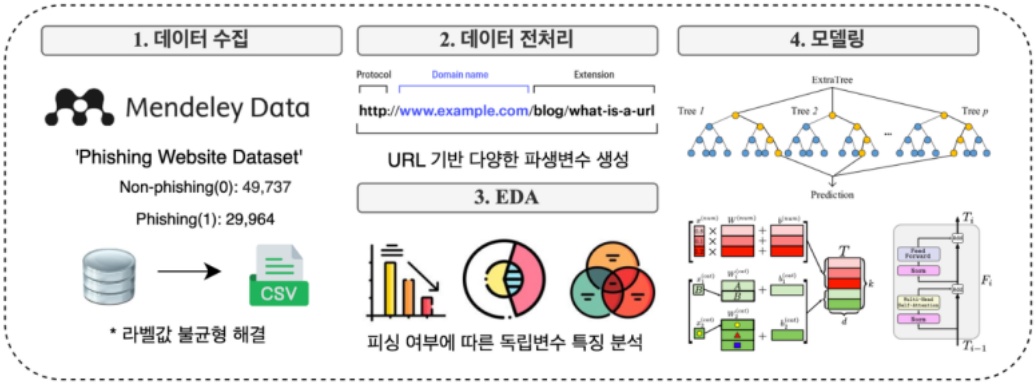
공모 제안서 요약문

큐싱(Qshing)은 QR 코드를 악용한 피싱 범죄로, 스마트폰과 QR 코드 사용이 보편화된 대한민국에서 피해가 꾸준히 증가하고 있다. 사용자가 악성 URL을 사전에 식별하기 어렵고, 기존 보안 솔루션으로는 실시간 경고와 차단이 어려워 예방 기술이 필수적이다. AI 기술을 활용해 QR 코드 URL의 이상 패턴을 분석하고 악성 여부를 실시간으로 탐지함으로써 피해를 줄이고 사용자 신뢰를 회복할 수 있다. 이를 통해 디지털 환경의 안전성을 강화하고 사회적 비용을 절감할 수 있을 것이다.



본 프로젝트의 데이터 분석 프로세스는 데이터 수집, 데이터 전처리, EDA, 모델링 단계로 이루어지며, 개발된 모델을 바탕으로 큐싱 범죄 차단을 위한 QR코드 URL 실시간 경고 시스템의 UI를 함께 제안한다. ‘Mendeley Data’의 ‘Phishing Website Dataset’ 데이터를 사용해 csv파일 형태로 변환시키고, 라벨값 불균형 해결을 위해 언더샘플링을 적용하여 피싱 URL와 정상URL 각각 29,964개, 총 59,928개를 활용한다. 이어, URL을 기반으로 얻을 수 있는 다양한 정보들을 추출하여 총 15개의 파생변수들을 생성한다. 예를 들어, URL 길이를 계산하여 ‘url_length’ 변수를 추출한다. 그 후, EDA과정을 통해 피싱 여부에 따른 파생변수들의 특성을 분석하여 ‘피싱 URL은 정상 URL보다 더 긴 URL과 호스트 이름을 가지는 경향을 가지며 이는 악성 코드를 포함한 복잡한 경로를 감추기 위한 것으로 추정’ 등과 같은 다양한 인사이트를 얻는다. 다음으로 AI 알고리즘을 활용해 URL 피싱 여부 분류 모형을 구축한다. 이를 위해 4개의 머신러닝 알고리즘(그래디언트 부스팅, 랜덤 포레스트, 익스트림 랜덤 트리, 의사결정 트리)과 피쳐 토크나이저를 사용해 정형 데이터에 대해 높은 성능을 보이는 트랜스포머 기반 딥러닝 알고리즘인 FT-Transformer모형을 학습시키고, 정확도, 정밀도, 재현율, F1 점수에 대해 가장 우수한 모형을 도출한다. 이후, 베이지안 최적화 기법을 통해 모델의 성능을 향상시킨다.

그동안 악성 QR 코드는 QR과 링크된 URL 주소를 눈으로 확인할 수 없어, QR을 카메라로 인식시켜 해당 URL 사이트에 접속하는 과정에서 피해자가 속출했다. 본 시스템을 통해 사용자는 간단한 QR 코드 스캐닝만으로 악성 QR 코드로 인한 피싱 공격과 악성 소프트웨어 설치를 사전에 차단할 수 있다. 이는 사용자 개인의 재산과 일상을 보호할 뿐만 아니라, 안전한 QR 제공자의 신뢰도를 높이는 데에도 기여한다. 또한, 서비스 내 새로운 QR 코드 데이터를 지속적으로 분석한다면 새로운 피싱 기법을 탐지하고 이를 사법기관 및 보안업계와 공유함으로써 공공 안전을 강화하는 데에도 기여할 수 있을 것이다.