



a security insight

캡스톤 디자인 5조 어시스트



01

프로젝트 목표

02

진행 상황

03

계획 및 제한요소

01

프로젝트 목표

02

진행 상황

03

계획 및 제한요소

asi - 핵심 아이디어

```

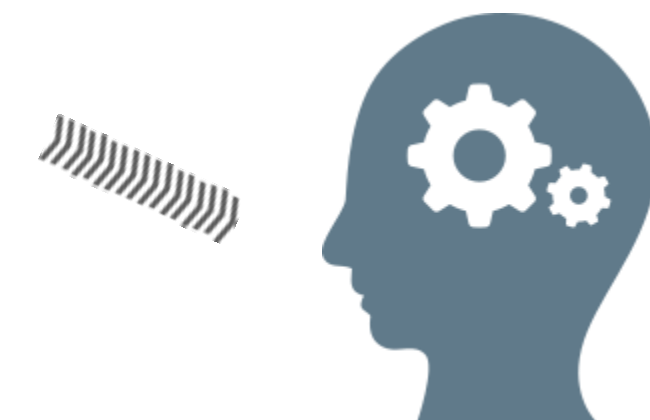
10001180: 55      push %ebp
10001181: 8b ec   mov %esp,%ebp
10001183: 6a ff   push $0xffffffff
10001185: 68 21 80 00 10 push $0x10008021
1000118a: 64 a1 00 00 00 00 mov %fs:0x0,%eax
10001190: 50      push %eax
10001191: 83 ec 08 sub $0x8,%esp
10001194: a1 00 40 02 10 mov 0x10024000,%eax
10001199: 33 c5   xor %ebp,%eax
1000119b: 89 45 f0 mov %eax,-0x10(%ebp)
1000119e: 53      push %ebx
1000119f: 56      push %esi
100011a0: 57      push %edi
100011a1: 50      push %eax
100011a2: 8d 45 f4 lea -0xc(%ebp),%eax
100011a5: 64 a3 00 00 00 00 mov %eax,%fs:0x0
100011ab: 6a 14   push $0x14
100011ad: 6a 0c   push $0xc
100011af: ff 15 44 b2 01 10 call *0x1001b244
100011b5: 83 c4 08 add $0x8,%esp
100011b8: 89 45 ec mov %eax,-0x14(%ebp)
100011bb: 8b c8   mov %eax,%ecx
100011bd: c7 45 fc 00 00 00 00 movl $0x0,-0x4(%ebp)
100011c4: e8 c7 04 00 00 call 0x10001690

```

disassemble



RNN



CNN



elastic

```

10001180: 55      push %ebp
10001181: 8b ec   mov %esp,%ebp
10001183: 6a ff   push $0xffffffff
10001185: 68 21 80 00 10 push $0x10008021
1000118a: 64 a1 00 00 00 00 mov %fs:0x0,%eax
10001190: 50      push %eax
10001191: 83 ec 08 sub $0x8,%esp
10001194: a1 00 40 02 10 mov 0x10024000,%eax
10001199: 33 c5   xor %ebp,%eax
1000119b: 89 45 f0 mov %eax,-0x10(%ebp)
1000119e: 53      push %ebx
1000119f: 56      push %esi
100011a0: 57      push %edi
100011a1: 50      push %eax
100011a2: 8d 45 f4 lea -0xc(%ebp),%eax
100011a5: 64 a3 00 00 00 00 mov %eax,%fs:0x0
100011ab: 6a 14   push $0x14
100011ad: 6a 0c   push $0xc
100011af: ff 15 44 b2 01 10 call *0x1001b244
100011b5: 83 c4 08 add $0x8,%esp
100011b8: 89 45 ec mov %eax,-0x14(%ebp)
100011bb: 8b c8   mov %eax,%ecx
100011bd: c7 45 fc 00 00 00 00 movl $0x0,-0x4(%ebp)
100011c4: e8 c7 04 00 00 call 0x10001690

```

이상 탐지

md5	cosine	edit
02a7993fcd5fea4442271e91e12d2df7	0.85	0.73
07FADB006486953439CE0092651FD7A6	0.21	0.32
344fbbbedc59a0a5108da10d4afd2152	0.90	0.87

유사도 검사

악성코드 의심 파일 분석

asi - 웹 서비스



<파일 업로드 초기 시각화 안>

asi - 웹 서비스



dasjfiwf.exe 1034 lines

Line	Hex	Op	Operand
17	6a ff	push	\$0xffffffff
18	68 21 80 00 10	push	\$0x10008021
19	64 a1 00 00 00 00	mov	%fs:0x0,%eax
20	50	push	%eax
21	83 ec 08	sub	\$0x8,%esp

< 2 / 12 > 목록으로 전체보기

<분석 결과 초기 시각화 안>

01

프로젝트 목표

02

진행 상황

03

계획 및 제한요소

데이터 수집



microsoft
malware prediction

정상 파일 55,000



정보보호 R&D
데이터챌린지



금융보안원
FINANCIAL SECURITY INSTITUTE

악성 파일 10,000개



자체 개발
웹 크롤러

시스템 DLL 파일
STEAM사 게임 인스톨러

니모닉 추출



disassemble
IDA

10001180:	55	push	%ebp
10001181:	8b ec	mov	%esp,%ebp
10001183:	6a ff	push	\$0xffffffff
10001185:	68 21 80 00 10	push	\$0x10008021
1000118a:	64 a1 00 00 00 00	mov	%fs:0x0,%eax
10001190:	50	push	%eax
10001191:	83 ec 08	sub	\$0x8,%esp
10001194:	a1 00 40 02 10	mov	0x10024000,%eax
10001199:	33 c5	xor	%ebp,%eax
1000119b:	89 45 f0	mov	%eax,-0x10(%ebp)
1000119e:	53	push	%ebx
1000119f:	56	push	%esi
100011a0:	57	push	%edi
100011a1:	50	push	%eax
100011a2:	8d 45 f4	lea	-0xc(%ebp),%eax
100011a5:	64 a3 00 00 00 00	mov	%eax,%fs:0x0
100011ab:	6a 14	push	\$0x14
100011ad:	6a 0c	push	\$0xc
100011af:	ff 15 44 b2 01 10	call	*0x1001b244
100011b5:	83 c4 08	add	\$0x8,%esp
100011b8:	89 45 ec	mov	%eax,-0x14(%ebp)
100011bb:	8b c8	mov	%eax,%ecx
100011bd:	c7 45 fc 00 00 00 00	movl	\$0x0,-0x4(%ebp)
100011c4:	e8 c7 04 00 00	call	0x10001690

parser
python, IDA

push
mov
push
push
mov
push
sub
sub
mov
xor
mov
push
push
push
push
lea
mov
push
push
call
add
mov
mov
movl
call

File

Assembly code

Mnemonic

단어 임베딩



단어 임베딩

| **실험 조건** **gensim** 라이브러리 사용

- 01 윈도우 크기 : 10
- 02 최소 단어 수 : 50
- 03 에폭 : 10
- 04 학습률 : 0.002
- 05 특징 벡터 차원 : 8/16/32/64/128

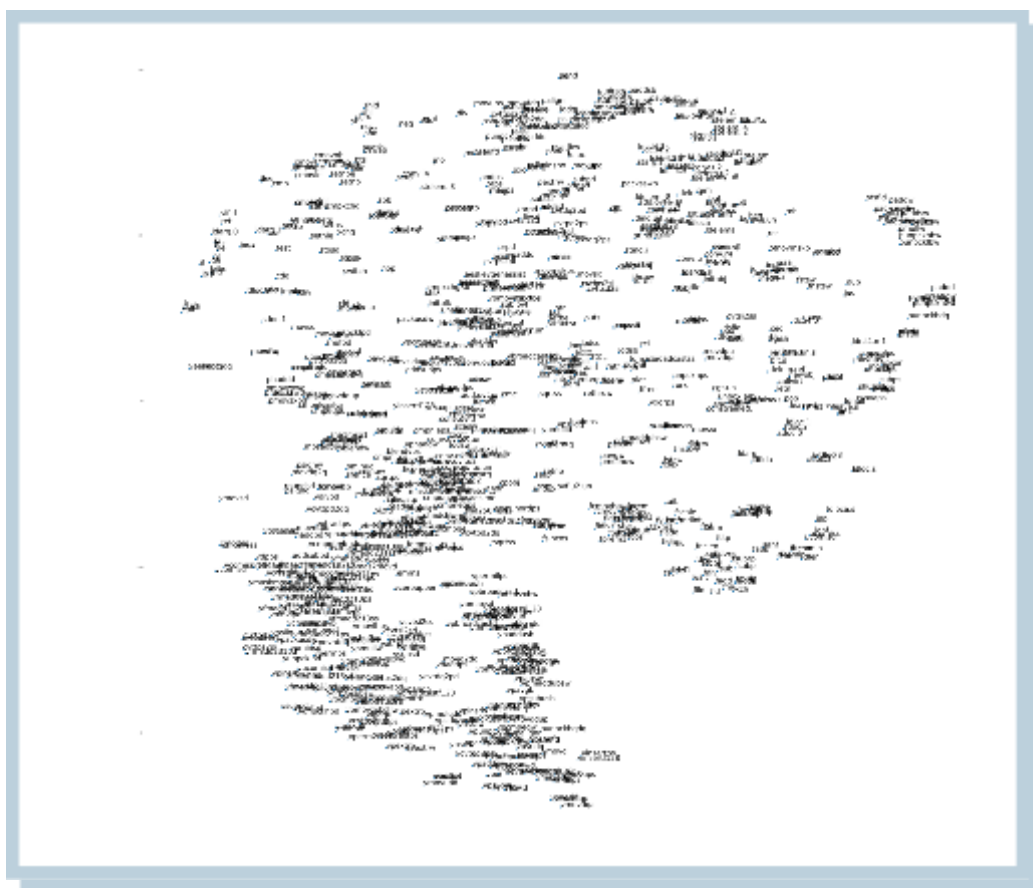
mov jmp add pop push

단어 임베딩

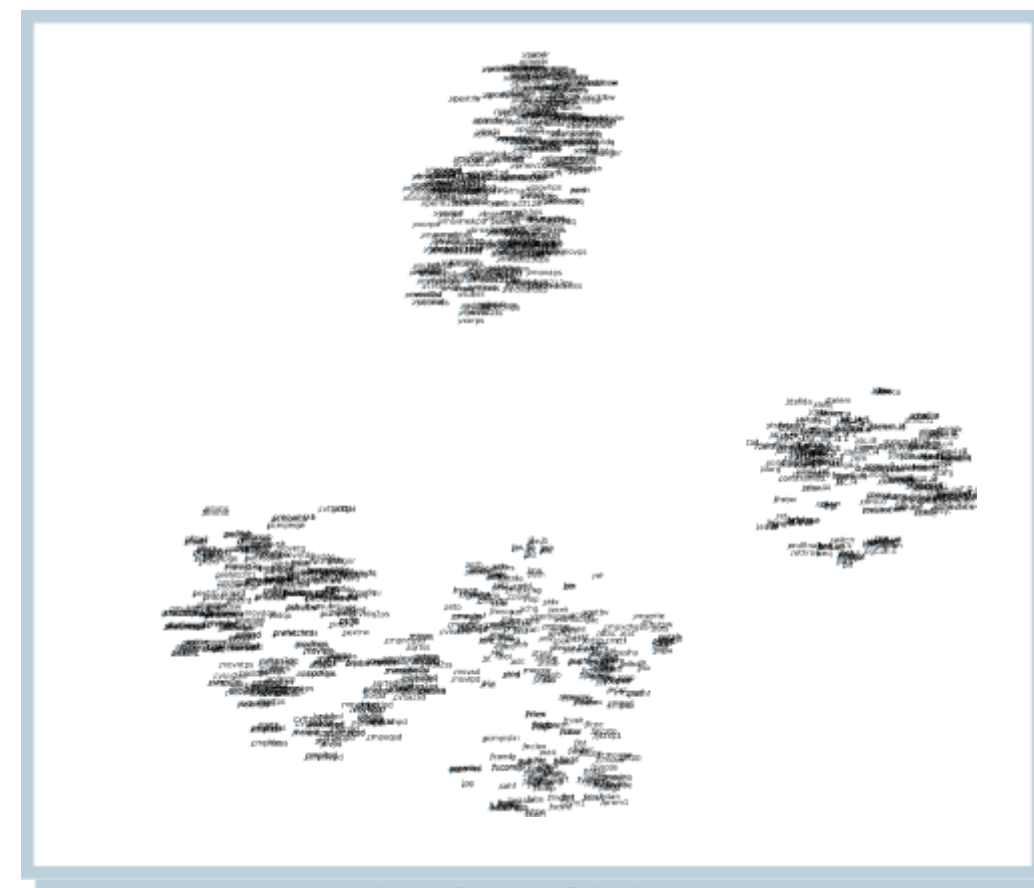
| 실험 결과

SkipGram

특징 벡터 차원 : 8/16/32/64/128



특징 벡터 차원 : 16



특징 벡터 차원 : 64

단어 임베딩

| 실험 결과

	vec 8	vec16	vec64
jmp	xend	xor	mov
	inc	mov	jnz
	fcmovnu	inc	jz
	vfmadd213pd	or	cmp
	xor	jnz	lea
	dec	lea	test
	lgdt	test	push
	ht jge	cmp	xor
	setno	movzx	retn
	cvtt2pi	jz	inc

오토인코더 기반 이상 탐지

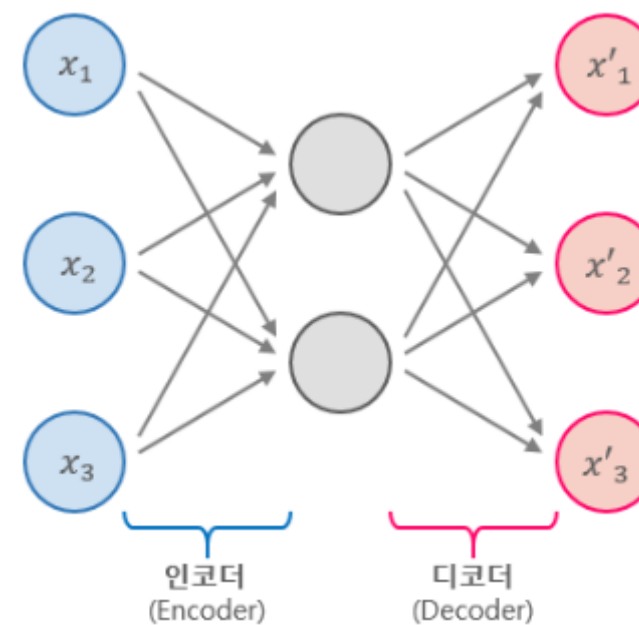


File

disassemble
IDA, Parser

push
mov
push
push
mov
push
sub
mov
xor
mov
push
push
push
push
lea
mov
push
push
call
add
mov
mov
movl
call

Mnemonic



RNN 기반
오토인코더

function 1

0.2

push
sub
add
str
ldr
ldr
ldr
:
bx

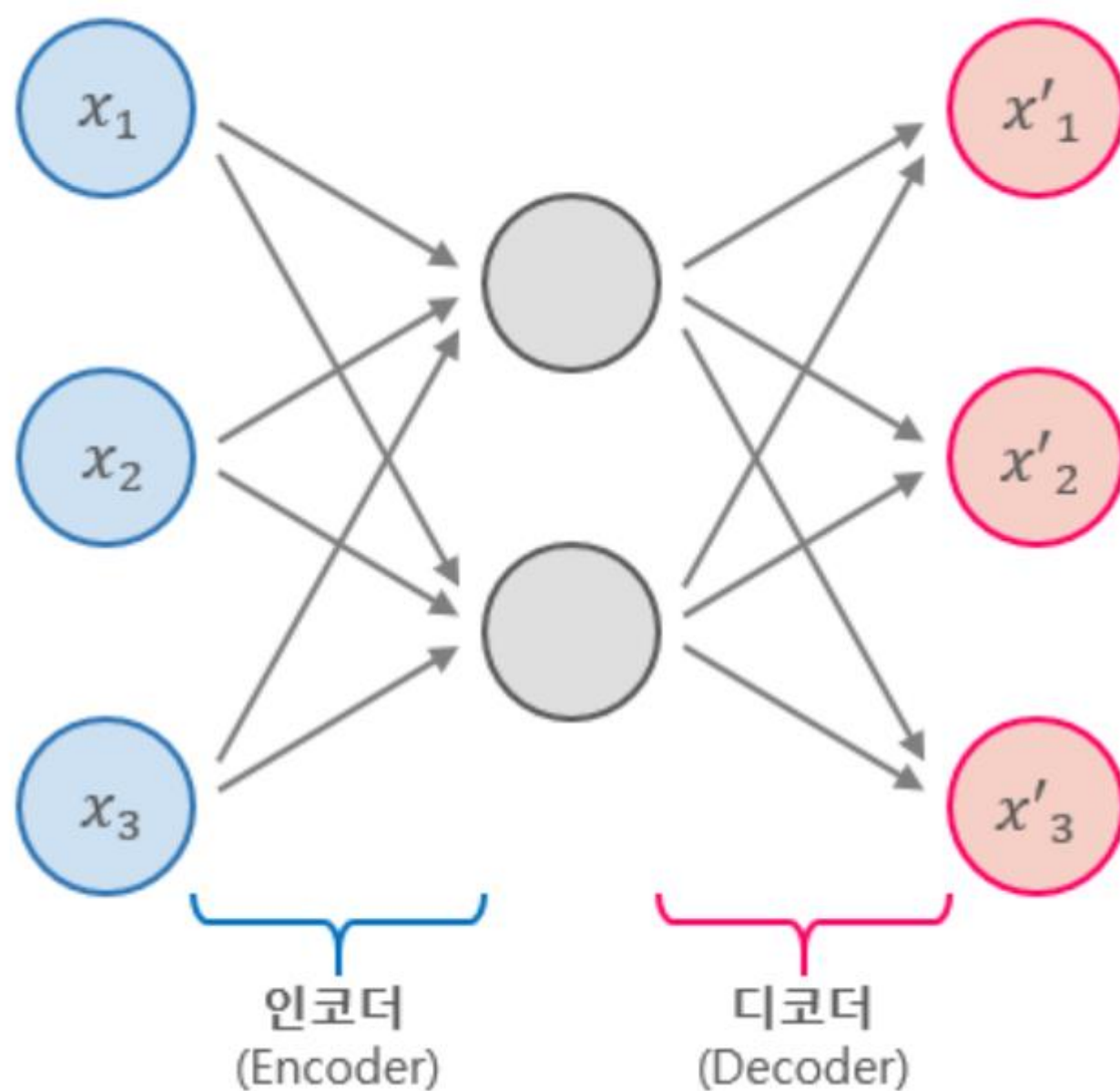
function 2

0.7

push
sub
add
mov
strb
ldrb
lsls
:
bx

이상탐지

이상 탐지

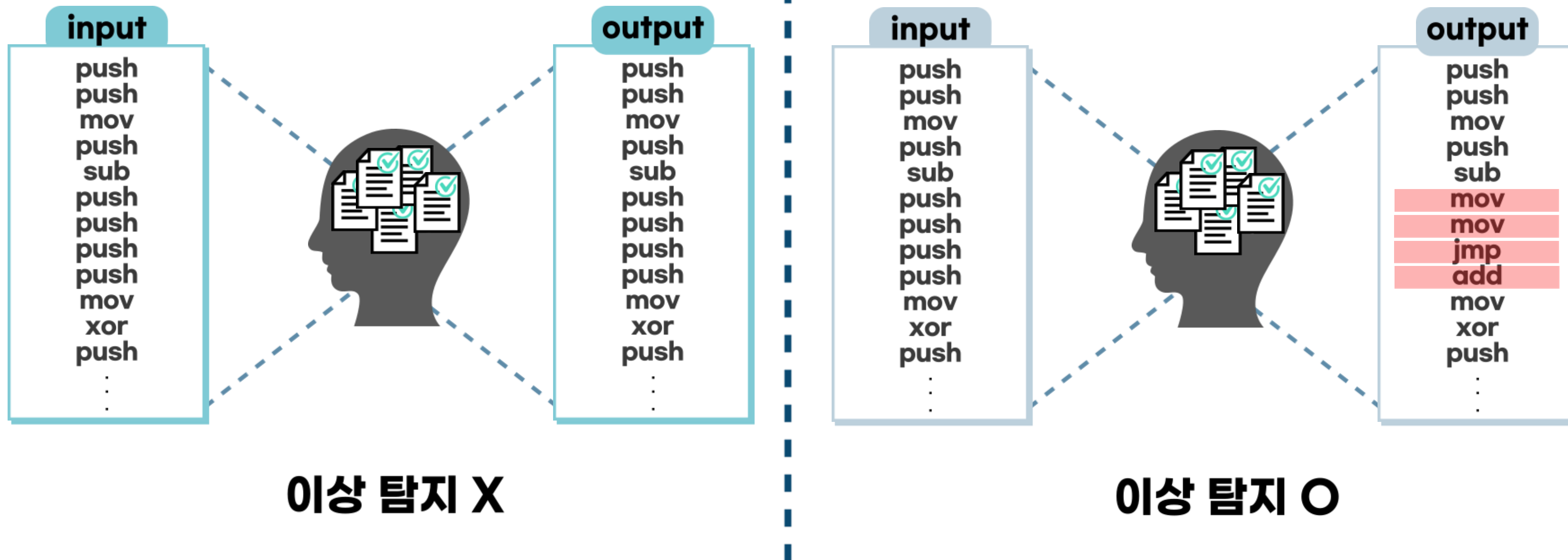


오토 인코더 이상 탐지

대표적 비지도 학습법

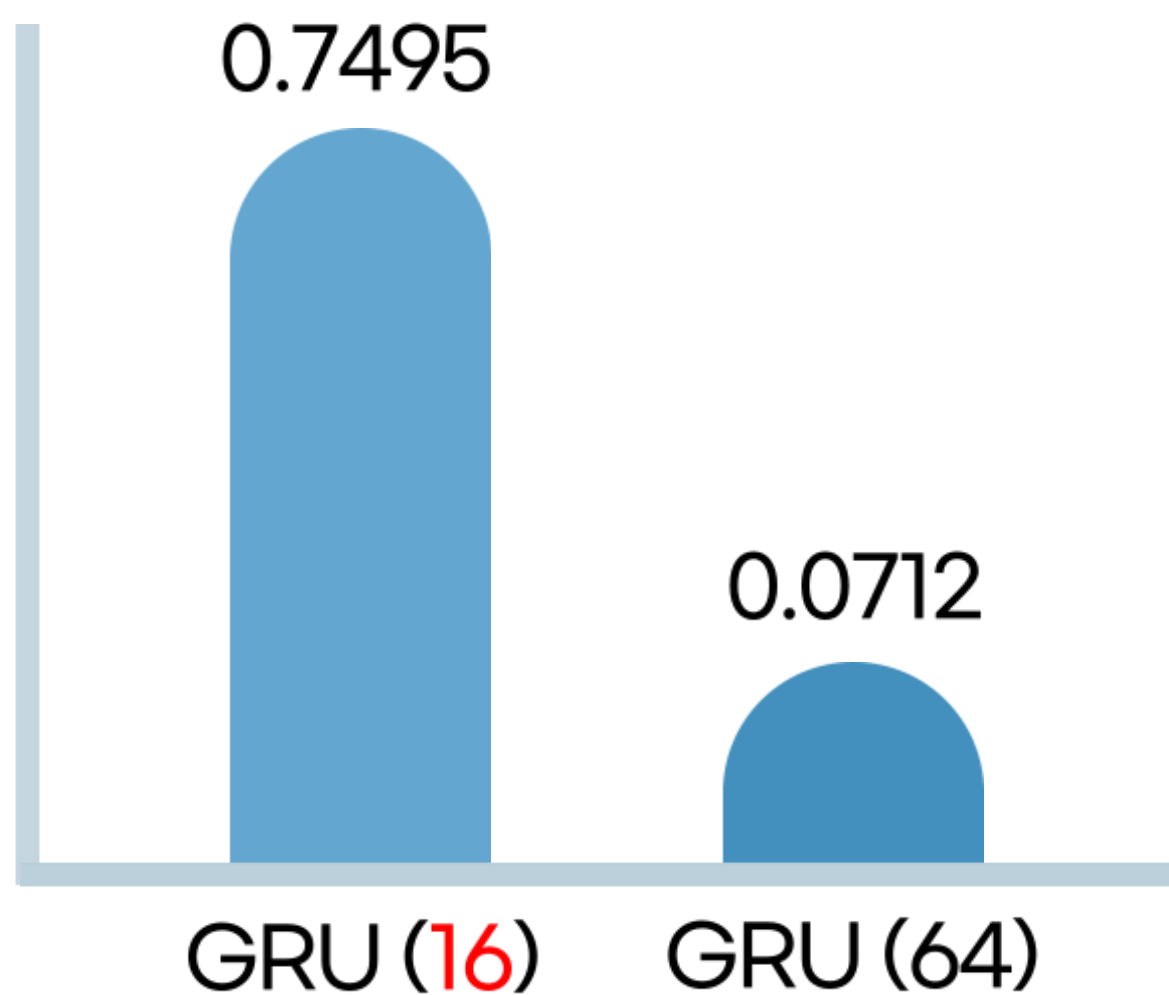
입력 값과 출력 값을 같게 함

이상 탐지



이상 탐지

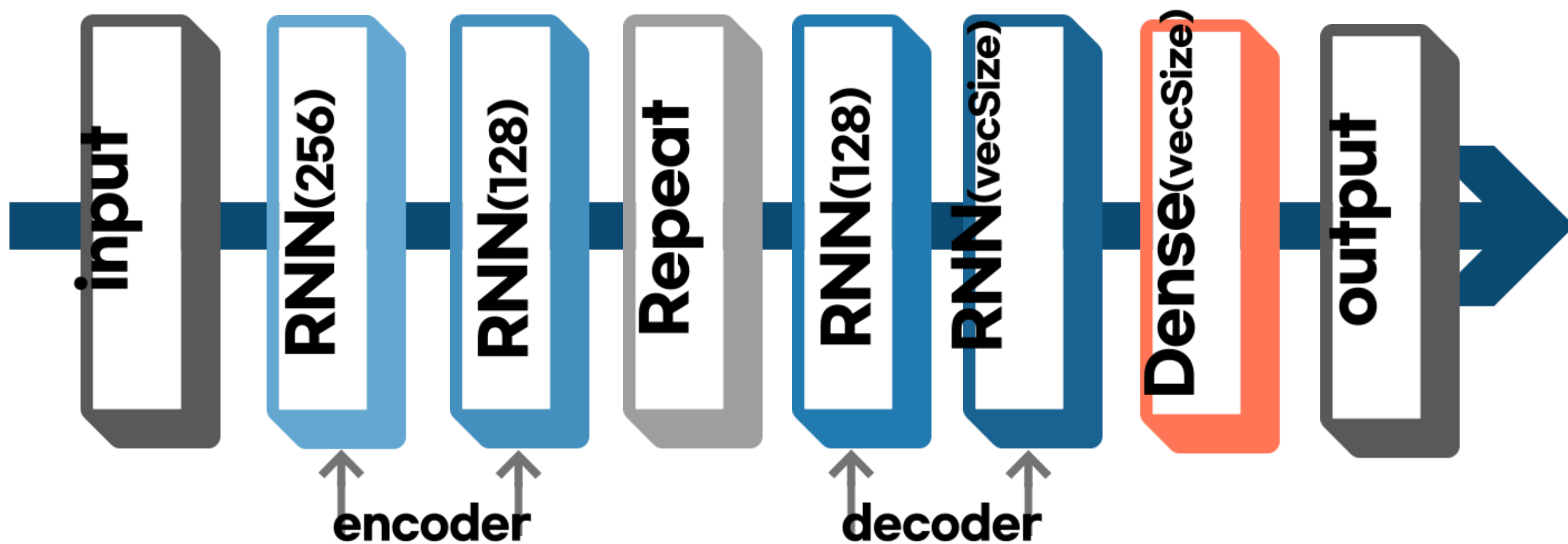
| 실험 결과 - 손실값



특징벡터 차원 → 64 차원

이상 탐지

| 신경망 구조



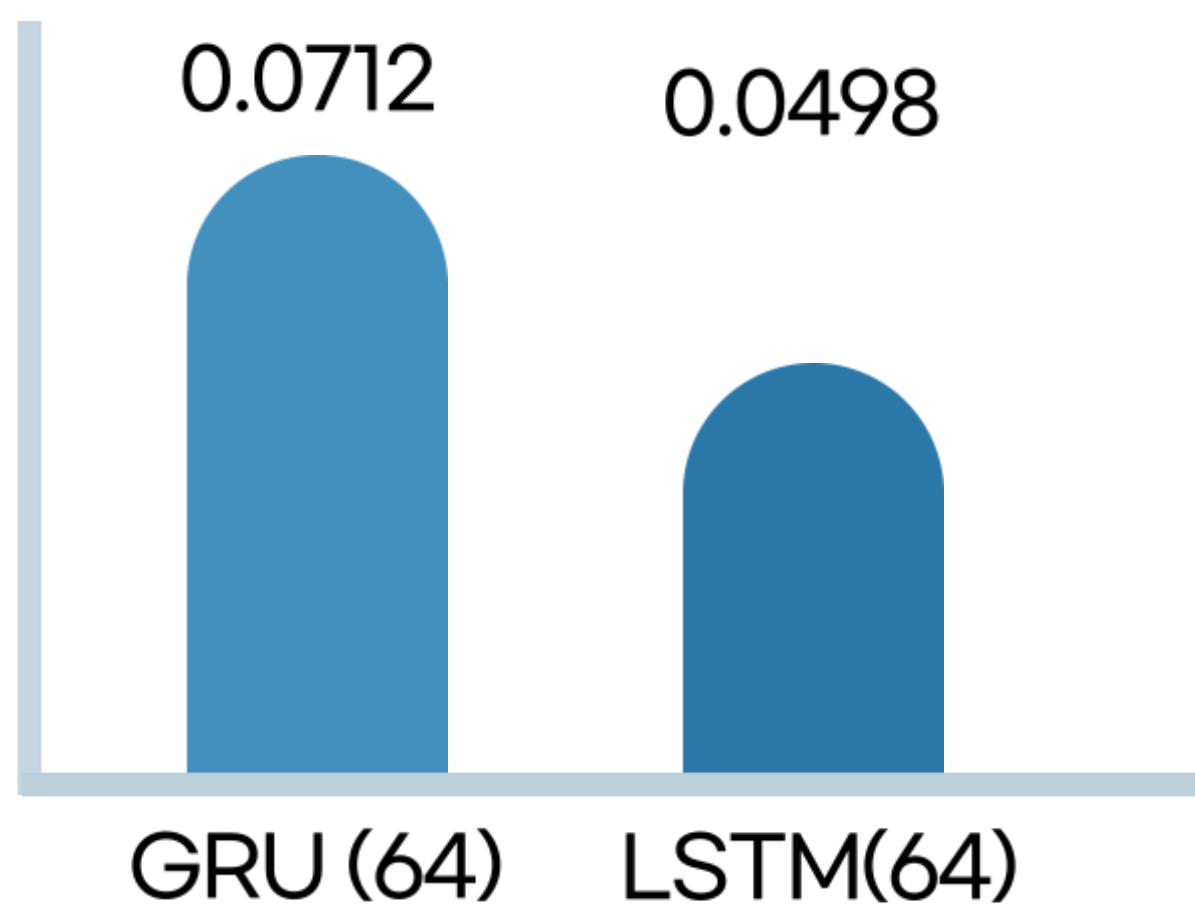
이상 탐지

| 실험 조건

- 01 정상 10,000개
- 02 벡터 크기 : 16 / 64
- 03 신경망 : GRU / LSTM
- 04 결과 스코어 값이 임계값(0.2)보다 크면 이상탐지
 - ↳ 스코어 산출 방식 : MAE(mean absolute error)

이상 탐지

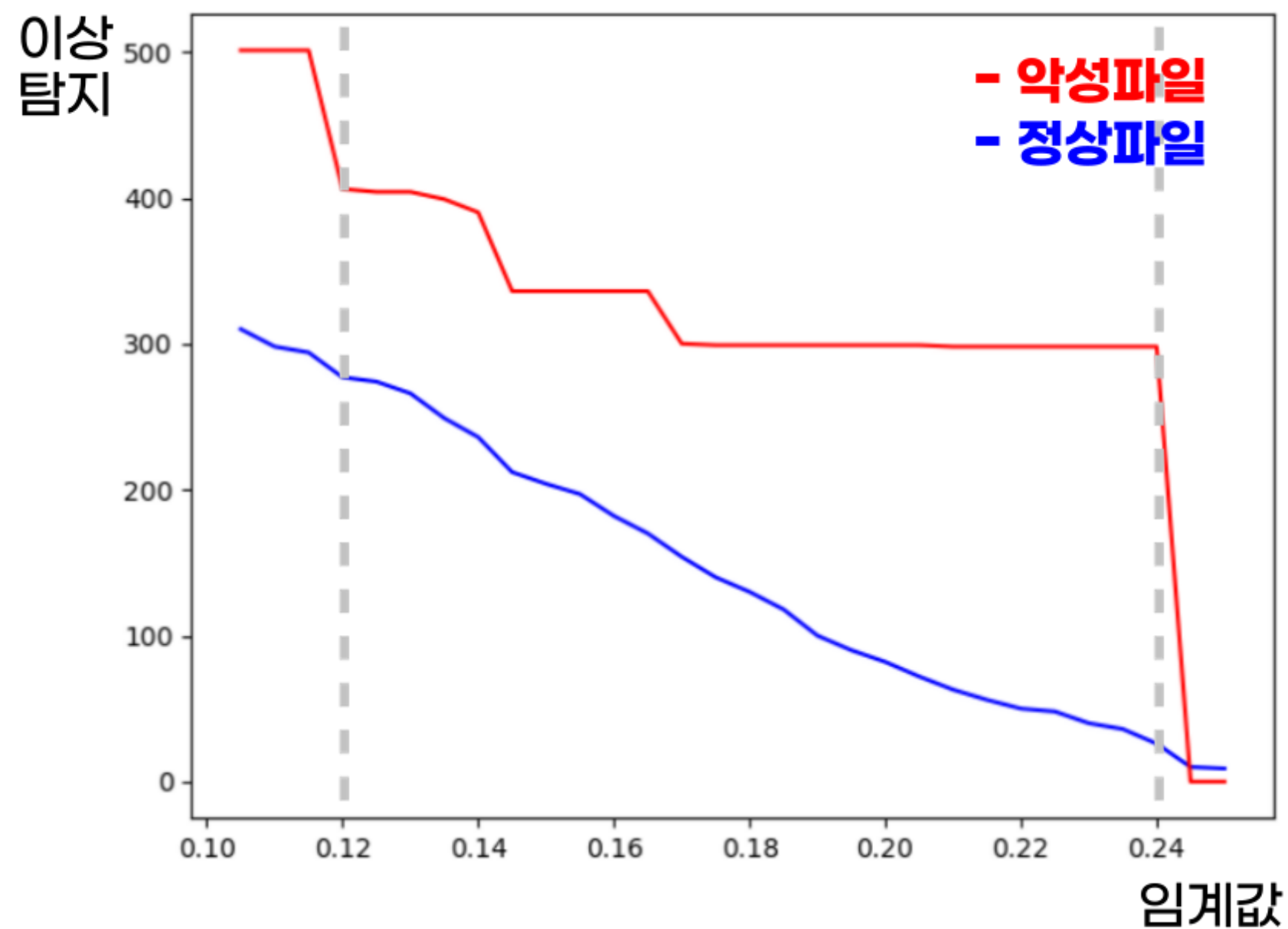
| 실험 결과 - 손실 값



GRU 보다 **LSTM**이 효과적

이상 탐지

| 실험 결과 - 이상 탐지



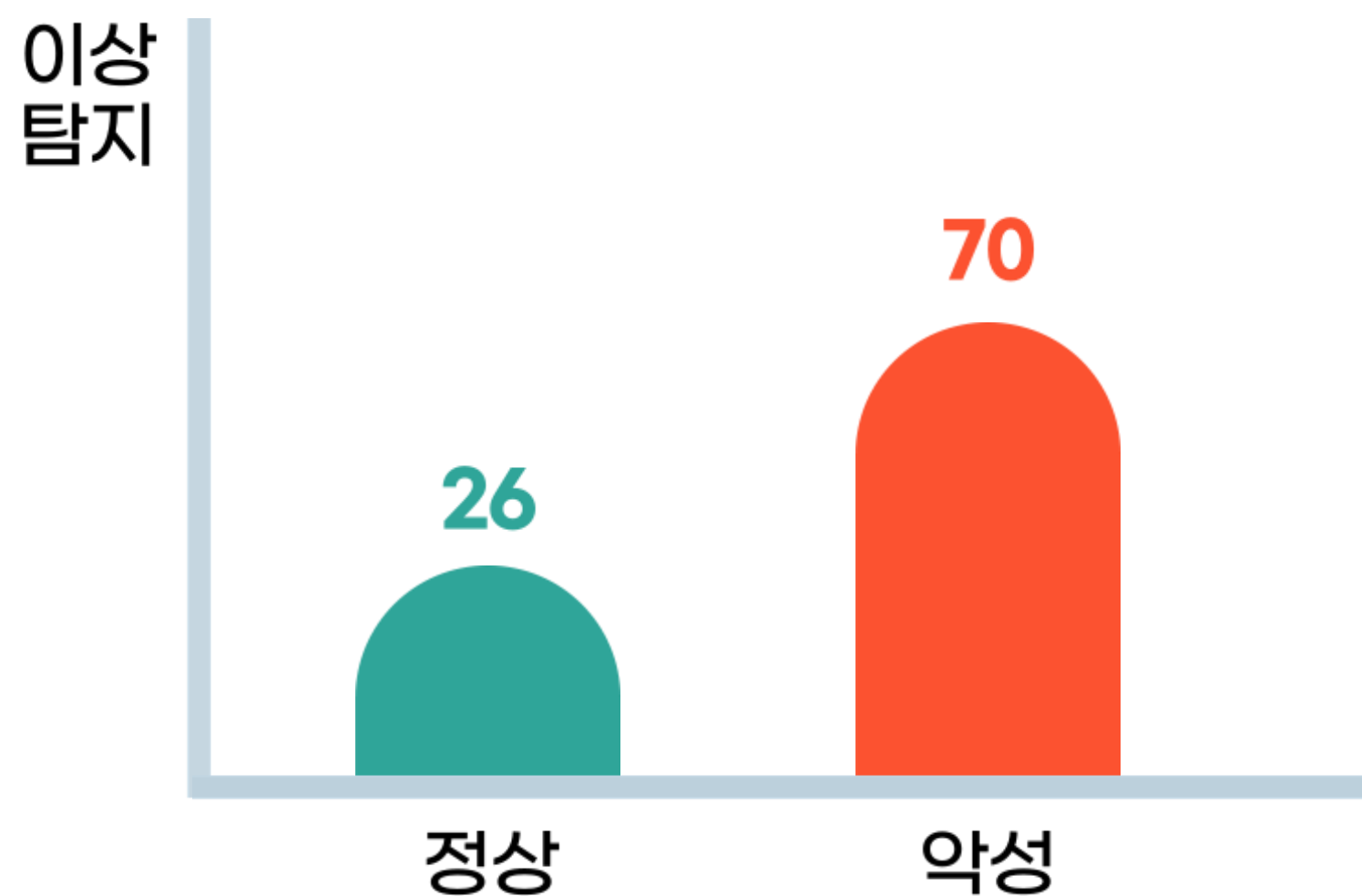
임계값 < 0.12 → 오탐 多

임계값 > 0.24 → 미탐 多

임계 값 : 0.2

이상 탐지

| 실험 결과 - 이상 탐지



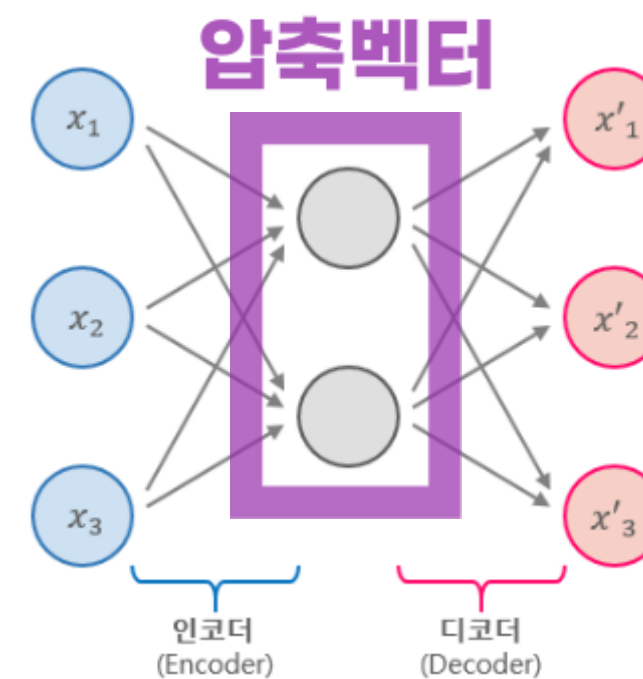
이상 탐지 비율

정상 < 악성

압축벡터 기반 유사도 검사

push
mov
push
push
mov
push
sub
sub
mov
xor
mov
push
push
push
push
lea
mov
push
push
call
add
mov
mov
movl
call

Mnemonic



CNN 기반
오토인코더

압축벡터 기반 유사도 검사

```
"vector" : [
  0.5706483,
  0.7005931,
  0.40261483,
  0.6934074,
  0.54094464,
  0.4278791,
  0.13182177,
  0.40346572,
  0.37148055,
  0.45193487,
  0.44311434,
]
```

압축벡터

cosine 유사도



정상: 200만개 악성: 120만개

유사도 검색

정상 : 20%

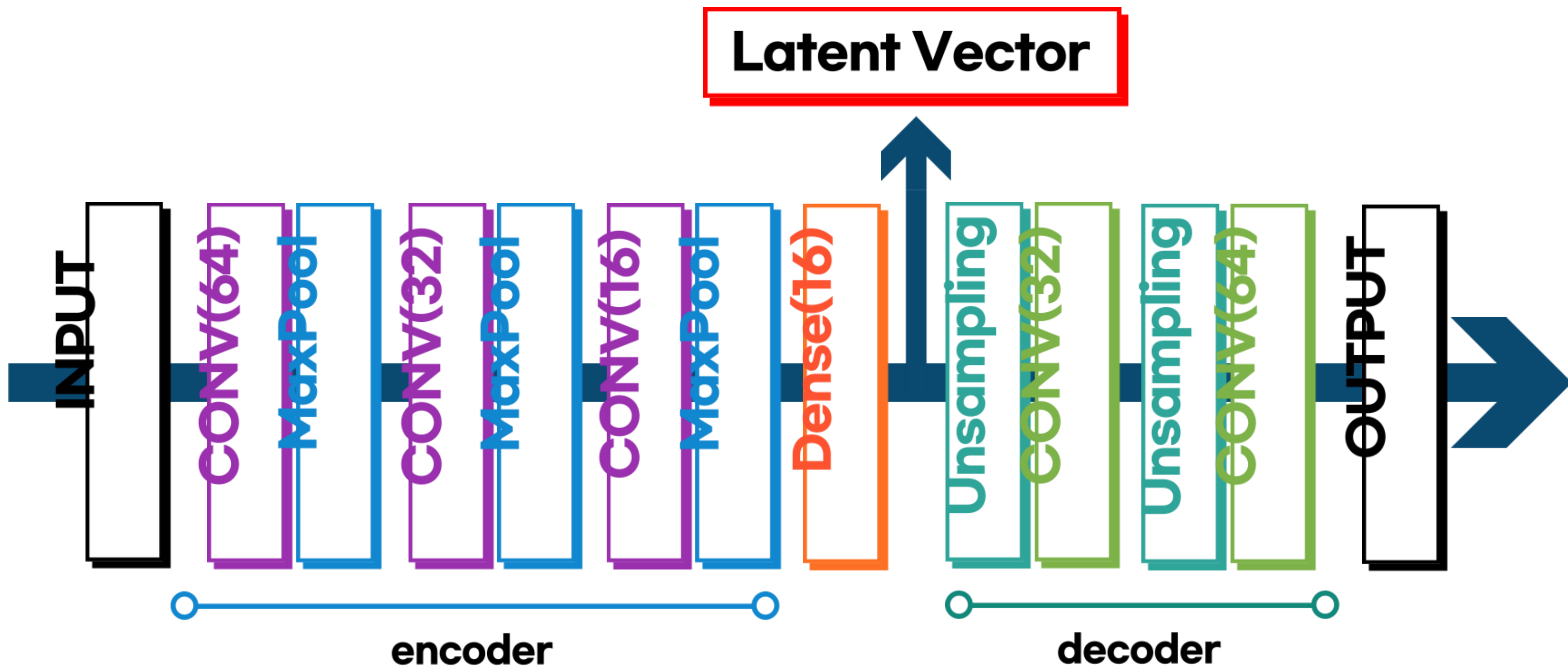
```
md5 : 02a7993fcd5fea4442271e91e12d2df7
md5 : 07FADB006486953439CE0092651FD7A6
md5 : 344fbbedc59a0a5108da10d4afd2152
⋮
```

악성 : 80%

```
md5 : 02a7993fcd5fea4442271e91e12d2df7
md5 : 07FADB006486953439CE0092651FD7A6
md5 : 344fbbedc59a0a5108da10d4afd2152
md5 : 02a7993fcd5fea4442271e91e12d2df7
md5 : 07FADB006486953439CE0092651FD7A6
md5 : 344fbbedc59a0a5108da10d4afd2152
⋮
```

악성/정상 비율

압축벡터 기반 유사도 검사

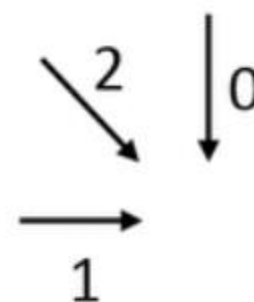


압축벡터 기반 유사도 검사

| 코사인 유사도 검증 - 편집 거리

		M	O	N	K	E	Y
	0	1	2	3	4	5	6
M	1	0	1	2	3	4	5
O	2	1	0	1	2	3	4
N	3	2	1	0	1	2	3
E	4	3	2	1	1	1	2
Y	5	4	3	2	2	2	1

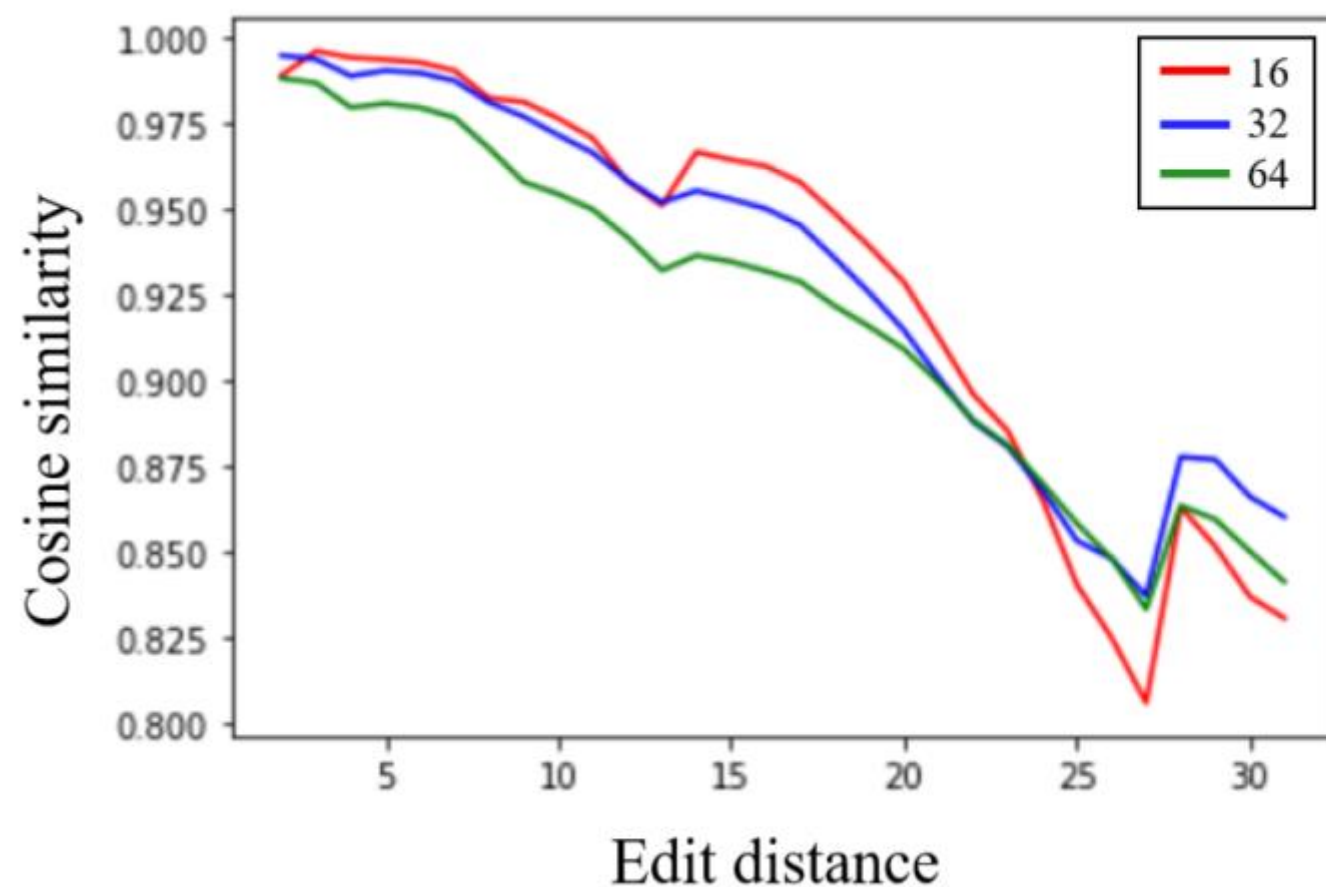
		M	O	N	K	E	Y
	0	1	1	1	1	1	1
M	0	2	1	1	1	1	1
O	0	0	2	1	1	1	1
N	0	0	0	2	1	1	1
E	0	0	0	0	2	2	1
Y	0	0	0	0	0	0	2



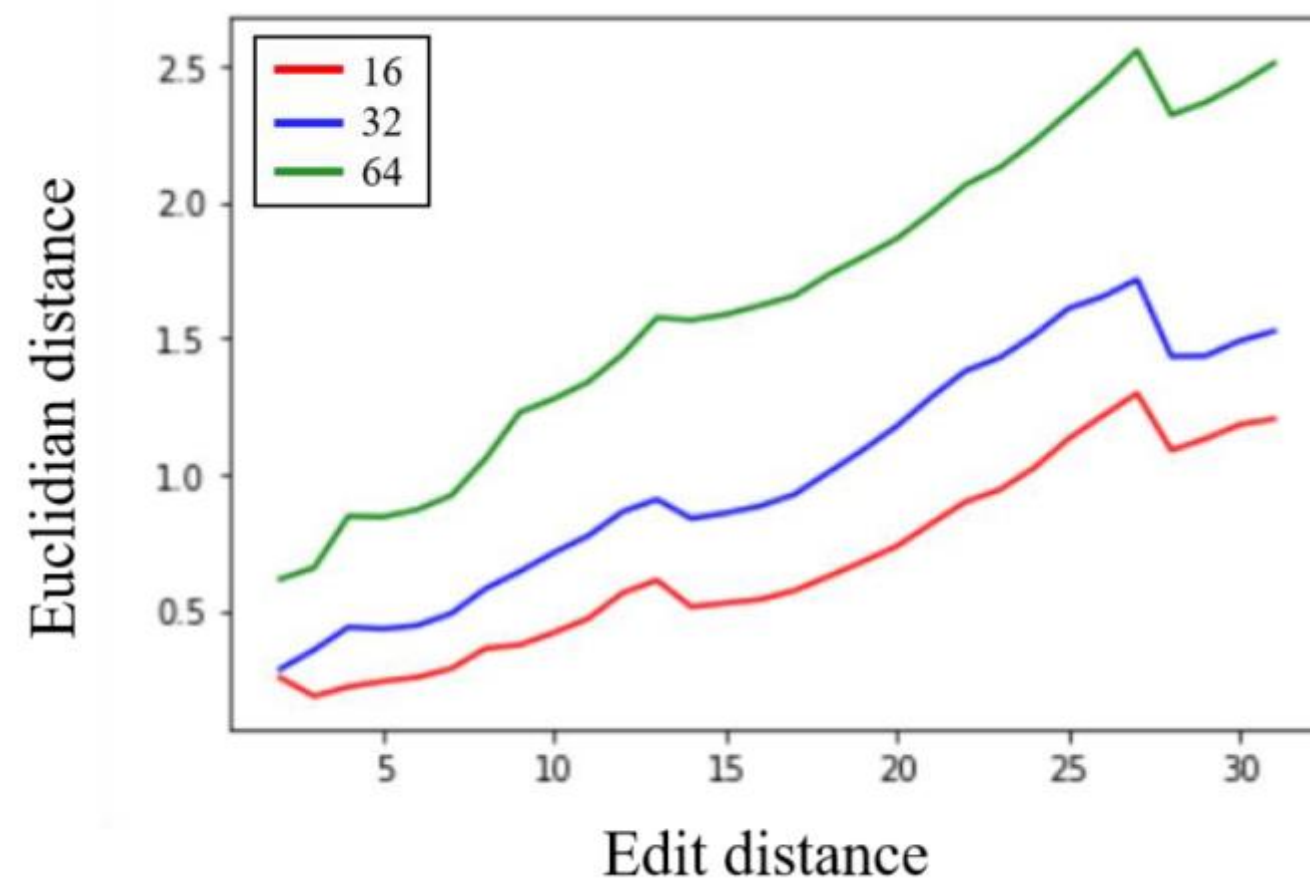
**일치하게 만들기 위해
편집하는 횟수**

압축벡터 기반 유사도 검사

| 실험 결과 - 코사인 유사도 검증



편집거리와 코사인 유사도 반비례



편집거리와 유클리드 거리 비례

압축벡터 기반 유사도 검사

| 실험 결과 - 코사인 유사도 검증

```
ldarg.0  
callvirt  
callvirt  
call  
stloc.0  
ldarg.1  
callvirt  
callvirt  
call  
stloc.1  
ldloc.0  
...
```

MD5 : 000091e9cc8946301647fbd01fed6ce1
Index : 4814

유사도 높음

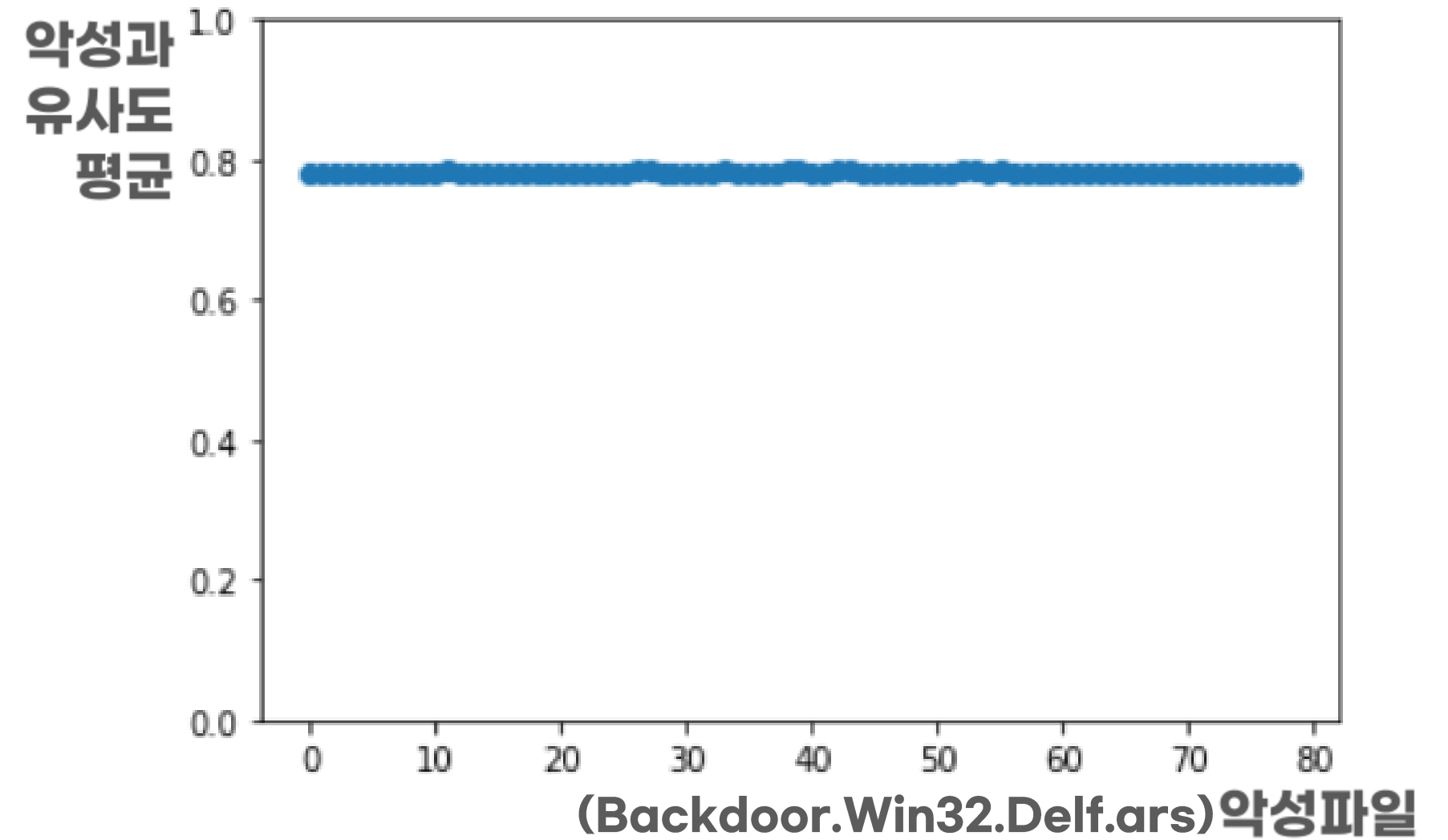
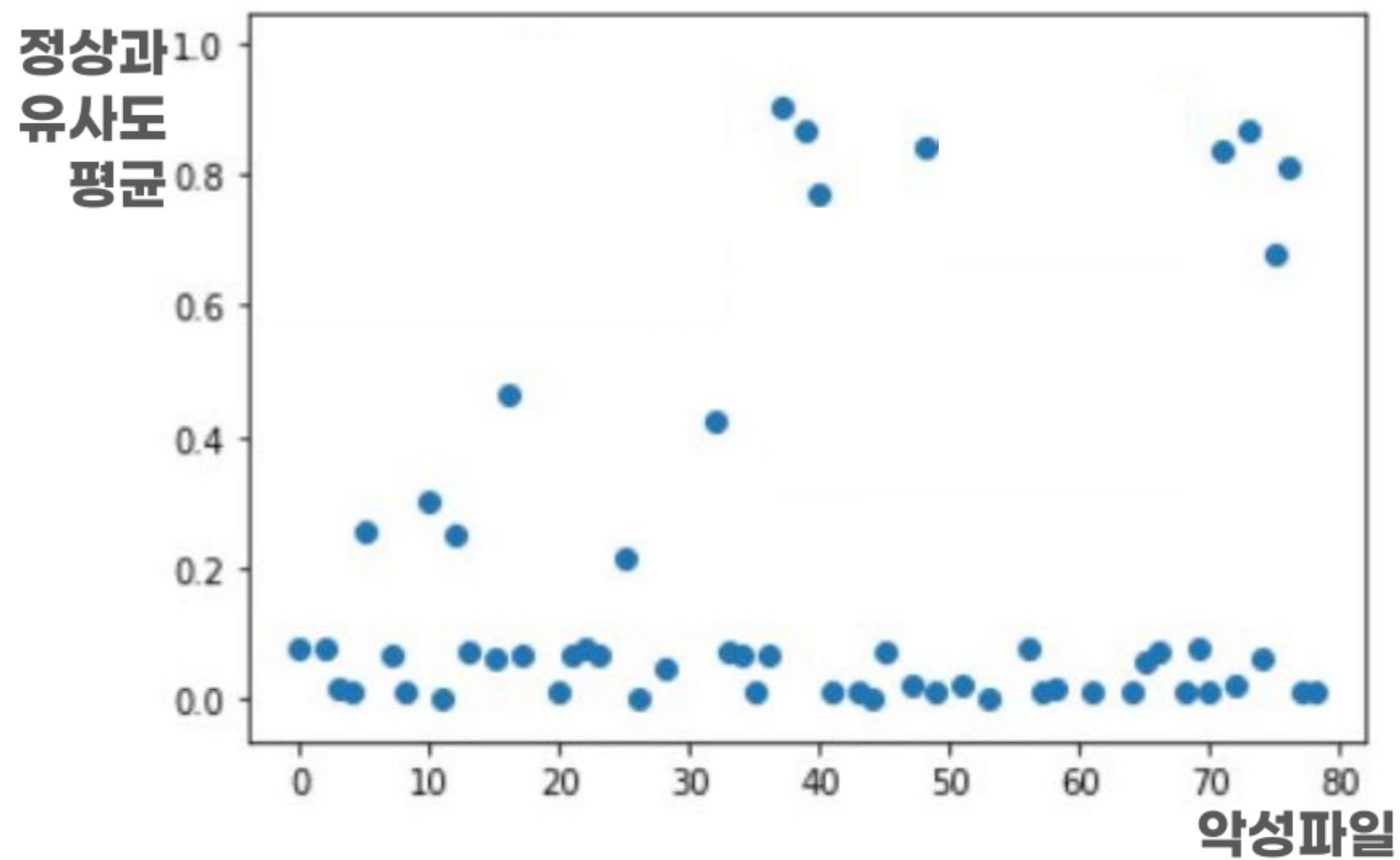


```
ldarg.0  
call  
callvirt  
ldarg.1  
ldfld  
callvirt  
stloc.0  
ldloc.0  
brtrue  
ret  
ldarg.1  
...
```

MD5 : 83b24d182dec262ce606c4ab46894c59
Index : 10977

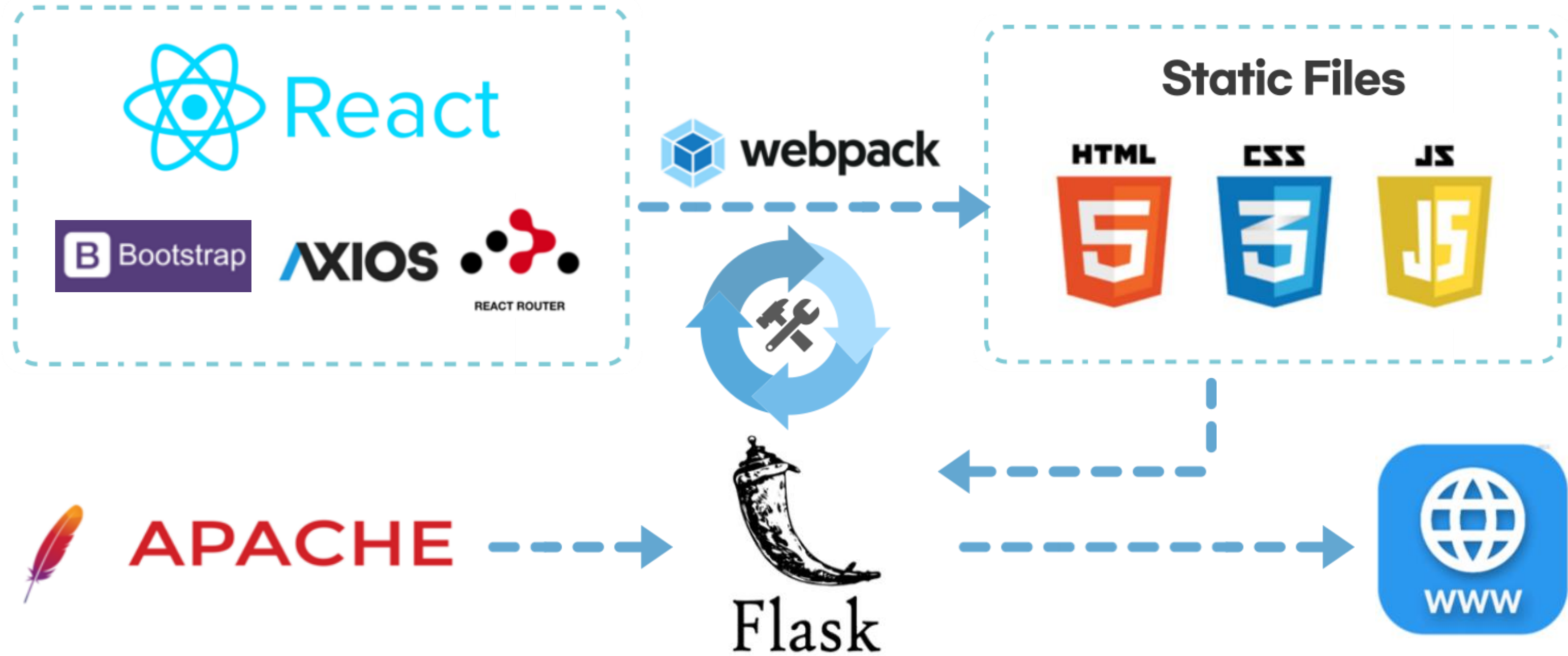
압축벡터 기반 유사도 검사

| 실험 결과



악성 파일간의 유사도가 높음

웹 구현



웹 구현

```

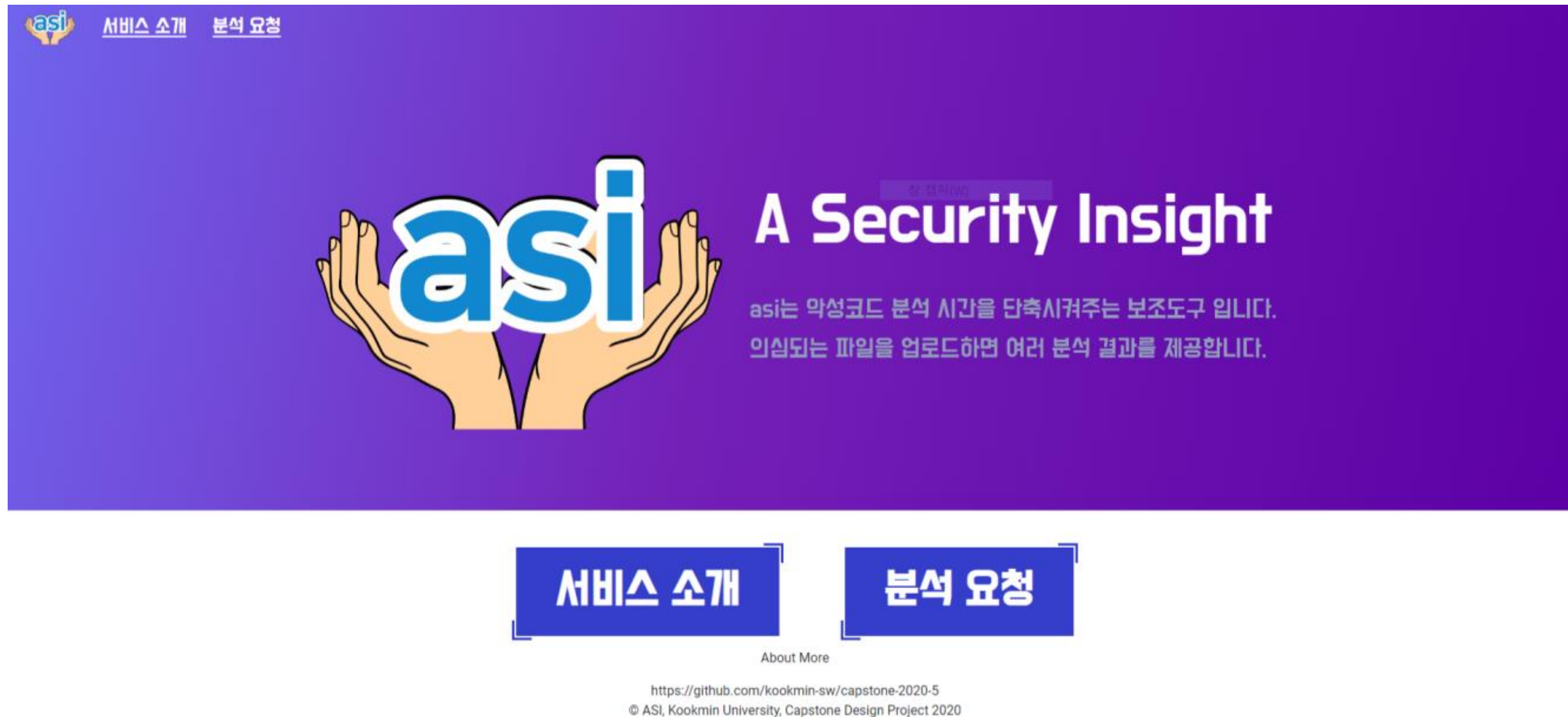
"#FDD5B1" : ["ble", "ble.s", "ble.un", "ble.un.s", "blt", "blt.s", "blt.un.s", "bne.un", "bne.un.s", "br", "br.s"],
"#1A4876" : ["ja", "jb", "jbe", "jecxz", "jg", "jge", "jl", "jle", "jmp", "jnb", "jno", "jnp", "jns", "jnz", "jo", "jp", "js", "jz"],
"#1DACD6" : ["cmp", "cmpsb", "cmpsd", "cmpxchg", "cmpxchg8b", "comisd", "comiss", "switch"],
"#EFDECD" : ["aaa", "aad", "aam", "aas", "adc", "add", "add.ovf", "addpd", "addps", "addsd", "adds", "and", "andnps", "andpd", "andps"],
"#000000" : ["div", "divsd", "divss"],
"#00B9FB" : ["fdiv", "fdivp", "fdivr", "fdivrp"],
"#4CB7A5" : ["fsub", "fsubp", "fsubr", "fsubrp"],
"#D68A59" : ["sub", "sub.ovf", "subpd", "subps", "subsd", "subss", "idiv", "imul"],
"#B4674D" : ["fcom", "fcomi", "fcomip", "fcomp", "fcompp", "fucom", "fucomi", "fucomip", "fucomp", "fucompp"],
"#DD9475" : ["pand", "test", "pandn", "pxor", "xor", "xorpd", "xorps", "xor", "xorpd", "xorps"]

```

...

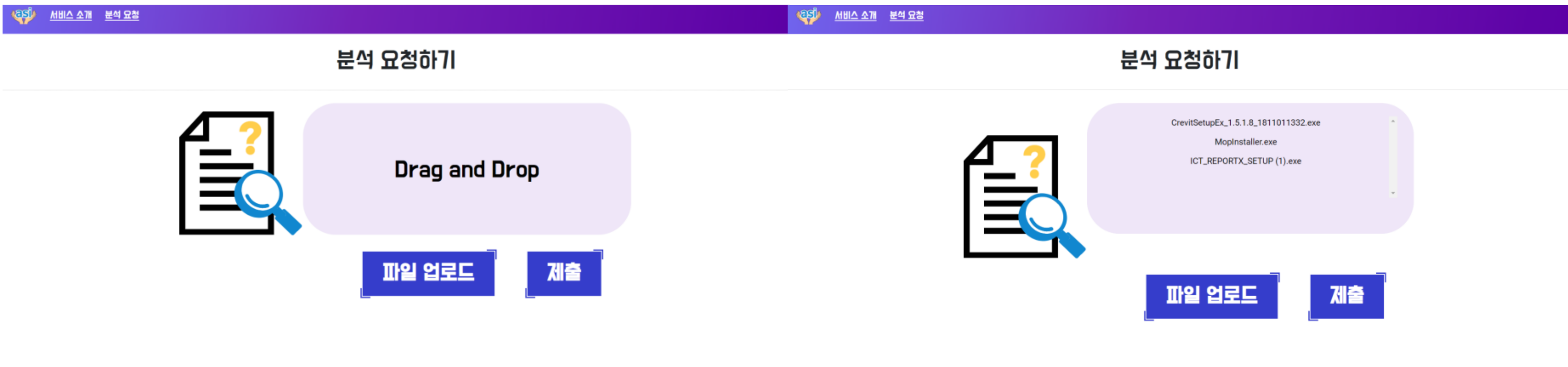
니모닉 계열 별 컬러

웹 구현 - 메인 페이지




메인페이지 추가로 UI개선

웹 구현 - 업로드









드래그 앤 드롭 방식 & 여러파일 업로드 가능

웹 구현 - 업로드 목록

 [서비스 소개](#) [분석 요청](#)

Upload / File List

Download ZIP

파일명	결과	정보
 00d7ab9a8e5c9a84cfa19ad9e583e6f.exe	결과 보기	@meta
 00d9dada816ad75caf97f9e09de2a521.exe	결과 보기	@meta
 00d8914ba4c475e568a292afbd7b35d1.exe	결과 보기	@meta
 00dde3d759f73ba093da9a375d725f47.exe	결과 보기	@meta
 0000f5c78ed2442813ceef6afaf436c3.exe	결과 보기	@meta
 00d2c06a552f782c1f16acf77db765a5.exe	결과 보기	@meta

List of users

업로드 파일 별 결과 확인

웹 구현 - 분석 결과


[서비스 소개](#)
[분석 요청](#)



05a00e66bc0d98a777f8c34e922274d1.exe
md5 : 05a00e66bc0d98a777f8c34e922274d1
sha256 : 4C03F4178F010AE68DBD1894C6F49CBE82F31334307EEC8A04C3BB713956E141
file size : 319909

[유사도 검사](#)
[이상탐지](#)

Function

☞0041103a95122ed285027277d434cb03	유사도 검사
☞02b0bbd0d4d6b3eec24b36926395145d	유사도 검사
☞02b7a586ceb26b9b1860df1b9918f794	유사도 검사
☞07024b5520aaf7f784aaeada57c87b96	유사도 검사
☞08bbc85b8676b23bd59144dba2b348c0	유사도 검사

메타 데이터 표시

함수단위로 disassemble

웹 구현 - 분석 결과

유사도 검사 이상탐지

Function

0041103a95122ed285027277d434cb03

유사도 검사

정상

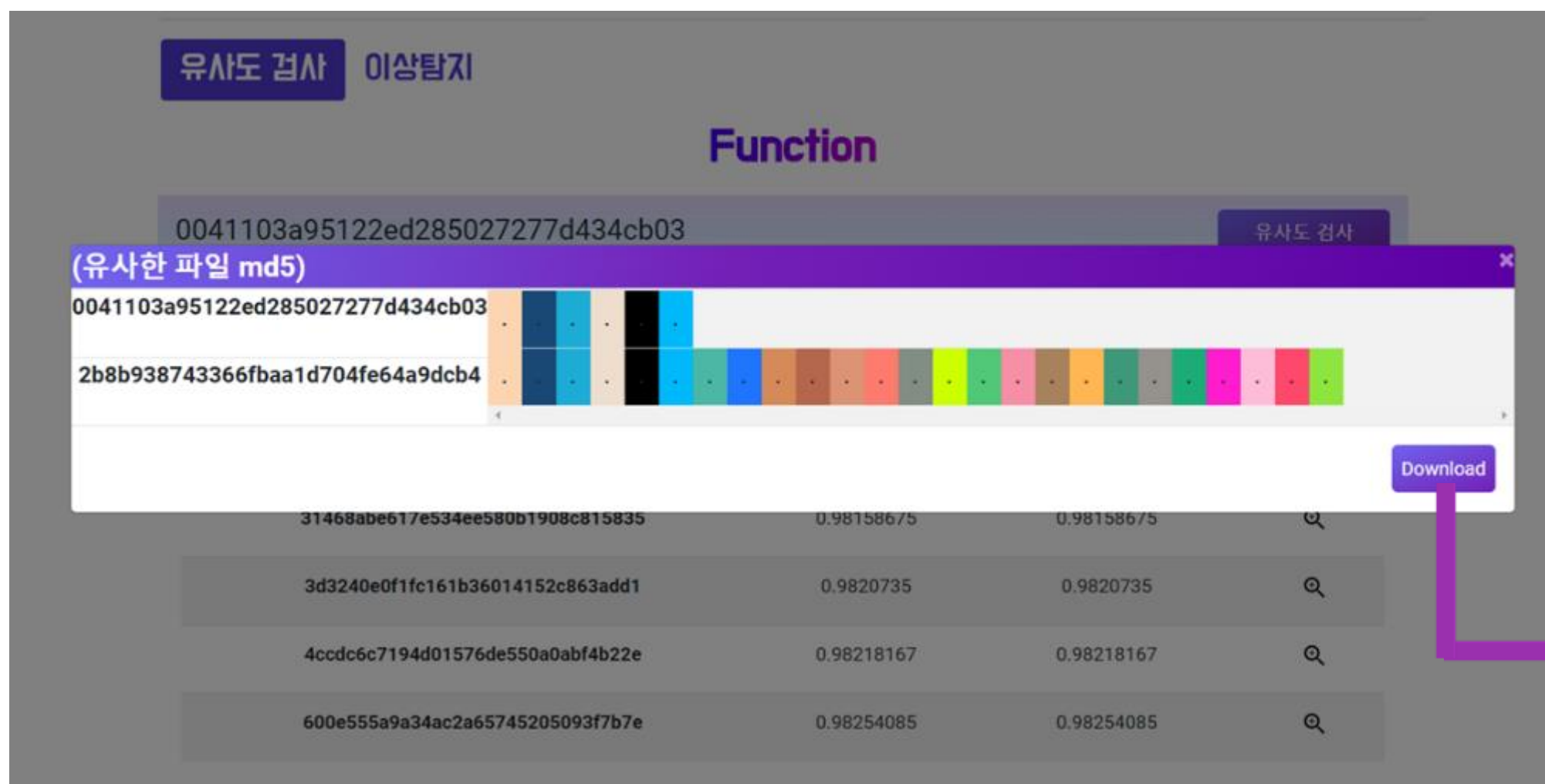
악성

File(md5)	Function(md5)	Cosine	Jaccard-distance	Details
2b8b938743366fbaa1d704fe64a9dcb4	24ee3823b3b5af887fc3a889f12a48dd	0.98153335	0.98153335	🔍
31468abe617e534ee580b1908c815835	dfc9f2c352ad3b2d4adcf5ffe7e00308	0.98158675	0.98158675	🔍
3d3240e0f1fc161b36014152c863add1	81849e90aa0da1f885548c8011e6fc04	0.9820735	0.9820735	🔍
4ccdc6c7194d01576de550a0abf4b22e	04edaa7d2b874d62284814d8ea7528bc	0.98218167	0.98218167	🔍
600e555a9a34ac2a65745205093f7b7e	700556142f7fb220a8661d90b861e251	0.98254085	0.98254085	🔍

유사한 파일 중 정상 악성 비율 표시

유사한 함수를 포함하는 파일 검색

웹 구현 - 분석 결과



함수 간 유사성
니모닉 계열 컬러로 시각화

함수 니모닉 다운로드 기능

웹 구현 - 분석 결과



05a00e66bc0d98a777f8c34e922274d1.exe

md5 : 05a00e66bc0d98a777f8c34e922274d1

sha256 : 4C03F4178F010AE68DBD1894C6F49CBE82F31334307EEC8A04C3BB713956E141

file size : 319909

유사도 검사

이상탐지

Function

94bff5a8d74f83f027287026841db878

유사도 검사

이상탐지 탭

이상탐지된 함수만 표시



01

프로젝트 목표

02

진행 상황

03

계획 및 제한요소

계획

| 검증



한계점

분석 결과에 대한
신빙성있는 검증방식 도입 필요



고려사항

악성코드 보고서 분석
분류기로써의 성능 확인

월별 구현 계획

항목	세부내용	1월	2월	3월	4월	5월	6월
요구사항분석	요구 분석	☑					
	SRS 작성	☑					
관련분야연구	딥러닝 기술 연구		☑	☑			
	관련 논문 동향조사		☑	☑			
설계	시스템 설계				↻	↻	
구현	코딩 및 모듈 테스트				↻	↻	
테스트	시스템 테스트						☑

팀원 별 역할 분담



손현기

크롤러 & 파서 개발
신경망 구현 및 튜닝, 임베딩
웹 백엔드 개발
ELK 구축



김주환

논문 동향조사
제안서 및 보고서 작성
단어 임베딩



김호준

자료 조사
회의록 등 문서 작성
웹 프론트 개발



오예린

발표자료 등 디자인
웹 기획/ 퍼블리싱
ELK 구축



이동운

정상파일 크롤러 개발
신경망 구현 및 튜닝



Ruslan

opcode 파서 개발
웹 프론트 개발





감사합니다.

