

NLP기반 powershell
악성스크립트
탐지 툴 개발

조원 김서영 / 박지연 / 최연재



01 배경 및 필요성



PowerShell?

- Microsoft가 개발한 스크립팅 언어
- .NET Framework에 기반하여 개발됨
- 비교적 짧은 소스코드로 효과적인 성능을 낼 수 있어 사이버 공격자의 공격도구로 사용됨



PowerShell



Key trends: Cybercriminals pivot, taking on new strategies and tactics

In Q4 of 2017, McAfee Labs recorded on average eight new malware samples per second—an increase from four new samples per second in Q3. Overall, the quarter was characterized by newer tools and schemes, such as PowerShell malware and cryptocurrency mining, which surged along with the value of Bitcoin.

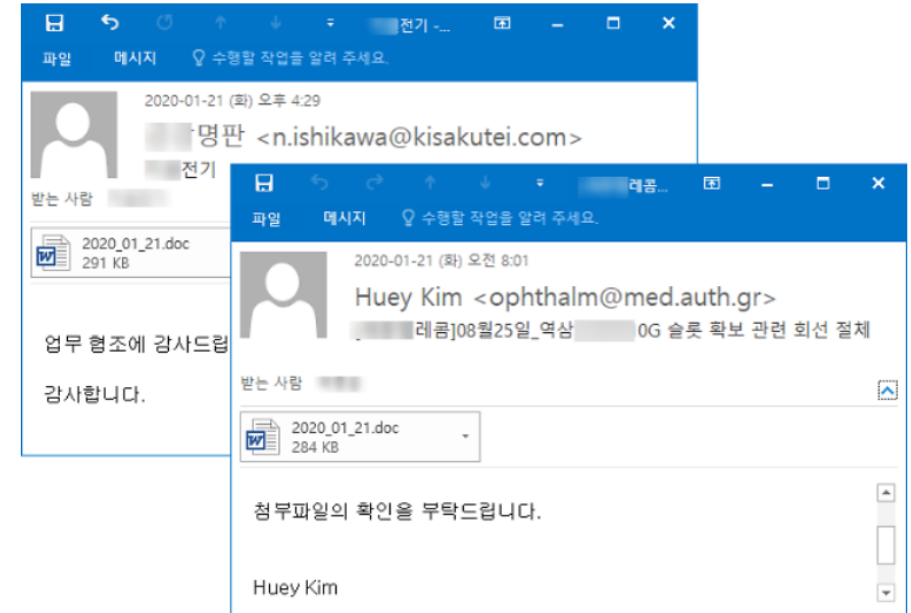
PowerShell: In 2017, McAfee Labs saw PowerShell malware grow by 267% in Q4, and by 432% year over year, as the threat category increasingly became a go-to toolbox for cybercriminals. The scripting language was irresistible, as attackers sought to use it within Microsoft Office files to execute the first stage of attacks.

In December, [Operation Gold Dragon](#), a malware campaign targeting the 2018 Winter Olympics, was uncovered. The campaign is an exemplary implementation of PowerShell malware in an attack.

출처 : McAfee labs threats report march 2018

Powershell 이용 악성코드 증가 현황 기사


악성코드 이모넷 예시



→ 2014년부터 Powershell이 악성코드에서 공격도구로서 활용되기 시작


출처 : 보안뉴스

→ 2017년 Powershell을 이용한 악성코드 증가 현황 기사 (Emotet)'의 국내 대량 유포 사건



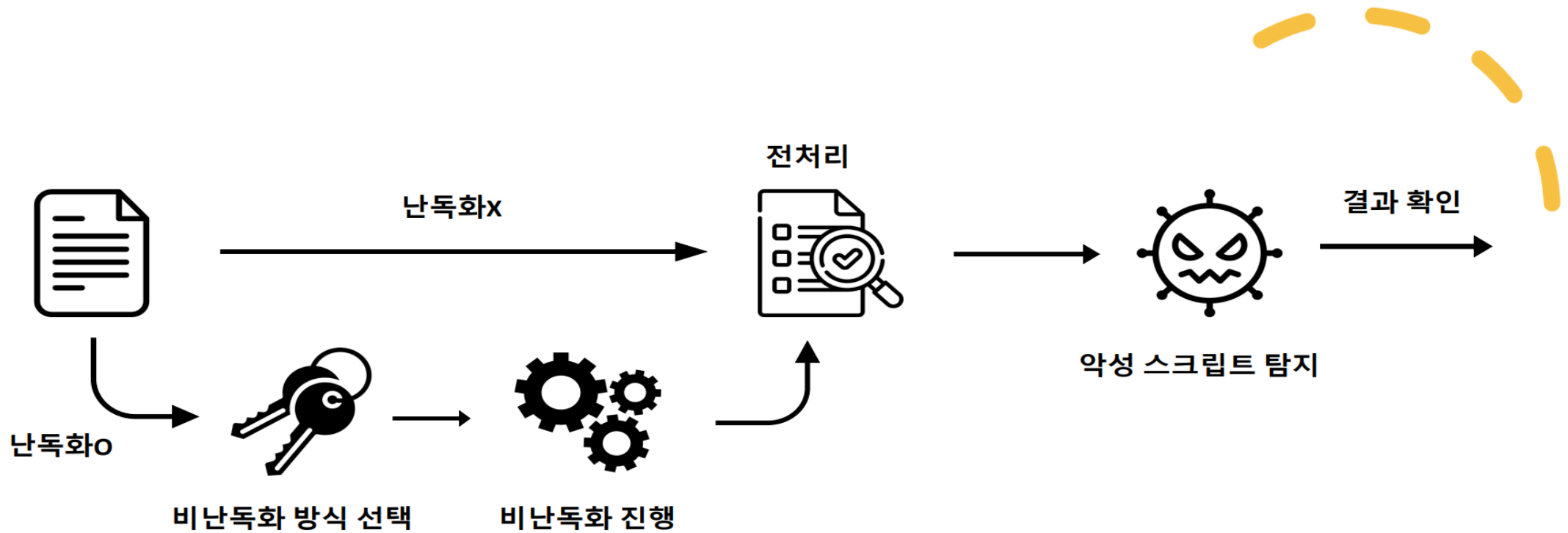
⚠ Powershell의 유연한 기능이 다양한 난독화 가능하게 하기
때문에 기존의 시그니처 기반 탐지를 어렵게 함

⚠ 사이버 공격자들이 지속적으로 탐지 기술 우회방안을 찾아 기존
방식으로 탐지가 어려움

- ✓ 난독화된 Powershell 기반 악성스크립트 비난독화 모듈 개발
 - ✓ 비난독화는 Base64, 16진수 방식 이용
 - ✓ 머신러닝을 이용한 모델 활용하여 악성 스크립트 탐지
- 

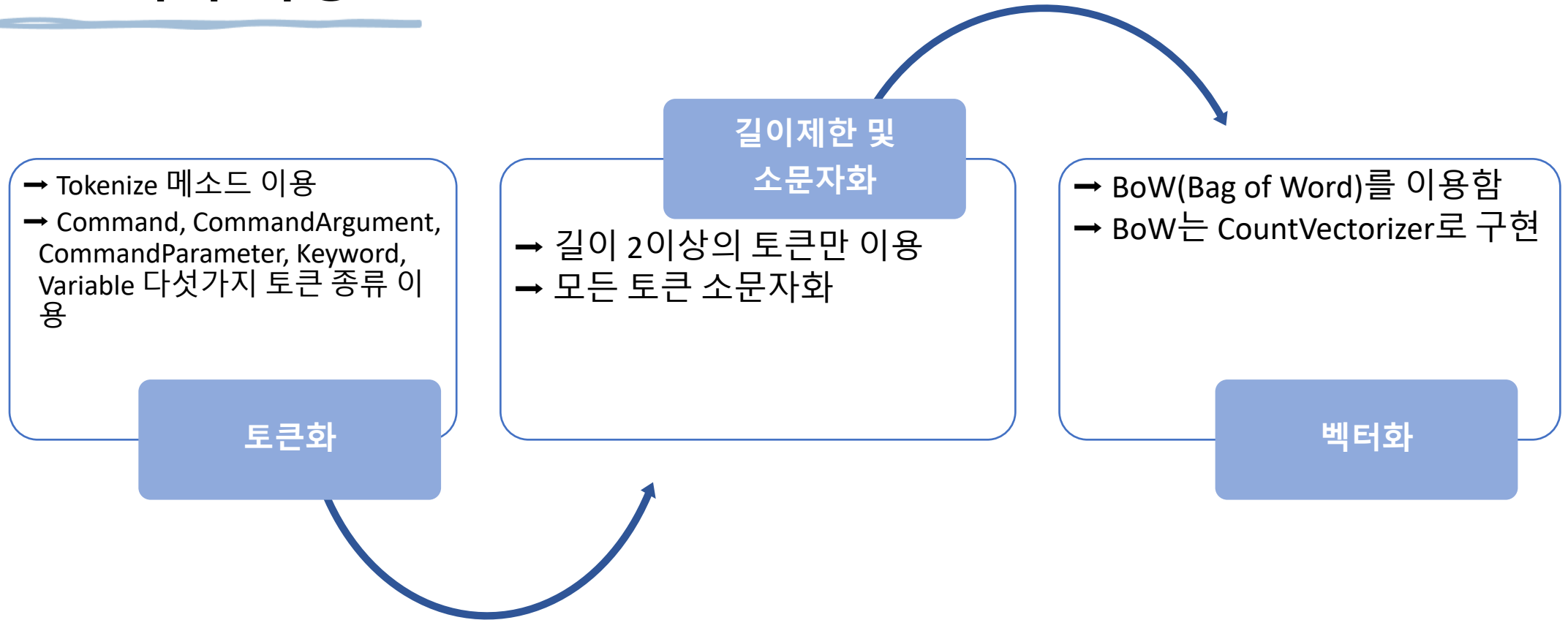
02 하름도





- 비난독화 모듈과 머신러닝을 활용한 악성스크립트 분류 모델, GUI 개발
- 비난독화는 Base64, 16진수 방식 이용
- 앙상블 모델은 Boosting 기법을 이용한 AdaBoost, GBM, XGBoost 이용

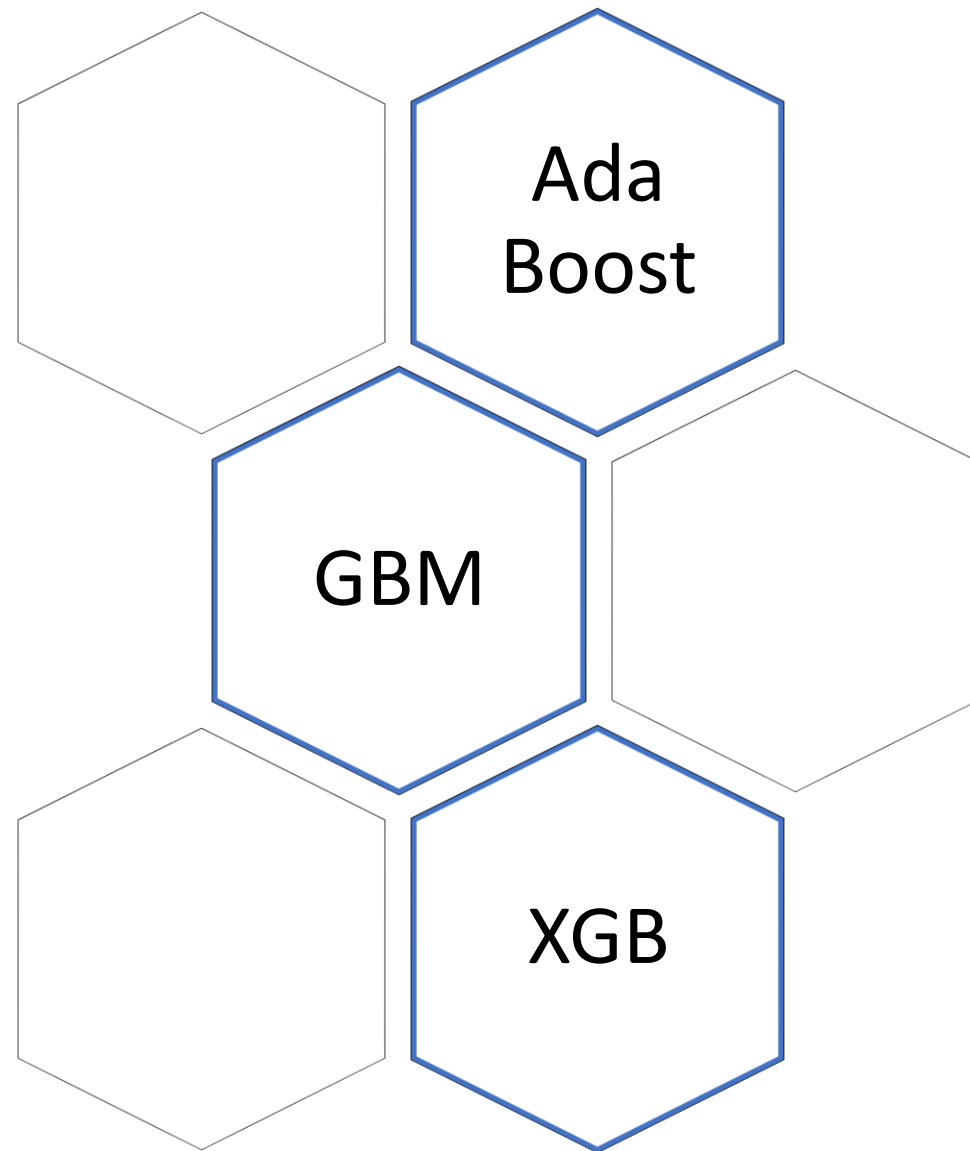
전처리 과정



앙상블 모델 구성

앙상블? 여러개의 분류기를 생성하고 예측 결과를 결합하여
보다 정확하고 신뢰성이 높은 예측값을 도출하는 기법

- Boosting 방식을 사용하는 모델로 구현
- AdaBoost : Boosting 기법 중 가장 기본적인 모델
- GBM : Gradient Descent 기법 이용
- XGBoost : GBM과 같은 기법 이용, 손실함수도 고려




03


연구 결과






모델	정확도	정밀도	재현율	F1
RF	0.979	1.00	0.68	0.809
GBM	0.986	0.981	0.8	0.881
Ada	0.984	0.922	0.83	0.874
XGB	0.988	0.977	0.84	0.903

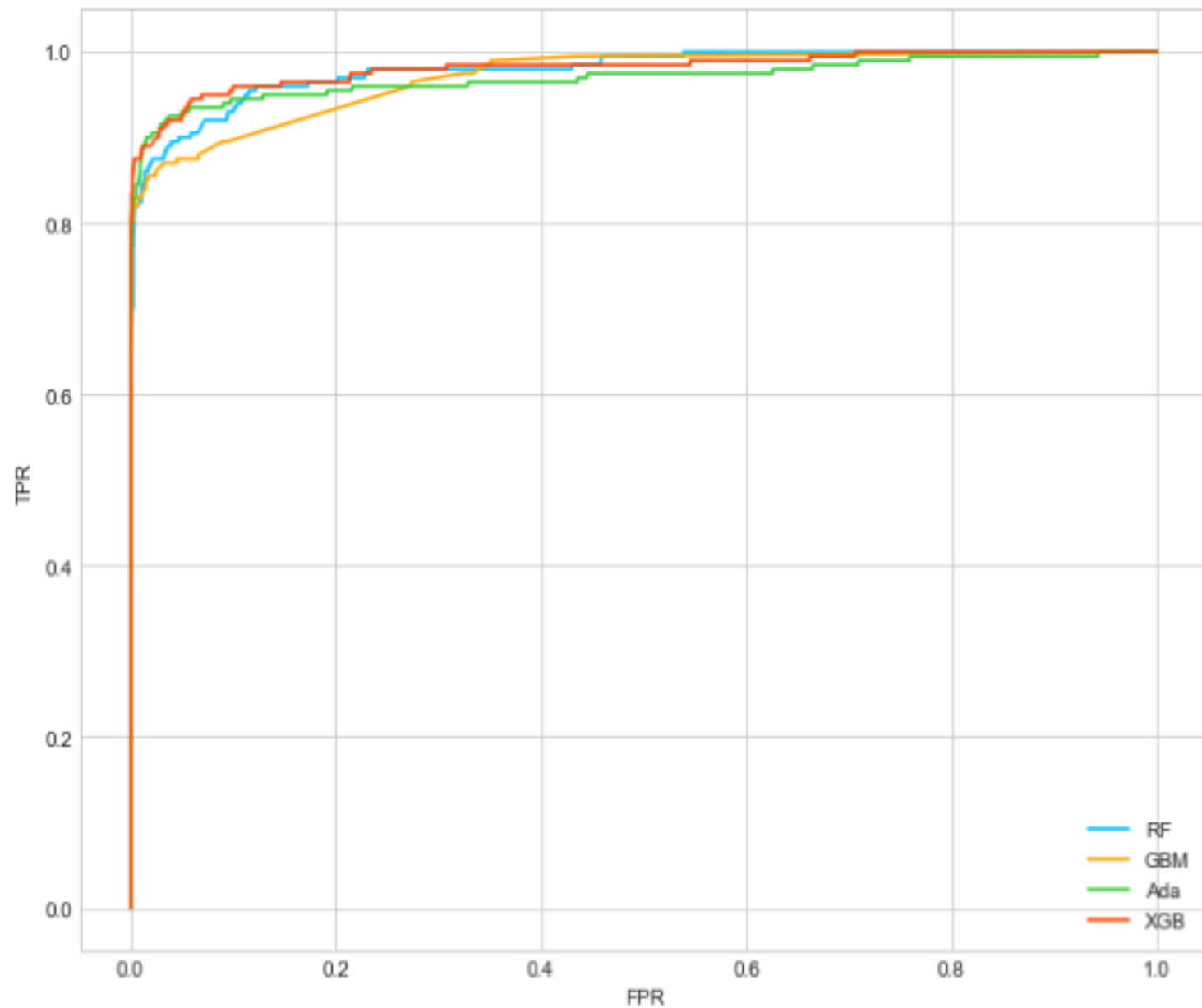
- 
- 정밀도 : 모델이 악성이라 예측한 데이터 중 실제 악성
 - 재현율(TPR) : 실제 악성데이터 중 모델이 악성이라 판단
 - F1 : 정밀도와 재현율의 조화 평균



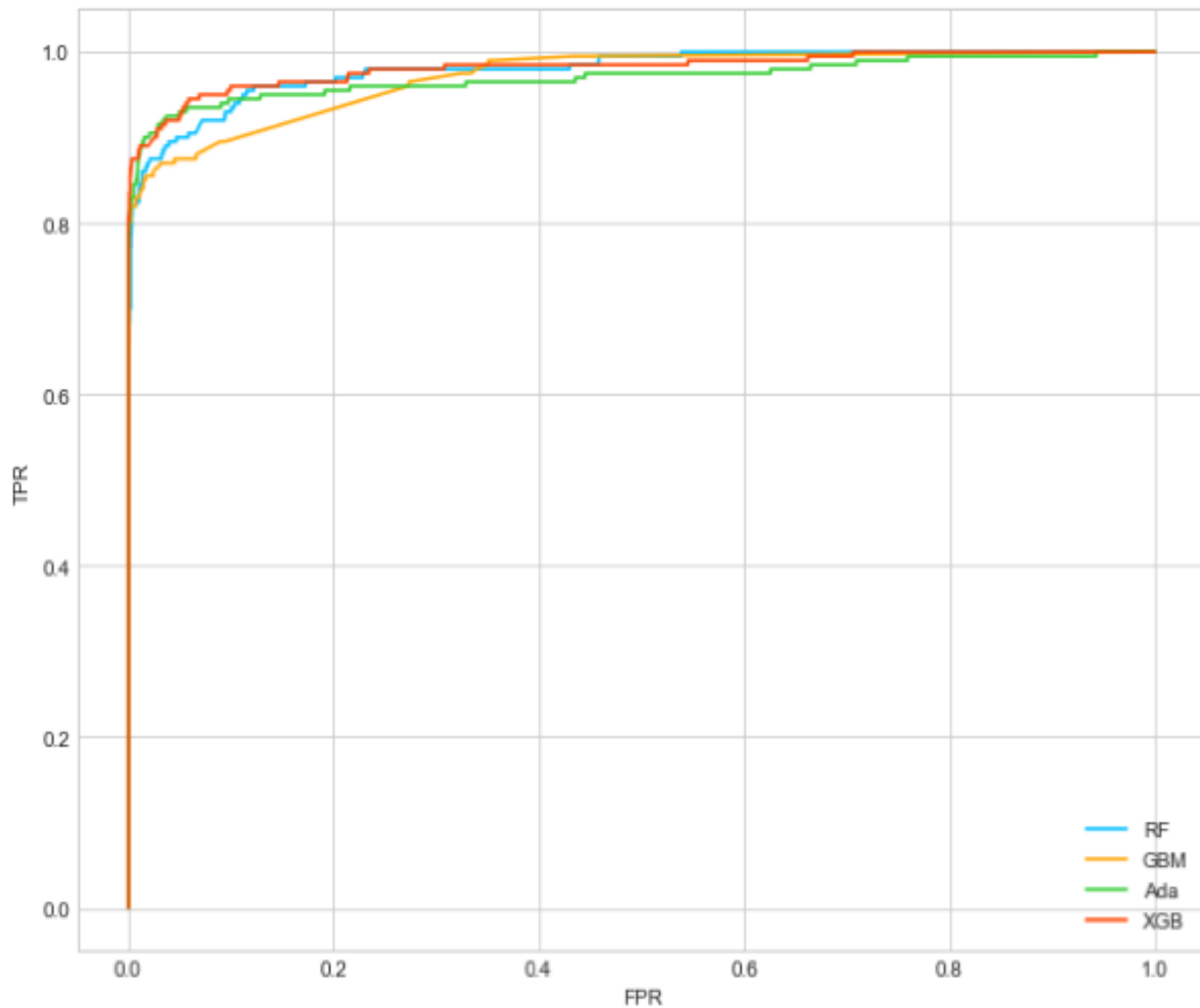
모델	정확도	정밀도	재현율	F1
RF	0.979	1.00	0.68	0.809
GBM	0.986	0.981	0.8	0.881
Ada	0.984	0.922	0.83	0.874
XGB	0.988	0.977	0.84	0.903

- 
- 정밀도 : 모델이 악성이라 예측한 데이터 중 실제 악성
 - 재현율(TPR) : 실제 악성데이터 중 모델이 악성이라 판단
 - F1 : 정밀도와 재현율의 조화 평균

모델	AUC
GBM	0.97
Ada	0.97
XGB	0.981
RF	0.976



모델	AUC
GBM	0.97
Ada	0.97
XGB	0.981
RF	0.976





모델	정확도	정밀도	재현율	F1
GBM + XGB	0.989	0.988	0.84	0.908
GBM + Ada	0.986	0.987	0.80	0.883
Ada + XGB	0.988	0.976	0.84	0.903
GBM + Ada + XGB	0.989	0.988	0.84	0.908

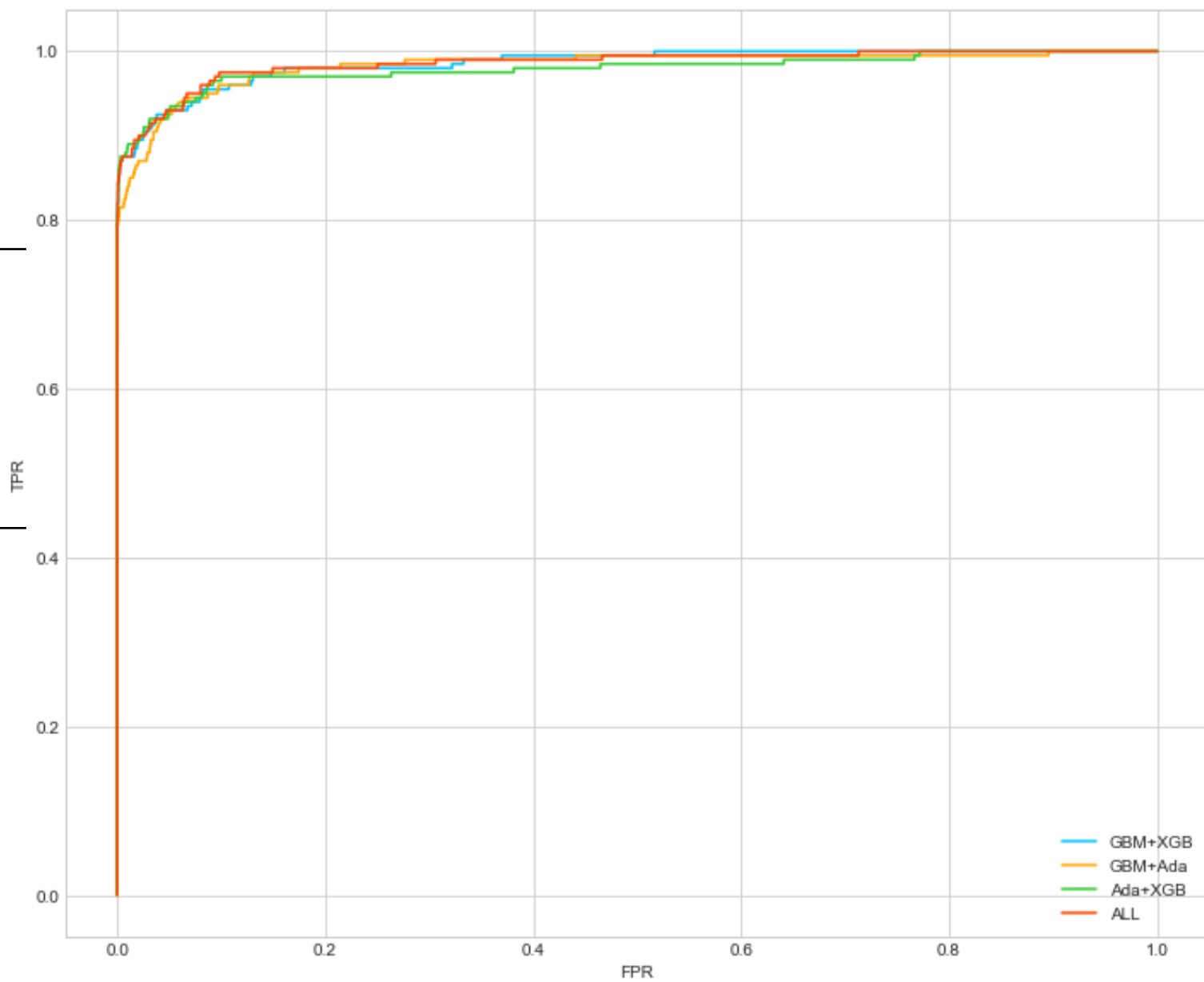
- 정밀도 : 모델이 악성이라 예측한 데이터 중 실제 악성
- 재현율(TPR) : 실제 악성데이터 중 모델이 악성이라 판단
- F1 : 정밀도와 재현율의 조화 평균



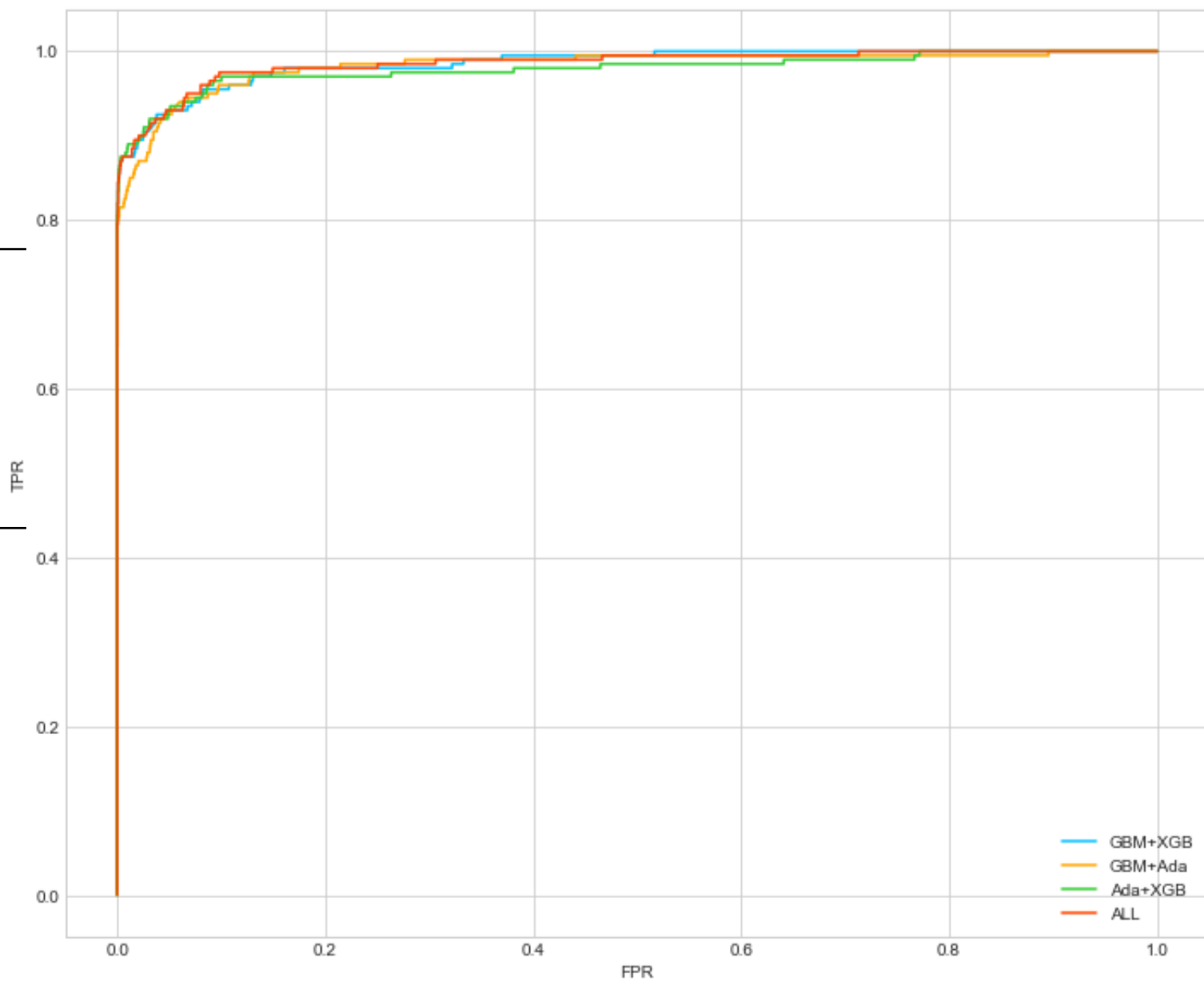
모델	정확도	정밀도	재현율	F1
GBM + XGB	0.989	0.988	0.84	0.908
GBM + Ada	0.986	0.987	0.80	0.883
Ada + XGB	0.988	0.976	0.84	0.903
GBM + Ada + XGB	0.989	0.988	0.84	0.908

- 정밀도 : 모델이 악성이라 예측한 데이터 중 실제 악성
- 재현율(TPR) : 실제 악성데이터 중 모델이 악성이라 판단
- F1 : 정밀도와 재현율의 조화 평균

모델	AUC
GBM + XGB	0.9853
GBM + Ada	0.9829
Ada + XGB	0.9789
GBM + Ada + XGB	0.9867



모델	AUC
GBM + XGB	0.9853
GBM + Ada	0.9829
Ada + XGB	0.9789
GBM + Ada + XGB	0.9867



04 시연 영상



An abstract composition of various geometric shapes. In the top left, a green-outlined triangle points right. To its right is a solid blue circle. Below the triangle is a blue-outlined circle. In the center is a large orange semi-circle. To the right of the semi-circle is a vertical yellow dashed line. In the bottom left is a large solid orange circle. Above it are three short, curved yellow dashes. In the bottom right is a green-outlined square.