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

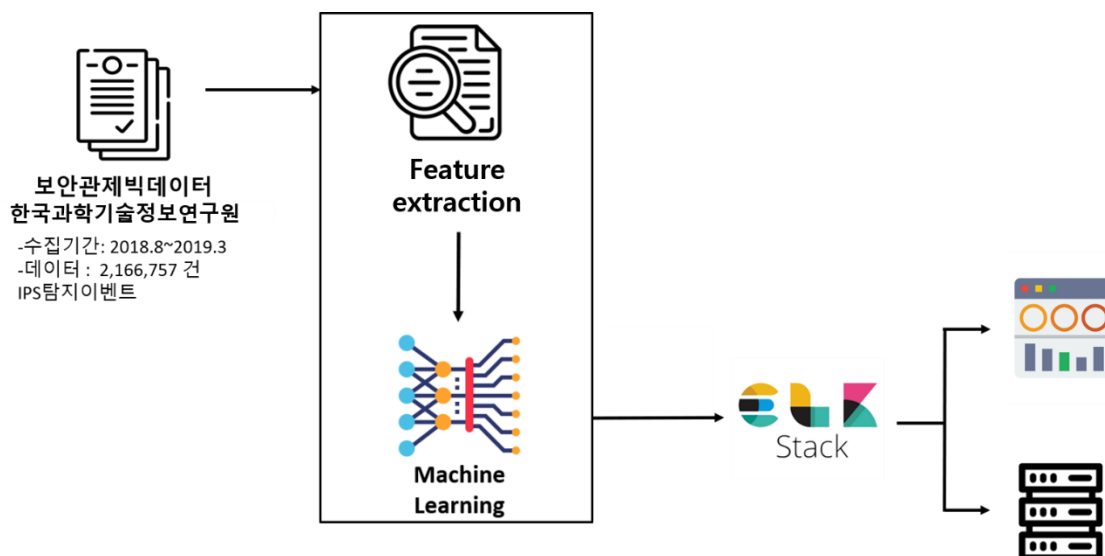
## 2 개발 목표 및 내용

### 2.1 목표

이 프로젝트는 수집된 로그데이터를 받아서 두 가지 일을 처리한다. 우선 딥러닝을 통해 학습된 모델로 정오탐 판별을 하고 한 쪽에서는 ELK Stack을 구축하여 로그데이터를 저장한다. 이후 정탐 데이터를 분석하여 사용자에게 분석결과를 보여준다. 이때 효율적인 시각화 도구인 Kibana 대시보드를 이용한다. Kibana의 대시보드를 웹과 연동시켜 사용자에게 보여지게 된다.

### 2.2 연구/개발 내용

#### 2.2.1 개발 시나리오




[그림 9] 개발 시나리오

본 프로젝트에서는 ELK Stack을 활용한다. IPS 장비로부터 나온 이벤트는 총 두가지 관계를 갖게 된다.

1. 시각화를 위한 ELK Stack PUSH
2. 정오탐 판별을 위한 데이터 Preprocessing

#### 2-1) Deep Learning 모델을 위한 Preprocessing

딥러닝 모델에서 데이터 정오탐 판별을 위해 이벤트 로그에서 Payload 필드만 가져와서 전처리한다. 전처리된 데이터는 딥러닝 모델을 거치게 되고 모델은 floating vector를 출력하게 된다.

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

## 2-2) Machine Learning 모델을 위한 Preprocessing

이벤트 로그에서 Payload와 무의미한 필드를 제외하여 전처리를 한다. 전처리 된 데이터는 기계학습을 거치게 되고 모델은 predicted Vector를 출력하게 된다.

## 2-3) Ensemble 모델을 위한 Preprocessing

딥러닝 모델에서 나온 floating vector와 기계학습 모델에서 나온 predicted vector는 다시 한번 전처리(concated) 되어 앙상블 모델에 들어가게 된다. 데이터를 입력 받은 앙상블 모델은 해당 데이터가 정탐인지 오탐인지 정오탐 판별을 하게 된다.

정오탐 판별이 된 데이터 중 정탐인 데이터에 한하여 ELK Stack에 UPDATE 시킨 후 해당 이벤트와 관련된 데이터를 Kibana 통해 웹에서 시각화 하여 보여지게 된다.

## 2.2.2 연구/개발 방법

### 2.2.2.1 기계 학습

#### (a) 데이터 설명

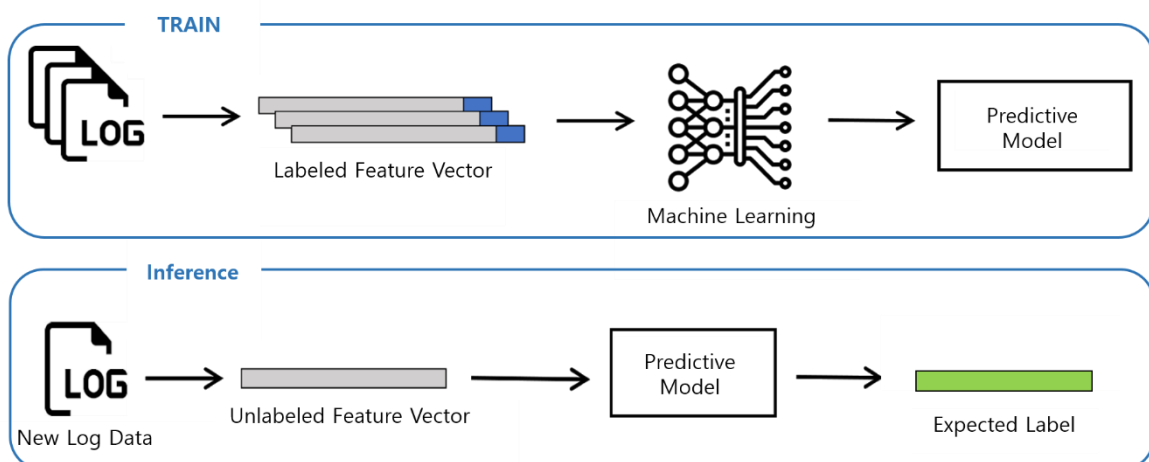
KISTI 사이버 안전센터에서 IPS 로그로부터 2018.08 ~ 2019.03 수집한 이벤트 로그 빅데이터 2,166,757개를 사용한다. 해당 데이터는 보안관제사들이 수작업으로 정오탐을 판별한 라벨이 있다.

#### (b) 전처리

IPS 장비 특성상 대상 기관에 최적화되어 규칙이 적용되고, IPS 장비가 다양하게 있으므로 대다수의 IPS장비에서 적용할 수 있는 일반적인(General) 필드들 예를 들면, Payload, Source IP, Destination IP 등을 딥러닝을 통해 학습한다.

#### (c) 기계 학습

기계 학습으로는 딥러닝을 사용한다. 딥러닝으로 학습된 모델을 바탕으로 새로운 로그 데이터에 대해 예측 과정을 거친 후 라벨링을 한다. 해당 내용을 바탕으로 다양한 딥러닝 신경망과 기계학습 모델을 사용하여 실험을 진행할 예정이다. 다양한 모델을 실험한 후 성능이 좋은 모델을 사용하거나, 앙상블 모델을 사용할 예정이다.



[그림 10] 기계 학습

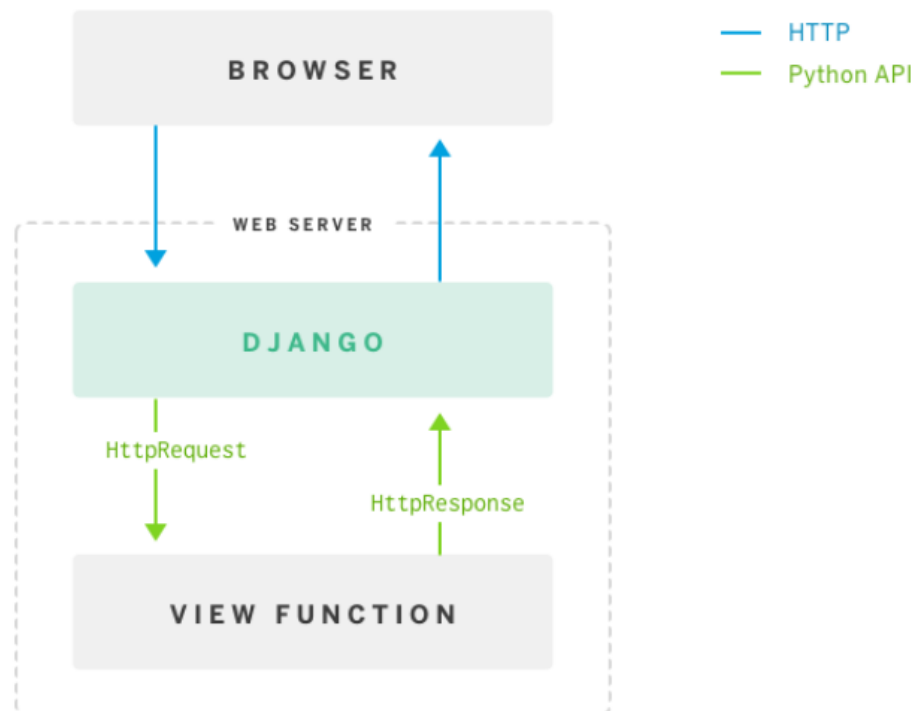
 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

### 2.2.2.2 ELK Stack 구축

학습된 모델로 판별한 데이터를 시각화하기 위해 ELK를 구축한다. Logstash는 로그 파일로부터 지속적으로 입력을 받아서 Elasticsearch 서버에 데이터를 Push한다. 모델을 거치고 나온 데이터는 판별값으로 라벨링이 되어 나온다. 이 때, 판별값이 정답인 데이터의 'id' 컬럼을 매칭 인덱스로 사용하여 Elasticsearch 데이터에 라벨링 컬럼을 추가한다. 최종적으로 악성 로그를 분석하기 위해 Kibana는 정답으로 판별된 데이터만 시각화한다.

### 2.2.2.3 웹

웹은 재학습, 분석, 시각화 기능을 제공하는 플랫폼을 Django를 기반으로 구현한다. ELK로부터 얻어진 데이터는 로컬 환경에 저장되고, 저장된 파일은 딥러닝을 통해 정오답 판별과 분석이 된다. 분석 결과는 브라우저에 나타난다.



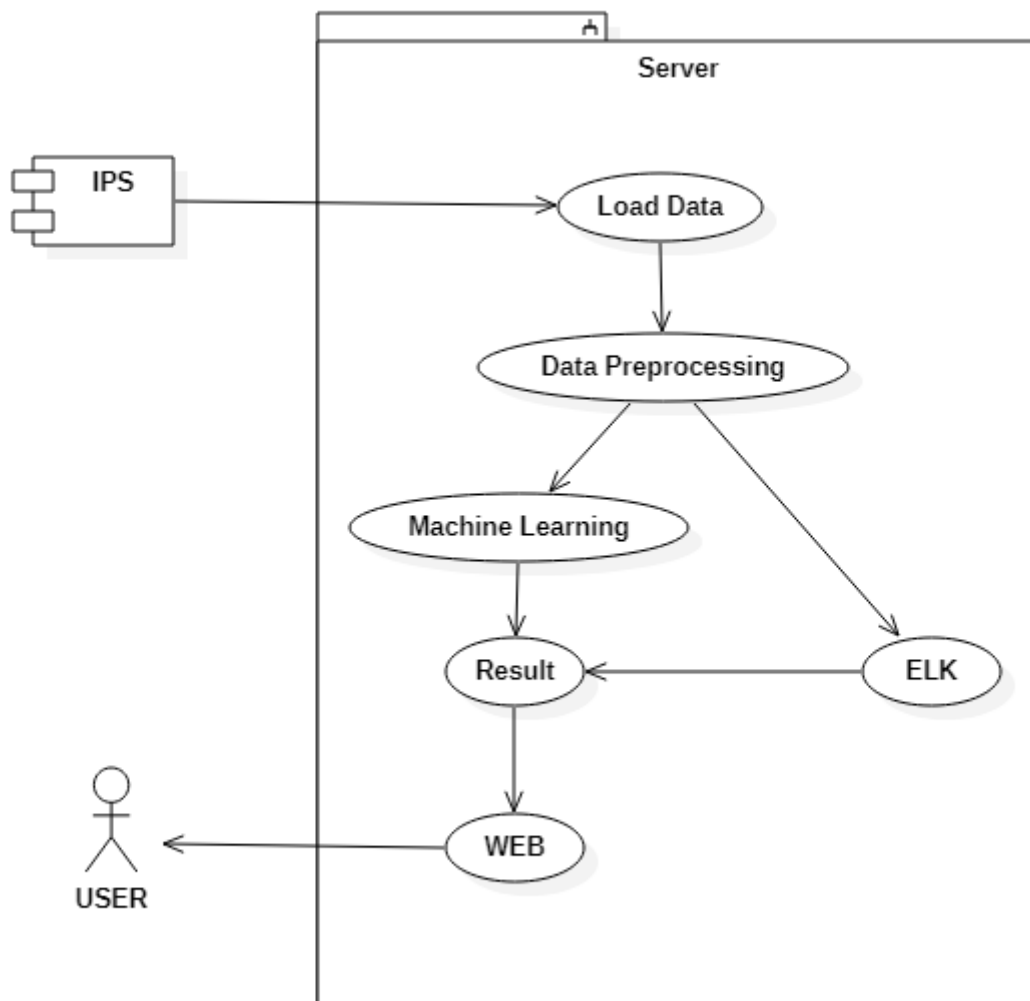
[그림 11] 웹 서비스 구조  
(출처 = <https://vinay13.github.io/>)



계획서		
프로젝트 명	SeMo	
팀 명	Do Mo! (Do Monitoring)	
Confidential Restricted	Version 1.6	2020-MAR-26

## 2.3 개발 결과

### 2.3.1 시스템 기능 요구사항



[그림 12] 유즈 케이스(Use Case)

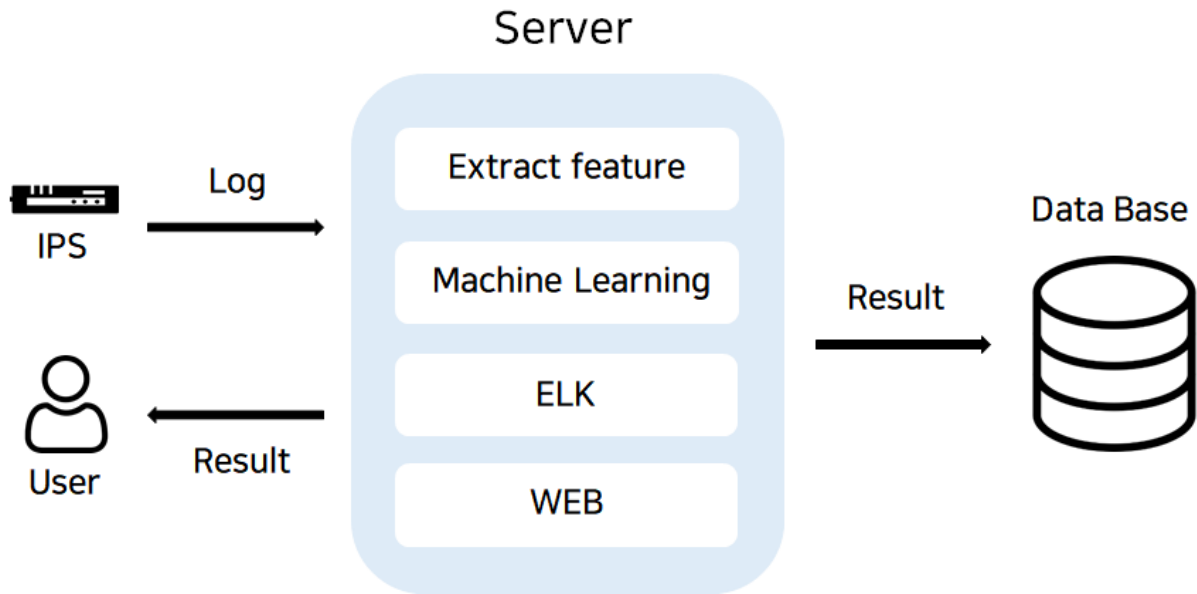
 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

### 2.3.2 시스템 비기능(품질) 요구사항

비기능적 요구사항		
제품요구사항	구현성	- Python 사용, ELK Stack 활용
	보안성	- 데이터는 사용자만 접근 가능하다.
	유지보수성	- 깃허브를 활용하여 소프트웨어의 수정, 개선을 용이하게 한다.
	이식성	- 시스템은 데스크 탑, 랩 탑 디바이스에서 작동하고, 모바일 디바이스에서는 작동되지 않는다.
	상호운용성	- 웹은 Django를 활용, 시각화&데이터는 ELK Stack을 활용한다.
	가용성	- 서비스는 하루 24시간 제공된다.
조직요구사항	배포 요구사항	<ul style="list-style-type: none"> <li>- 온라인 서비스(오픈소스)로 가이드라인을 제공한다.</li> <li>- 해당 프로젝트는 계속 서비스하며 수정사항을 제안받거나 오류가 발생시 버전을 높혀서 재배포한다.</li> </ul>
	구현 요구 사항	<ul style="list-style-type: none"> <li>- 개발도구로 Github를 사용한다.</li> <li>- 서버의 구현위치는 Django+ELK Stack을 이용한다.</li> </ul>
안전요구사항	안정성	- DBMS는 sqlite3와 Elastic search를 사용한다.

 국민대학교 소프트웨어학부 캡스톤 디자인 I	계획서		
	프로젝트 명	SeMo	
	팀 명	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26


### 2.3.3 시스템 구조



[그림 13] 시스템 구조

웹 서비스는 파일 업로드가 가능한 형태로 구현한다. 웹 브라우저를 통해 로그 데이터 파일을 업로드 하고, 업로드 된 파일은 두 가지 처리 과정을 거친다. 우선 딥러닝을 통해 학습된 모델로 해당 파일에 대해 정탐과 오탐을 판별하고, 한 쪽에서는 ELK Stack을 구축하여 로그데이터를 저장한다. 이후 효율적인 시각화 도구인 Kibana 대시보드를 이용해 클라이언트에게 정탐 분석결과를 보여준다.



 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

### 2.3.4 결과물 목록 및 상세 사양

대분류	소분류	기능	형식	비고
웹서비스	파일 업로드	현재 열린 파일을 저장한다.	DLL/함수	
	파일 다운로드	다른 이름으로 파일을 저장한다	DLL/함수	
	대시 보드	문서 파일을 연다.	모듈	
	데이터베이스 (로그인) [추가목표]			
정오탐 분류 기	머신러닝	정오탐 분류	모델	
	딥러닝	정오탐 분류	모델	
	앙상블 모델	정오탐 분류	모델	

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

## 2.4 기대효과 및 활용방안

### 2.4.1 기대효과

네트워크 트래픽의 양이 방대해지면서 보안분야에서는 보안 관제의 역할이 더욱 중요해지고 있다. 본 프로젝트는 보안관제사들의 자동 처리 규칙 업데이트를 도와주기 위해 고안됐다. 이 플랫폼으로 인해 보안관제사들은 효율적으로 트리거를 업데이트 할 수 있다. 뿐만 아니라 수작업으로 이루어지던 정오탐 판별의 시간적 낭비가 사라진다.


우선적으로, 매일 새롭게 생성되는 방대한 보안 이벤트 분석을 자동화함으로써, 보안 업무의 효율성을 높일 수 있다. 우선 처리해야 할 고위험 이벤트를 선별함으로써 방대한 보안 이벤트 분석에서 소요되는 시간을 단축하여 보다 빠른 대응이 가능해진다. AI 알고리즘에 적용할 학습 데이터를 생성하고 AI 시스템에 내린 결과에 피드백을 주는 과정을 반복함으로써 예측의 정확성을 끌어올릴 수 있다. 예를 들어, 기존 상태에서 약간의 변화만 있는 신·변종 악성코드의 경우, 이전에는 보안 관리자가 일일이 보안 이벤트를 분석해 처리해야 했다. 하지만, 방대한 데이터를 학습한 AI 시스템을 구축해 단순한 공격은 자동 처리하게 하고 보안 담당자는 고위험군 이벤트 분석에 집중한다면, 이벤트 처리 효율성을 비약적으로 높일 수 있게 될 것이다.

더불어, 장기간 축적된 보안 데이터를 AI 알고리즘을 통해 분석함으로써, 날로 진화하는 고도화된 보안 위협을 보다 빠르고 정확하게 탐지하고 위협 대응에 소요되는 시간을 단축시킬 수 있다. 공격자들이 장기간에 걸쳐 기업 내부 시스템들을 교묘히 옮겨 다니며 공격하는 만큼, 단기간 수집된 보안 데이터 분석만으로는 공격자의 행위를 정확하고 빠르게 탐지하기 어려웠던 것이 사실이다.

머신러닝 기반의 AI 알고리즘을 통해 보안 데이터, 최신 위협 정보, 취약점 등 관련된 정보를 신속하게 연관 분석함으로써, 기업 전반을 아우르는 폭넓은 가시성을 확보할 수 있게 된다. 또한 악의적 행위·공격자 특성 등이 담긴 양질의 학습 데이터에 대한 비지도 학습을 통해 심각한 위협으로 발전할 수 있는 알려지지 않은 변칙 활동 및 이상행위를 보다 빨리 식별할 수 있게 된다.

### 2.4.2 활용방안

오픈 소스로 배포하여 보안관제 플랫폼 가이드라인을 제시함으로써 기존의 자체적인 보안관제 시스템이 구축되어 있지 않은 회사들은 자체적인 보안관제 시스템을 구축하여 활용할 수 있다. 또한 인공지능 보안관제 기술을 개발함으로써 정탐과 오탐의 자동 분석이 가능해짐으로써 보안관제사들의 업무 능력을 향상시킬 수 있다.

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

## 3 배경 기술

### 3.1 기술적 요구사항

- ❑ 개발 언어
  - Python 3.6.5
- ❑ 개발 환경
  - CPU : Intel(R) Core(TM) i7-6850K CPU @ 3.60GHz 3.60GHz
  - GPU : NVIDIA GeForce GTX 1080 Ti
  - Storage : Samsung SSD 850 Pro 1TB
  - RAM : Samsung DDR4 128GB
- ❑ 라이브러리
  - Django 3.0.4
  - Logstash 7.6.1
  - Elasticsearch 7.6.1
  - Kibana 7.6.1
  - TensorFlow 2.1.0
  - numpy 1.14.5
  - Pandas 0.23.4
  - Pytorch 1.0.1
- ❑ 데이터
  - KISTI IPS 이벤트 로그 데이터 (2018.08 ~ 2019.03)

### 3.2 현실적 제한 요소 및 그 해결 방안

#### 3.2.1 하드웨어

웹 서버에 엔진이 정상적으로 확장될 수 있는 서버 사양이 되어야 한다. 기계 학습 모델을 학습할 때 많은 양의 데이터를 학습할 경우 오랜 시간이 소요된다. 이는 GPU를 사용함으로써 모델이 학습되는데 걸리는 시간을 단축할 수 있다.

#### 3.2.2 소프트웨어


실시간 로그분석이 되는 것이 주된 목적이므로 웹 서버와 엔진 간의 통신이 원활하게 되도록 해야 한다.

IPS 장비마다 로그 파일 형식이 다르기 때문에 범용적인 소프트웨어를 만들어야 한다. 따라서 로우 데이터에서 수집 가능한 가장 general한 feature를 추출한다.

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26


## 4 프로젝트 팀 구성 및 역할 분담

이름	역할
전하훈	<ul style="list-style-type: none"> <li>- Software Project Leader</li> <li>- Deep Learning 모델 설계 및 구축</li> <li>- Deep Learning 모델 파라미터 튜닝</li> <li>- ELK 구축</li> <li>- 앙상블 모델 구현</li> <li>- 데이터 전처리</li> <li>- Deep Learning 재학습 구현</li> </ul>
김성은	<ul style="list-style-type: none"> <li>- 앙상블 모델 구현</li> <li>- Deep Learning 모델 파라미터 튜닝</li> <li>- Deep Learning 재학습 구현</li> </ul>
최운호	<ul style="list-style-type: none"> <li>- 웹 서버 구축 및 관리</li> <li>- 웹 서버와 ELK 연동</li> <li>- 웹 서버와 모델 연동</li> <li>- 웹 안정화</li> </ul>
최현인	<ul style="list-style-type: none"> <li>- 웹 구축 및 안정화</li> <li>- 웹 프론트 제작</li> <li>- ELK 구축</li> <li>- 문서 관리</li> </ul>
허윤서	<ul style="list-style-type: none"> <li>- 웹 프론트 제작</li> <li>- 웹 서버 관리</li> <li>- 웹 서버와 ELK 연동</li> <li>- 웹 안정화</li> </ul>

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

## 5 프로젝트 비용

항목	예상치 (MD)
ELK 환경 구축	15
웹 환경 구축 및 연동	40
웹 - ELK - 모델 연동	15
웹 서비스 시각화	15
분석 환경 구축	20
학습 모델 구성	40
재학습 환경 구축	25
프로젝트 테스트 및 유지보수	20
프로젝트 평가 및 보고서 작성	15
합	205

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

## 6 개발 일정 및 자원 관리


### 6.1 개발 일정

항목	세부내용	1월	2월	3월	4월	5월	6월	비고
요구사항분석	요구 분석							
	정보 수집							
관련분야연구	주요 기술 연구							
	관련 시스템 분석							
설계	시스템 설계							
구현	코딩 및 모듈 테스트							
테스트	시스템 테스트							

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

## 6.2 일정별 주요 산출물


마일스톤	개요	시작일	종료일
계획서 발표	개발 환경 완성 (GCC 설치, 기본 응용 작성 및 테스트 완료) <b>산출물 :</b> 1. README.md 2-1. 계획서 발표 슬라이드 쇼 2-2. 계획서 발표 슬라이드 화일 2-3. 수행 계획서 화일	2020-01-10	2020-03-27
설계 완료	시스템 설계 완료 <b>산출물 :</b> 1. 시스템 설계 사양서	2020-03-27	2020-04-10
중간 보고	서버구축 및 인공지능 모델링 <b>산출물 :</b> 1. 프로젝트 중간 보고서 2. 프로젝트 진도 점검표 3. 구현 소스 코드	2020-04-10	2020-04-23
구현 완료	시스템 구현 완료 <b>산출물:</b> 보안관제 분석 플랫폼	2020-04-23	2020-05-01
테스트	시스템 통합 테스트 <b>산출물:</b> 1. 최종버전 시스템 2. 웹페이지	2020-06-02	2020-06-11
최종보고서	최종 보고 <b>산출물:</b> 1. 전시용 자료(포스터 및 소개책자) 2. 온라인 평가용 자료 3. 최종결과 보고서	2020-06-12	2020-06-19

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

### 6.3 인력자원 투입계획


이름	개발항목	시작일	종료일	총개발일(MD)
전원	Project Study	2020-01-10	2020-01-30	20
전원	프로젝트 절차 구성	2020-02-01	2020-02-06	5
전하훈	딥러닝 모델 설계 딥러닝 모델 구현 딥러닝 모델 파라미터 튜닝 ELK 구축 앙상블 모델 구현 데이터 전처리 딥러닝 재학습 구현	2020-03-02	2020-06-01	80
김성은	딥러닝 모델 파라미터 튜닝 앙상블 모델 구현 데이터 전처리 딥러닝 재학습 구현	2020-03-02	2020-06-01	80
최운호	웹 구축 및 관리 웹 서버와 ELK 연동 웹 서버와 모델 연동 웹 안정화	2020-03-02	2020-06-01	80
최현인	웹 구축 및 안정화 웹 프론트 제작 ELK 구축 문서 관리	2020-03-02	2020-06-01	80
허윤서	웹 프론트 제작 웹 서버 관리 웹 서버와 ELK 연동 웹 안정화	2020-03-02	2020-06-01	80



 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26


## 6.4 비 인적자원 투입계획

항목	Provider	시작일	종료일	Required Options
개발용 PC 1대	조립 PC	2020-01-10	2020-06-12	
서버용 PC 1대	조립 PC	2020-03-02	2020-06-12	
개발용 개인 노트북 5대	Lenovo*3, Samsung*2	2020-03-02	2020-06-12	

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

## 7 참고 문헌

번호	종류	제목	출처	발행년도	저자	기타
1	논문	ADELE: Anomaly Detection from Event Log Empiricism	<a href="https://ieeexplore.ieee.org/document/8486257">https://ieeexplore.ieee.org/document/8486257</a>	2018		
2	논문	DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning	<a href="https://dl.acm.org/doi/10.1145/3133956.3134015">https://dl.acm.org/doi/10.1145/3133956.3134015</a>	2017		
3	논문	ATTACK2VEC: Leveraging Temporal Word Embeddings to Understand the Evolution of Cyberattacks	<a href="https://arxiv.org/abs/1905.12590">https://arxiv.org/abs/1905.12590</a>	2019		
4	논문	Tiresias: Predicting Security Events Through Deep Learning	<a href="https://dl.acm.org/doi/10.1145/3243734.3243811">https://dl.acm.org/doi/10.1145/3243734.3243811</a>	2018		
5	기사	2019국내 정보보안 산업매출 3조 2천700억원...수출액은 1천80억원 기록	<a href="https://www.dailysecu.com/news/articleView.html?idxno=106428">https://www.dailysecu.com/news/articleView.html?idxno=106428</a>	2020		
6	보고서	이글루시큐리티 SPiDER TM AI Edition - AI 기반 보안관제시스템 기대 효과	<a href="http://www.igloosec.co.kr/brochure/kr/SPiDER_TM_AI_Edition(kr).pdf">http://www.igloosec.co.kr/brochure/kr/SPiDER_TM_AI_Edition(kr).pdf</a>	2019		
7	참고 사이트	직무인터뷰-sk인포섹	<a href="https://blog.naver.com/skinfosec2000?blogId=skinfosec2000&amp;logNo=221407240489&amp;parentCategoryNo=&amp;categoryNo=&amp;viewDate=&amp;isShowPopularPosts=false&amp;from=postView">https://blog.naver.com/skinfosec2000?blogId=skinfosec2000&amp;logNo=221407240489&amp;parentCategoryNo=&amp;categoryNo=&amp;viewDate=&amp;isShowPopularPosts=false&amp;from=postView</a>	2018		
8	참고 사이트	다가오는 인공지능 기반의 보안관제, 그 전에 준비해야할 것은?	<a href="http://www.igloosec.co.kr/BLOG_%EB%8B%A4%EB%A0%80%EC%98%A4%EB%8A%94%20%EC%">http://www.igloosec.co.kr/BLOG_%EB%8B%A4%EB%A0%80%EC%98%A4%EB%8A%94%20%EC%</a>	2019		

 <b>국민대학교</b> <b>소프트웨어학부</b> <b>캡스톤 디자인 I</b>	<b>계획서</b>		
	<b>프로젝트 명</b>	SeMo	
	<b>팀 명</b>	Do Mo! (Do Monitoring)	
	Confidential Restricted	Version 1.6	2020-MAR-26

			<a href="#">9D%B8%EA%B3%B5%EC%A7%80%EB%8A%A5%20%EA%B8%B0%EB%B0%98%EC%9D%98%20%EB%B3%B4%EC%95%88%EA%B4%80%EC%A0%9C,%20%EA%B7%B8%20%EC%A0%84%EC%97%90%20%EC%A4%80%EB%B9%84%ED%95%B4%EC%95%BC%20%ED%95%A0%20%EA%B2%83%EC%9D%80[qs]?searchItem=&amp;searchWord=&amp;bsCatelId=1&amp;gotoPage=1</a>			
9	기사	지난해 국내 정보보안·물리보안 산업매출 규모, 10조5000억원	<a href="https://byline.network/2020/02/5-58/">https://byline.network/2020/02/5-58/</a>	2020		
10	논문	CNN and RNN based payload classification methods for attack detection	<a href="https://www.sciencedirect.com/science/article/pii/S0950705118304325#b14">https://www.sciencedirect.com/science/article/pii/S0950705118304325#b14</a>	2018		