



# DREAM

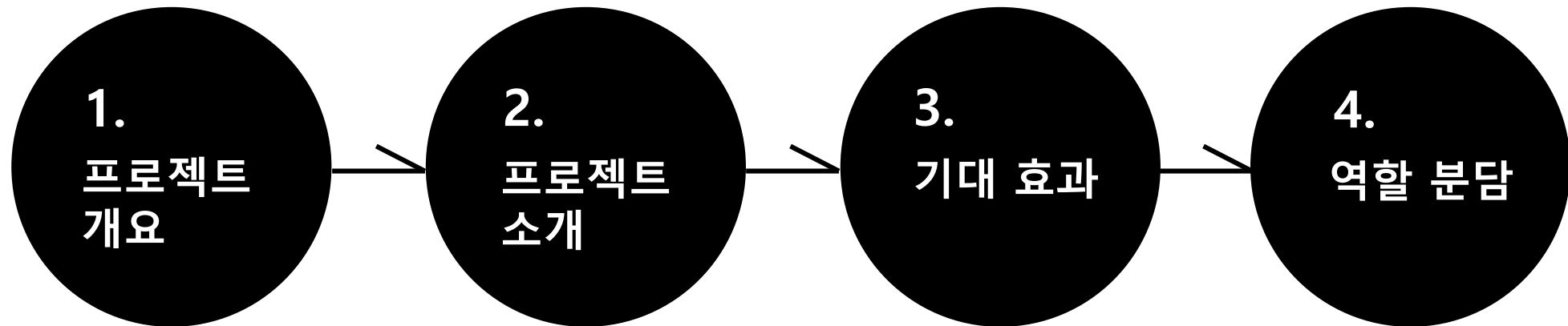
Do it !



문다민 김기환 김현석 정혜리 방유한



# — 목차



# 1. 프로젝트 개요

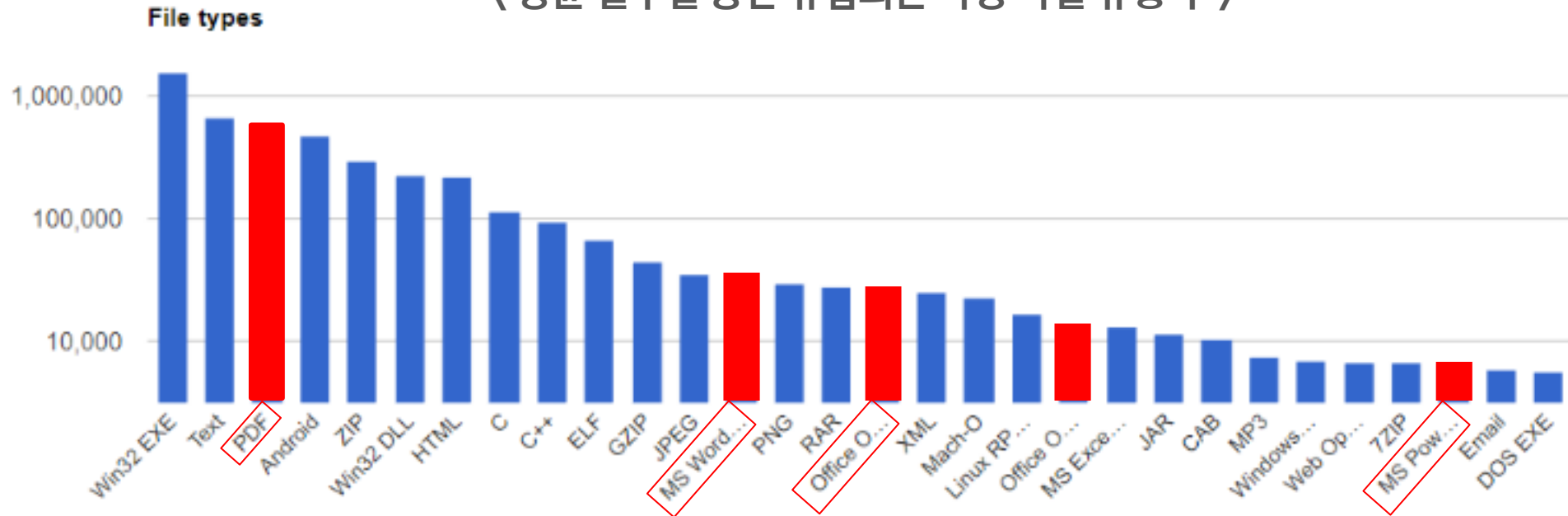
지능적으로 발전하는 악성코드, 이를 이용한 사이버 공격



# 1. 프로젝트 개요

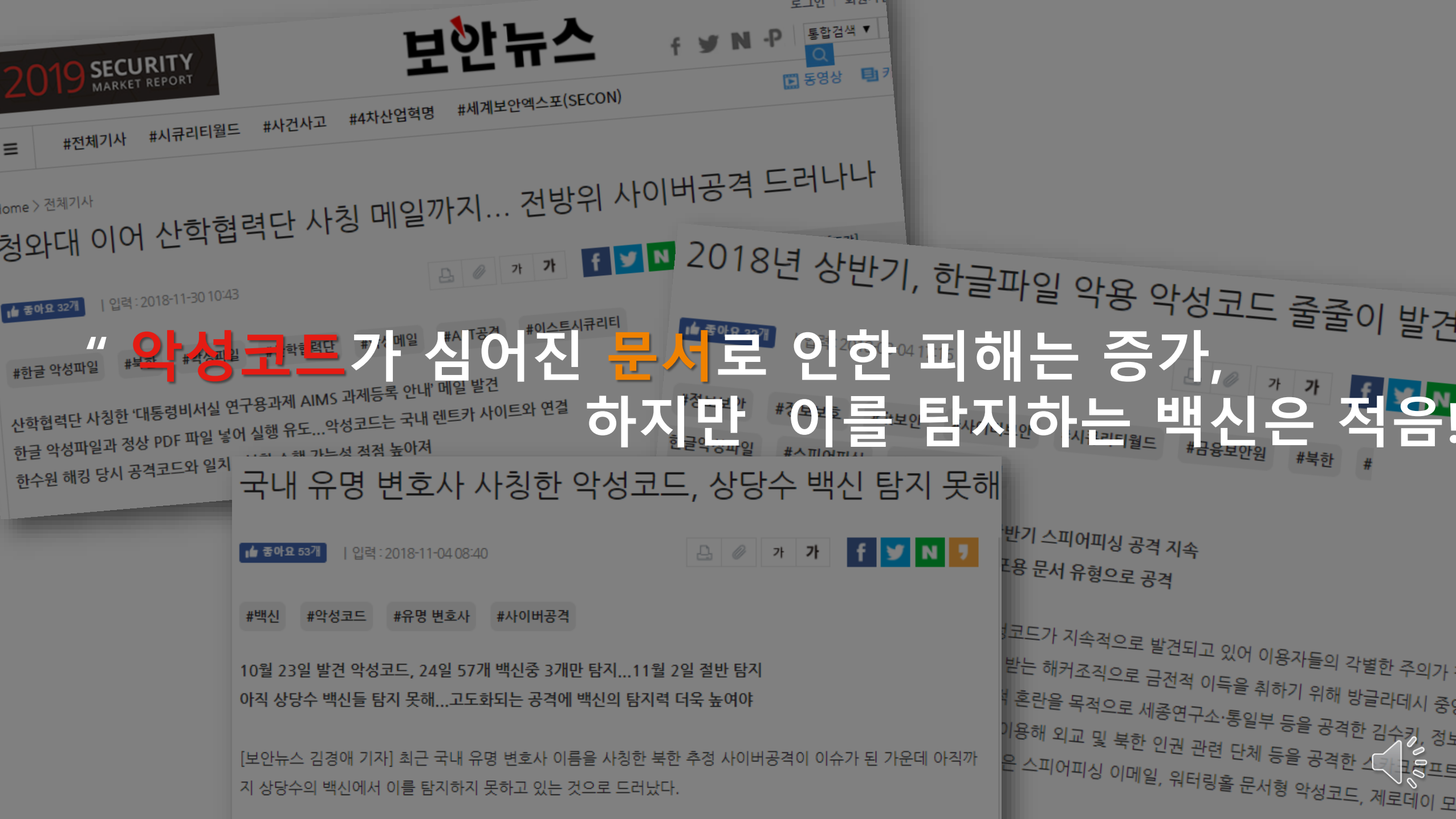
그 중의 상당 수는 우리가 자주 접하는 PDF, MS Office 등 문서에 숨겨진 악성코드

〈 평균 일주일 동안 유입되는 악성 파일 유형 수 〉



출처 : <https://www.virustotal.com/ko/statistics/>





# “악성코드가 심어진 문서로 인한 피해는 증가, 하지만 이를 탐지하는 백신은 적음!”

## 국내 유명 변호사 사칭한 악성코드, 상당수 백신 탐지 못해

좋아요 53개 | 입력: 2018-11-04 08:40

#백신 #악성코드 #유명 변호사 #사이버공격

10월 23일 발견 악성코드, 24일 57개 백신중 3개만 탐지...11월 2일 절반 탐지  
아직 상당수 백신들 탐지 못해...고도화되는 공격에 백신의 탐지력 더욱 높아야

[보안뉴스 김경애 기자] 최근 국내 유명 변호사 이름을 사칭한 북한 추정 사이버공격이 이슈가 된 가운데 아직까  
지 상당수의 백신에서 이를 탐지하지 못하고 있는 것으로 드러났다.



## 2. 프로젝트 소개

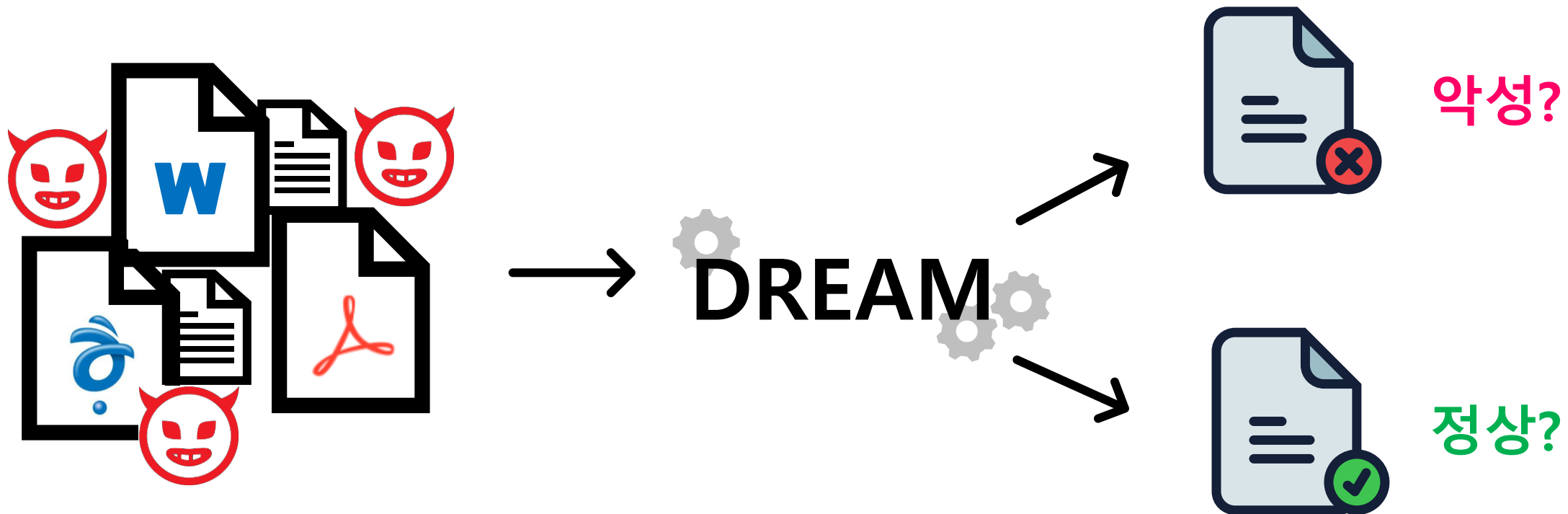
프로젝트 이름

**D**etecting  
in **R**eal – tim**E**  
**m****A**licious documents  
using **M**achine learning



## 2. 프로젝트 소개

프로젝트 목표



## 2. 프로젝트 소개

### 개발 내용



Dictionary Object << >> - contains the attribute of the object

Indirect Object ID - This gives the object unique object identifier by which other object can refer to it.

Stream Filter - Indicates how the data in the stream must be decoded.

**PDF**

**Header**

- Version Number

**Body**

- Page objects
- Image objects
- Font objects
- Bookmark objects
- Forms objects
- ...

**Cross-reference Table**

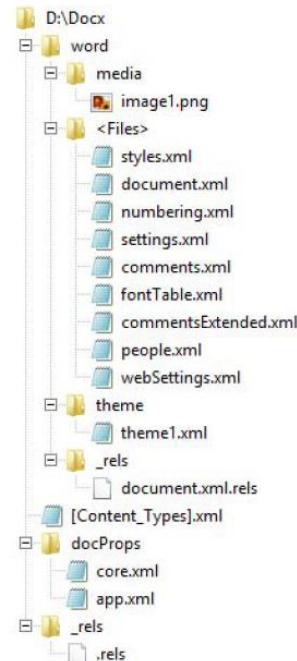
- Locations of objects within the file - for random access

**Trailer**

- Locations of certain objects within the file
- Location of the Cross-reference Table in the file

```

%PDF-1.6
32 0 obj
<</Length 78/C 85/Filter/FlateDecode/1 107/0 69/S 38>>stream
x0b...e' 04 0;00 '00' 00000000' 0000'0000'
00-00000000' 0000'0000' 0000'0000' 0000'0000'
endstream
Endobj
26 0 obj<</Names 27 0 R/Outlines 1 0 R
R/SpiderInfo 19 0 R/StructTreeRoot 4 0 R
endobj
27 0 obj<</IDS 11 0 R/URLS 12 0 R>>
Endobj
28 0 obj<</CropBox[0.0 0.0 612.0 792.0]/Parent 10 0 R/ID 31 0
R/StructParents 0/Contents 29 0 R/Rotate 0/MediaBox[0.0 0.0 612.0
792.0]/Resources<</Font<</F1_0 30 0 R>>/ProcSet[/PDF/Text]>>/Type/Page>>
Endobj
xref
6 9
0000000016 00000 n
0000000625 00000 n
0000000701 00000 n
0000000827 00000 n
0000000910 00000 n
0000001210 00000 n
0000001617 00000 n
0000001850 00000 n
0000000476 00000 n
trailer
<<
/Size 15
/Prev 5871
/Root 7 0 R
/Info 5 0 R
/ID[<-8472F86D2B7970DA5A126EABCD510B6D><-8ED405EE9ADA3D45893AFARD2BF4D9BE>]
>>
startxref
0
%%EOF
  
```



설명	구별 이름	길이(바이트)	레코드 구조	압축/암호화
파일 인식 정보	FileHeader	고정		
문서 정보	DocInfo	고정	✓	✓
본문	BodyText	가변	✓	✓
문서 요약	\005HwpSummaryInformation	고정		
바이너리 데이터	BinData	가변		✓
미리보기 텍스트	PrvText	고정		
미리보기 이미지	PrvImage	가변		
문서 옵션	DocOptions	가변		
스크립트	Scripts	가변		
XML 템플릿	XMLTemplate	가변		
문서 이력 관리	DocHistory	가변	✓	✓





## 2. 프로젝트 소개

### 개발 내용



```
obj 1 0
Type: /Catalog
Referencing: 2 0 R, 17 0 R, 1151 0 R, 1152 0 R

<<
  /Type /Catalog
  /Pages 2 0 R
  /Lang (ko-KR)
  /StructTreeRoot 17 0 R
  /MarkInfo
    <<
      /Marked true
    >>
  /Metadata 1151 0 R
  /ViewerPreferences 1152 0 R
>>

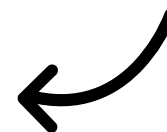
obj 2 0
Type: /Pages
Referencing: 3 0 R, 14 0 R

<<
  /Type /Pages
  /Count 2
  /Kids [ 3 0 R 14 0 R ]
>>
```



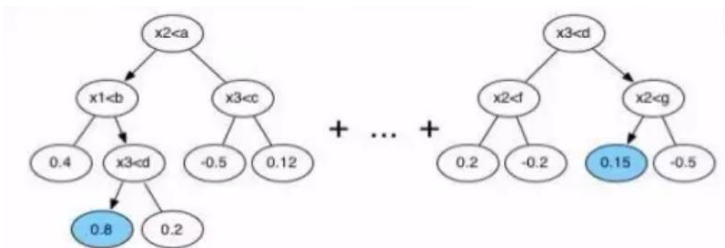
$F_1$	$F_2$	$F_3$	$F_4$	...	$F_n$
1.7	12	9	124	...	0

Feature Vector

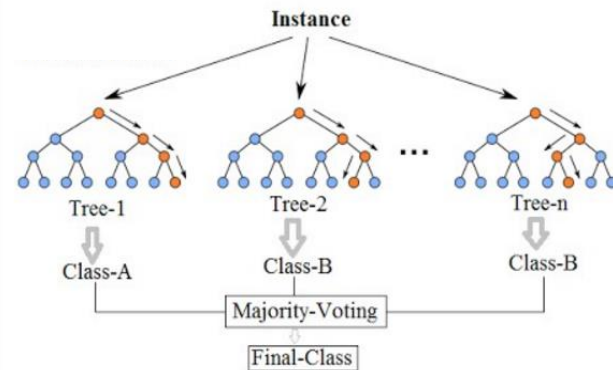


## 2. 프로젝트 소개

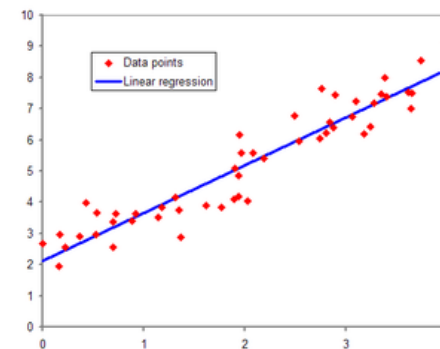
### 개발 내용



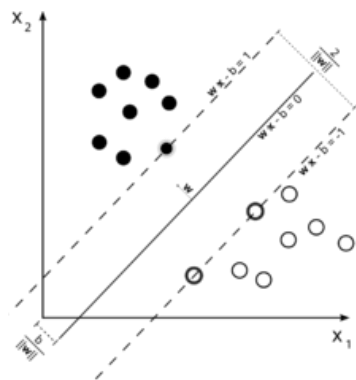
GBDT



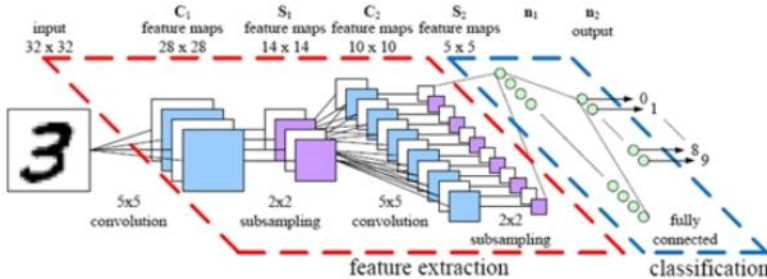
RandomForest



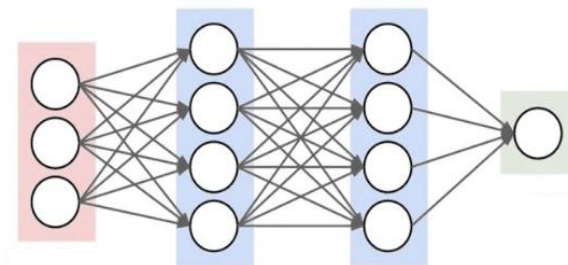
Linear Regression



SVM



CNN

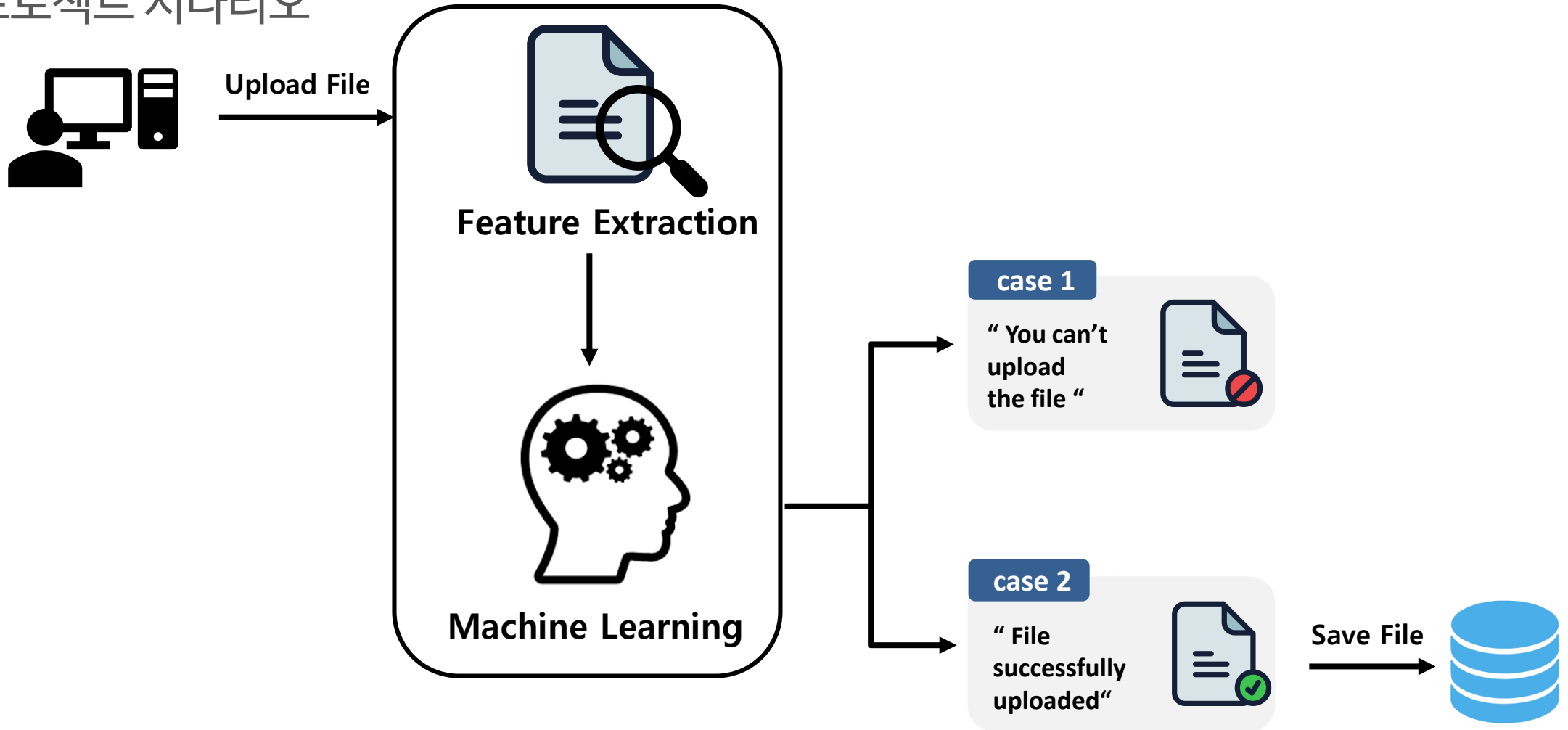


DNN



## 2. 프로젝트 소개

### 프로젝트 시나리오



### 3. 기대효과

#### 1. 문서형 악성코드 유포 방지

문서형 악성코드 유포 방지로 사회 문제 해소



#### 2. 오픈소스 소프트웨어

다른 시스템 환경에 맞게 코드 수정 및 사용 가능



## 4. 역할분담



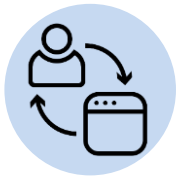
### 문다민 (팀장)

머신 러닝 모델 설계 및 구축



### 김기환

특징 추출 및 가공, 데이터 수집



### 방유한

유저 인터페이스 구현



### 김현석

데이터 라벨링, 웹 서버 구축



### 정혜리

특징 추출 및 가공, 문서 작업





# 감사합니다.

문다민 김기환 김현석 정혜리 방유한



# Reference

- GBDT 사진 : <https://ask.hellobi.com/blog/mlanddlanddm/7103>
- RandomForest 사진 : <https://yeo0.github.io/data/2018/11/06/6.-%EB%9E%9C%EB%8D%A4-%ED%8F%AC%EB%A0%88%EC%8A%A4%ED%8A%B8/>
- 선형회귀 사진 : [https://ko.wikipedia.org/wiki/%EC%84%A0%ED%98%95\\_%ED%9A%8C%EA%B7%80](https://ko.wikipedia.org/wiki/%EC%84%A0%ED%98%95_%ED%9A%8C%EA%B7%80)
- SVM 사진 : [https://ko.wikipedia.org/wiki/%EC%84%9C%ED%8F%AC%ED%8A%B8\\_%EB%B2%A1%ED%84%B0\\_%EB%A8%B8%EC%8B%A0](https://ko.wikipedia.org/wiki/%EC%84%9C%ED%8F%AC%ED%8A%B8_%EB%B2%A1%ED%84%B0_%EB%A8%B8%EC%8B%A0)
- CNN 사진: <http://physics2.mju.ac.kr/juhapruwp/?p=1517>
- DNN 사진: <https://www.slideshare.net/JinwonLee9/ss-70446412>

# Reference

- 기사 1 – <https://www.boannews.com/media/view.asp?idx=75093>
- 기사 2 –  
<https://www.boannews.com/media/view.asp?idx=71995&kind=1&search=title&find=%B9%AE%BC%AD%C7%FC>
- 기사 3 – <https://www.boannews.com/media/view.asp?idx=74302>