# 머신러닝 및 딥러닝 앙상블 모델을 활용한 피싱사이트 탐지 연구

Phishing site detection research using machine learning and deep learning ensemble models

저자: 김주령, 김현진, 김해찬, 박찬민, 이경림, 이소영, 홍혜원

**2024**

# 목차



## Abstract

최근 웹 사이트 내 악성코드가 심각한 보안 문제로 대두되고 있다. 피싱 사이트의 수법이 다양해지고 있을 뿐만 아니라 개인은 물론 기업도 피해 대상이 되어 피해액은 기하급수적으로 증가하고 있다. 본 논문에서는 피싱 사이트의 URL 주소와 콘텐츠 특징을 이용해 머신러닝과 딥러닝 앙상블 모델을 학습시켜, 해당 사이트가 피싱 사이트인지 정상 사이트인지 쉽게 탐지할 수 있는 방법을 제시한다. 악성 확률을 계산하여 결과를 시각화하는 방법으로 웹 사이트를 제작하였으며, 머신러닝 모델로는 LGBM, 딥러닝 모델로는 CNN을 선택하여 앙상블을 진행한 결과, 0.999의 정확도를 나타냈다. 본 연구를 통해 개인과 기업의 피해를 줄이는 데 기여할 수 있을 것으로 기대한다.

# 제 1장 서론

## 1.1 피싱 개념 소개

피싱(Phishing)은 사용자를 속여 민감한 정보를 도용하기 위한 사이버 공격 기법 중 하나로, 주로 이메일, 문자 메시지, 가짜 웹사이트 등을 통해 이루어진다. 피싱 사이트는 정상 사이트와 외관이 거의 동일하게 만들어져 사용자들이 이를 구분하기 어려워, 대규모 경제적 손실과 사회적 불안정을 초래할 수 있다.

## 1.2 피싱 사이트의 발전 및 위험성

최근 몇 년간 피싱 사이트는 급격하게 증가하며 더욱 정교해지고 있다. 피싱 공격은 다양한 형태로 진화하고 있으며, 공격자들은 탐지 기술을 회피하기 위한 새로운 수법을 끊임없이 개발하고 있다. 예를 들어, 웹사이트의 보안 인증서를 위조하거나, 최신 인공지능 기술을 이용해 실제 사용자와의 대화를 흉내내는 등의 방법으로 발전하고 있다. 이러한 피싱 공격의 발전으로 인해 기존의 블랙리스트 방식으로는 전체 피싱 공격의 20%를 탐지하지 못해 여전히 해결하기에 어려움이 있다. 피싱 공격으로 인한 경제적 피해는 심각한 수준에 이르렀으며, 일부 국가에서는 피싱 사이트 제작을 지원하는 사례도 나타나고 있다. 피싱 사이트로 인한 피해는 개인뿐만 아니라 대기업, 금융 기관, 정부 기관으로 확대되고 있어 그 위험성은 점점 더 커지고 있다. 특히, 기업의 경우 피싱으로 인한 데이터 유출과 재정적 손실은 기업 신뢰도 하락과 심각한 경제적 타격을 초래할 수 있다.

## 1.3 연구 필요성 및 목적

이러한 배경에서 피싱 사이트를 효과적으로 탐지하고 방지하기 위한 연구의 필요성은 시간이 지날수록 중요해지고 있다. 기존 탐지 기법들이 발전하는 피싱 공격 기법을 충분히 탐지하지 못하는 상황에서, 새로운 기술을 통해 피싱 사이트를 더 정확하게 탐지할 수 있는 방법이 요구되고 있다. 본 연구에서는 피싱 사이트의 주요 특징으로 알려진 URL의 차이, 보안 인증서 차이, 콘텐츠 차이를 활용하여 머신러닝 및 딥러닝 앙상블 모델을 학습시켜 탐지 모델을 개발하는 데 중점을 두었다. 이 연구는 피싱 공격에 대한 정교한 대응 방안을 제시하며, 향후 사이버 보안 연구에 중요한 기여를 할 것으로 기대된다.

# 제 2장 본론

## 2.1 피처 수집

본 연구에서 정상 URL과 피싱 URL을 구분하는 피처는 총 14개로, URL 내에서 직접 얻을 수 있는 특성(IP_LIKE, AT, URL_Depth, Redirection, Is_Https, Tiny_url, Check_Hyphen, Query)과 도메인 관련 특성(is_domain_created, Domain_age, Domain_end, Mouse_Over, Web_forwards, Hyperlinks, Domain_cons, URL_length, HTTP_Code, HTTP_Status)이 포함된다.

AT: URL 내 '@' 기호의 존재 여부를 확인하는 피처로, '@' 기호는 URL에서 드물게 사용되며, 피싱 사이트는 이를 활용해 사용자를 속이는 경우가 많기 때문에 '@'가 포함된 URL은 피싱 사이트일 가능성이 크다.

Redirection: URL 내에서 리다이렉션이 여러 번 발생하는지를 탐지하는 피처로, 피싱 사이트는 사용자를 다른 사이트로 이동시키기 위해 리다이렉션을 자주 사용하기 때문에 리다이렉션이 많을수록 피싱일 가능성이 높다.

Tiny_url: 단축 서비스(e.g., bit.ly, t.co 등)를 사용하여 생성된 URL을 탐지하는 피처로, 피싱 공격자는 단축 URL을 사용해 악성 웹사이트로 연결되는 링크를 숨기려 하기 때문에 이런 URL은 피싱 위험이 높다.

Query: URL의 쿼리 문자열 부분을 분석하는 피처로, 피싱 사이트는 쿼리 문자열에 악의적인 코드를 삽입하거나 사용자를 속이기 위한 정보를 포함시킬 수 있기 때문에 의심스러운 쿼리 패턴이 감지되면 피싱으로 판단할 수 있다.

Domain_age: 도메인이 생성된 후 경과한 시간을 측정하는 피처로, 피싱 사이트는 대개 단기간에 만들어져 빠르게 사라지므로, Domain_age가 작을수록 피싱 위험이 크다.

Mouse_Over: 마우스를 URL 위에 올렸을 때 나타나는 링크와 실제 연결되는 링크가 일치하는지 여부를 탐지하는 피처로, 피싱 사이트는 두 링크를 다르게 설정해 사용자를 속이려 하기 때문에 불일치가 감지되면 피싱 가능성이 높다.

Hyperlinks: 웹페이지 내에 외부 링크가 과도하게 포함되었는지를 탐지하는 피처로, 정상적인 웹사이트는 내부 링크와 외부 링크가 균형을 이루는 반면, 피싱 사이트는 외부 링크를 과도하게 포함하는 경향이 있기 때문에 외부 링크가 많으면 피싱으로 의심할 수 있다.

URL_length: URL의 전체 길이를 측정하는 피처로, 피싱 URL은 종종 복잡하고 길기 때문에 URL이 비정상적으로 길 경우 피싱 가능성이 높다고 판단할 수 있다.

IP_LIKE: URL에 IP 주소 형식이 포함되어 있는지를 확인하는 피처로, IP 주소는 일반적으로 정상적인 웹사이트에서 사용되지 않지만, 피싱 사이트는 도메인 등록 비용을 절감하거나 추적을 피하기 위해 IP 주소를 사용하는 경우가 많다. 따라서, URL에 IP 주소가 포함된 경우 피싱 사이트일 가능성이 크다.

URL_Depth: URL의 깊이를 측정하는 피처로, URL 내에서 슬래시('/')의 개수를 세어 경로의 깊이를 계산한다. 피싱 사이트는 사용자를 혼란스럽게 만들기 위해 복잡하고 깊은 경로를 사용하는 경우가 많기 때문에, URL의 깊이가 깊을수록 피싱 사이트일 가능성이 높다.

Is_Https: URL이 HTTPS 프로토콜을 사용하는지를 확인하는 피처로, HTTPS는 데이터를 암호화하여 전송하므로 보안성이 높다. 피싱 사이트는 HTTPS를 사용하지 않거나, 유효하지 않은 인증서를 사용할 가능성이 높다. 따라서, HTTPS가 아닌 URL은 피싱 사이트일 가능성이 있다.

Check_Hyphen: URL에 하이픈('-')이 포함되어 있는지를 확인하는 피처로, 하이픈은 정상적인 도메인 이름에서는 자주 사용되지 않는다. 그러나 피싱 사이트는 하이픈을 추가하여 정상 도메인처럼 보이도록 유도하는 경우가 많기 때문에, 하이픈이 포함된 URL은 피싱 사이트일 가능성이 높다.

is_domain_created: 도메인의 생성 날짜를 확인하는 피처로, 도메인이 최근에 생성되었는지를 판단한다. 피싱 사이트는 주로 새로 생성된 도메인을 사용하여 짧은 기간 동안 운영되므로, 도메인이 최근에 생성된 경우 피싱 사이트일 가능성이 높다.

Domain_end: 도메인의 만료일을 확인하는 피처로, 도메인의 만료일이 가까운 경우 피싱 사이트일 가능성이 크다. 피싱 사이트는 주로 짧은 기간 동안만 운영되기 때문에 도메인 만료일이 6개월 이내인 경우 피싱 사이트로 판단할 수 있다.

Web_forwards: URL이 여러 번 리디렉션되었는지 여부를 확인하는 피처로, 정상적인 웹사이트는 리디렉션이 거의 없거나 적지만, 피싱 사이트는 사용자를 악성 웹사이트로 유도하기 위해 다수의 리디렉션을 사용하는 경우가 많다. 따라서, 리디렉션 횟수가 많으면 피싱 사이트일 가능성이 높다.

Domain_cons: URL의 원본 도메인과 최종 리디렉션된 도메인이 일치하는지 여부를 확인하는 피처로, 피싱 사이트는 사용자를 속이기 위해 원본 도메인과 다른 도메인으로 리디렉션할 수 있다. 원본 도메인과 최종 도메인이 일치하지 않으면 피싱 사이트일 가능성이 있다.

HTTP_Code: 웹 서버가 URL 요청에 대해 반환하는 HTTP 상태 코드를 분석하는 피처로, 정상적인 웹사이트는 일반적으로 200 OK 상태 코드를 반환하지만, 피싱 사이트는 비정상적인 상태 코드나 리디렉션 관련 코드(예: 3xx)를 반환할 수 있다. 비정상적인 HTTP 상태 코드는 피싱 사이트의 가능성을 높인다.

HTTP_Status: HTTP 상태 코드와 연관된 추가적인 정보를 기반으로 URL의 정상 여부를 판단하는 피처로, 정상적인 웹사이트는 보통 일관된 상태 코드를 반환하지만, 피싱 사이트는 다양한 상태 코드를 통해 사용자에게 혼란을 줄 수 있다. HTTP 상태 코드의 비일관성이나 비정상적인 코드가 발견되면 피싱 사이트일 가능성이 높다.

## 2.2 머신러닝 모델(LGBM)

LGBM(Light Gradient Boosting Machine)은 대규모 데이터 처리에 최적화된 머신러닝 모델로, Gradient Boosting 알고리즘의 한 종류이다. LGBM은 기존의 레벨 중심 방식이 아닌 리프 중심 방식으로 트리를 성장시키며, 트리의 깊이를 증가시켜 과적합을 방지하고 더 나은 성능을 발휘할 수 있다. 또한, 서로 상관관계가 없는 특징들을 하나로 묶어 메모 리 사용량을 줄여 고차원 데이터에서 매우 유리하다. LGBM은 연속적인 데이터를 구간으 로 나누어 히스토그램을 생성하고, 이를 통해 연산을 단순화하고 속도를 높인다. 또한, 학습이 일정 횟수 이상 이루어지지 않으면 중단하는 Early Stopping을 통해 과적합을 방 지하고 학습 시간을 단축시킨다. 이러한 특징들로 인해 LGBM은 분류(Classification), 회귀(Regression), 순위화(Ranking) 등 다양한 문제에 적용할 수 있으며, XGBoost와 더불어 현 재 가장 많이 사용되는 부스팅 계열 알고리즘 중 하나이다.

## 2.3 딥러닝 모델

### 2.3.1 ANN (Artificial Neural Network):

ANN은 입력층, 1개 이상의 은닉층, 출력층으로 구성된 기본적인 신경망 구조다. 입력층에서는 데이터를 입력받고, 출력층에서는 결과를 출력한다. 은닉층에서는 데이터의 특징을 가중치(weight)와 바이어스(bias)를 추출하고 변형한다. 학습 방법으로는 일반적으로 역전파와 경사 하강법과 같은 알고리즘을 활용하여 가중치를 업데이트하며, 정확도를 높인다.

### 2.3.2 DNN (Deep Neural Network):

DNN은 여러 개의 은닉층을 가진 신경망 모델로, 각 은닉층은 이전 은닉층의 출력을 입력으로 받아 학습하고 다시 출력을 다음 은닉층의 입력으로 전달한다. DNN은 ANN과 마찬가지로 역전파와 경사 하강법으로 가중치를 업데이트하며, 은닉층이 많아 데이터셋 의 복잡한 관계를 모델링할 수 있다.

### 2.3.3 CNN (Convolutional Neural Network):

CNN은 주로 이미지의 특징을 추출하는 데 특화된 신경망으로, 합성곱 층, 풀링 층, 완 전 연결층으로 이루어져 있다. 합성곱 층은 커널을 이용해 입력 데이터의 지역적 특징을 감지하고, 풀링 층에서는 데이터의 차원을 축소해 연산량을 줄인다. 마지막으로 완전 연결층에서는 전체 데이터를 통합해 최종 출력을 만든다. CNN은 이미지 분류 뿐만 아니라 텍스트 데이터에서도 뛰어난 성능을 보인다.

## 2.4 데이터셋 구성

본 연구에서는 피싱과 정상 사이트를 구분하기 위해 URL과 콘텐츠를 수집하여 데이터셋을 구성하였다. 데이터셋은 다양한 웹사이트에서 수집한 피싱 및 정상 URL로 구성되며, 각 URL에 대한 특성 데이터를 포함한다. 이 데이터셋은 머신러닝 및 딥러닝 모델의 학습 과 평가에 사용된다. 데이터셋은 훈련셋과 테스트셋으로 나누어져, 모델의 성능을 평가하 는 데 사용되었다.

# 제 3장 결론 및 향후 연구 방향

피싱 사이트의 탐지 문제는 사이버 보안 분야에서 매우 중요한 문제로 대두되고 있으며, 본 연구에서는 URL과 콘텐츠 특징을 기반으로 머신러닝 및 딥러닝 앙상블 모델을 이용해 피싱 사이트 를 탐지하는 방법을 제시하였다. 머신러닝 모델로는 LGBM을 사용하였고, 딥러닝 모델로는 CNN 을 사용하여 앙상블 기법을 통해 성능을 향상시켰다. 결과적으로 제안된 모델은 0.999의 높은 정 확도를 기록하여 피싱 사이트 탐지에 효과적임을 입증하였다. 이 연구를 통해 사이버 보안 위협 에 대한 대응력을 높이고, 사용자와 기업의 피해를 줄이는 데 기여할 수 있을 것으로 기대한다.

# 제 4장 참고문헌


[1] 권현, "Design do detection method for malicious URL based on Deep Neural Network", 2021

[2] 강태화, "CNN 기반 피싱 사이트 탐지 시스템 구현", 2023

[3] 서희수, "피싱 웹사이트 탐지를 위한 신경망과 진화연산 기반 URL 특징 최적화 알고리즘의 결합", 2020

[4] JURNAL EMACS, "Phishing Site Detection Classification Model Using Machine Learning Approach", 2023

[5] Shweta Singh, "A Deep Learning-Based Framework for Phishing Website Detection", 2020

[6] Journal of Artificial Intelligence, "Phishing Website URL's Detection Using NLP and Machine Learning Techniques", 2023