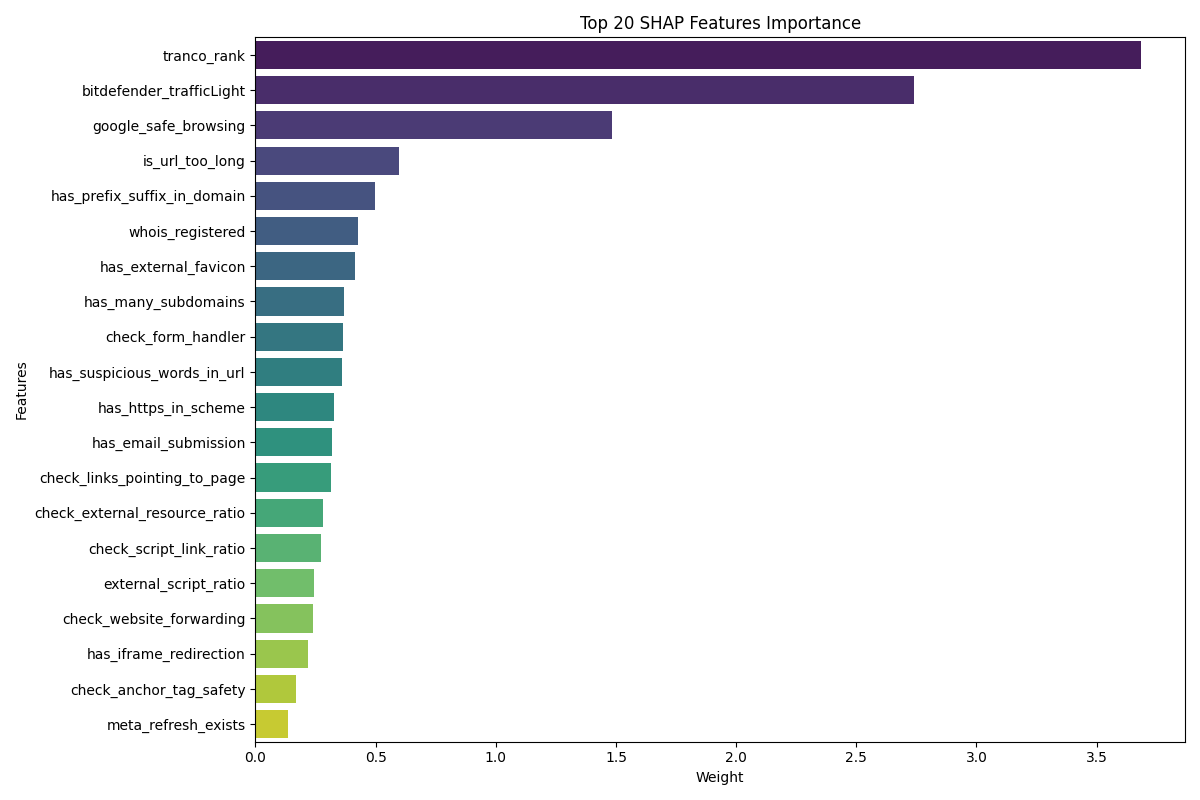
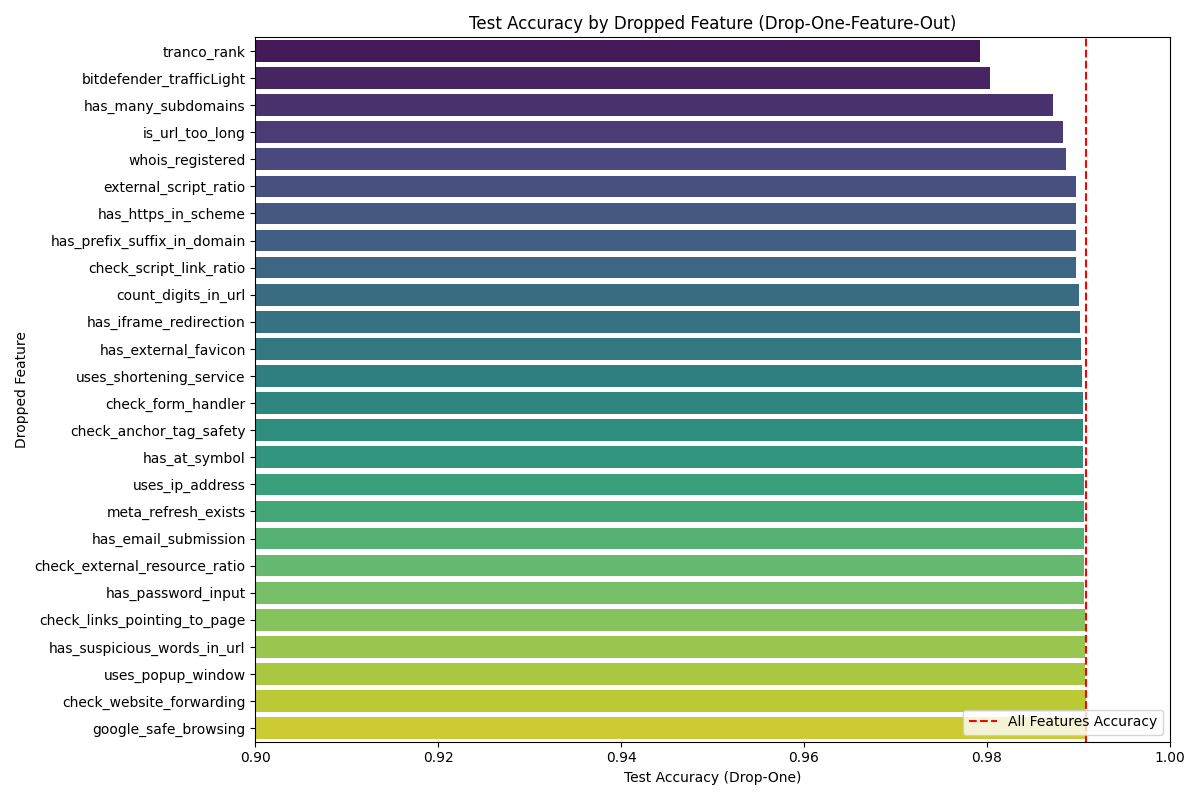
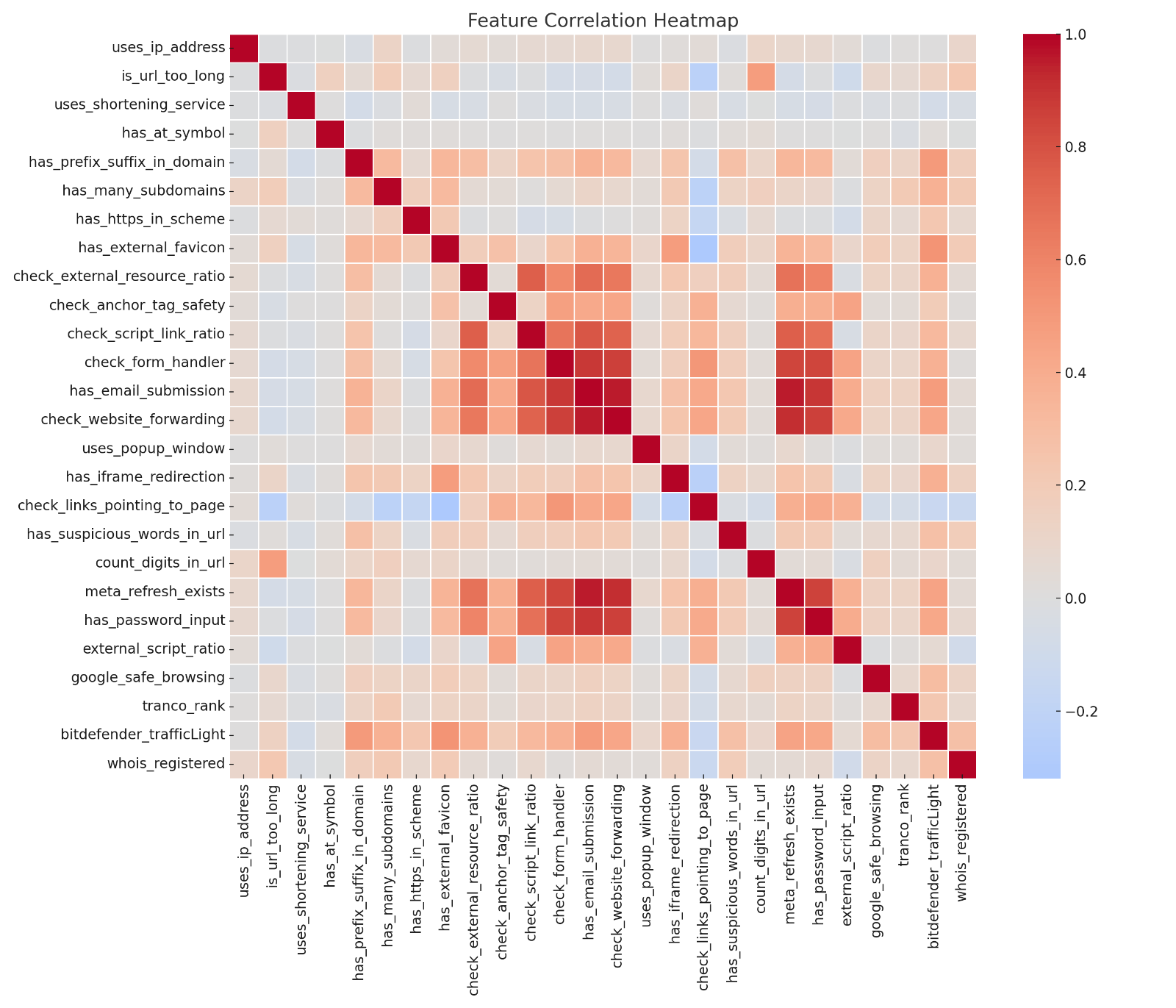


구 모델 (모델 조합, 파라미터 값 수작업) 피쳐 중요도

  
optuna 기반 모델 피쳐 중요도



optuna 기반 모델 기준 피쳐 1개씩 제거한 정확도



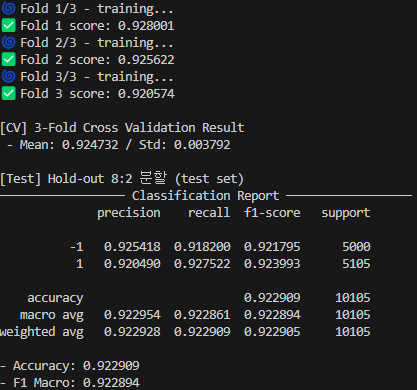
피쳐간 상관관계 히트맵

**----- 위 내용 설명 ------**

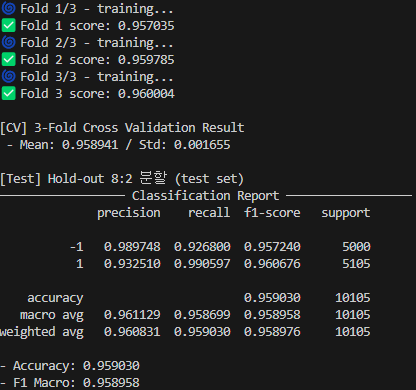
1.  
optuna로 모델 조합과 파라미터 값을 데이터셋에 맞게 최적화시키기 때문에 optuna 기반 모델과 구 모델의 피쳐 중요도가 각각 다르게 나오므로 중요도 하위권에 있는 피쳐가 데이터셋 구성에 따라서 상위권으로 중요도가 바뀔 수 있다.

2.  
optuna 기반 모델 중요도에서 구글 세이프 브라우징이 3번째로 중요도가 높지만 실제 구글 세이프 브라우징을 뺀 정확도를 보면 정확도 변동이 가장 낮기 때문에 중요도와 정확도가 직접적으로 연결되는건 아니다.

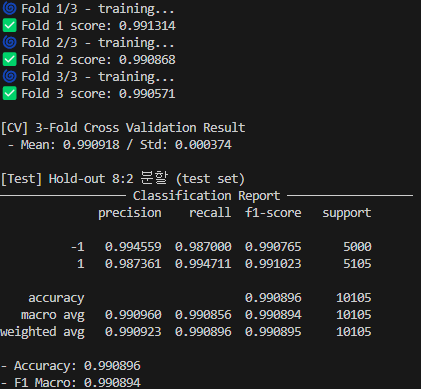
3.  
optuna 기반 모델 중요도에서 tranco\_rank와 bitdefender\_trafficlight의 피쳐 각각 제거한 정확도를 보면 전체피쳐 기준 99%에서 각각 제거한 정확도는 97.9%, 98%로 나오므로 중요도가 매우 높은 걸 제거하더라도 정확도에 그만큼의 큰 영향은 없다.  
그 이유는 주요 피쳐(예: tranco\_rank, bitdefender\_trafficlight 등)는 개별적으로 강한 중요도를 가지지만, 피쳐간 상관관계 히트맵을 보면 bitdefender\_trafficlight 의 경우 다른 피쳐와도 높은 상관관계를 가지기에 하나의 피쳐를 제거해도 유사한 정보를 가진 다른 피쳐들이 이를 보완하기에 성능 하락이 크지 않은것이고,  
tranco\_rank 의 경우 대부분의 피쳐들과 낮은 상관관계를 보이지만 그만큼 모델이 tranco\_rank 를 제외한 피쳐 조합만으로도 비슷한 수준의 분류를 해낼 수 있을 정도로 피쳐 선정과 모델이 잘 구성되있다는 의미이다.



정적 피쳐만으로 구성된 데이터셋 정확도 (optuna 기반 모델)



외부 서비스 피쳐만으로 구성된 데이터셋 정확도 (optuna 기반 모델)



정적 + 외부 서비스 통합 데이터셋 정확도 (optuna 기반 모델)

**----- 왜 외부서비스 피쳐를 썼을 때 정확도가 올라가는가? -----**

**1. 외부 서비스 피쳐의 정확도가 높은 이유**

외부 서비스 피쳐는 신뢰도 높은 "정답 신호"를 제공하며, 피쳐 개수가 적어도 정보의 질이 뛰어나다.

위 정적, 외부, 통합 테스트 정확도를 각각 살펴보면 정적: 0.922 / 외부: 0.959 / 통합: 0.9908 로 정적 피쳐가 22개, 외부 서비스 피쳐가 4개인거에 반해 외부 서비스의 정확도가 더 높은데 그 이유를 살펴보면

정적 피쳐는 개수가 많아도 "변조 및 우회"가 쉬운 값이 많고, 실제 피싱/정상 구분을 결정적으로 설명해주지 못하는 경우가 많지만, 외부 서비스 피쳐는 "전 세계 수억/수십억 유저의 행동 + 보안 전문가 분석 + 머신러닝"이 다 반영된 정보로, 개수가 적어도 단일 피쳐가 매우 강력한 분류 신호를 가지기에 피쳐의 개수가 적어도 높은 정확도를 보인다.

**2. 정적 피쳐와 외부서비스 피쳐의 시너지 효과**

외부 서비스 피쳐 테스트 결과에서 precision(정밀도) 와 recall(재현율) 을 각각 살펴보면 피싱에 대해 정밀도는 0.989, 재현율은 0.926 을 보이고 있다. 이것이 의미하는 바는 위에서 말했듯이 외부 서비스는 신뢰도 높은 “정답 신호”를 제공하기에 피싱이라고 분류한 것 중 98.9%는 일치 했지만, 테스트한 데이터셋에 있는 모든 피싱에 대해서는 92.6% 밖에 발견하지 못했다고 볼 수 있는데,

이는 몇 시간, 몇 일 사이에 만들어진 신규 피싱 사이트의 경우 대부분이 외부 서비스 DB에 등록되기 전이기 때문에 최근에 만들어진 피싱 사이트가 데이터 셋에 많을수록 그만큼 외부 서비스의 재현율은 낮아진다.

정적 피쳐는 URL/HTML 코드 수준의 “기초적 패턴”을 잡는 역할, 외부 서비스 피쳐는 “실제 보안 DB에 등록된 실전 신호”를 잡는 역할을 함.

외부 서비스 DB 반영 전이거나 새로 나온 피싱 기법 등은 정적 피쳐가 보완해줄 수 있고, 정적 피쳐만으로는 구분이 애매한 케이스는 외부 서비스 피쳐가 명확하게 보완해줌.

두 피쳐가 상호 보완적으로 동작해 최종적으로 정확도가 대폭 상승한다.

**3. 모델 최적화로 높아지는 성능**

Optuna 등 자동화 최적화 툴은 각 피쳐의 조합 및 모델 파라미터를 실험적으로 탐색하여 해당 데이터셋에서 가장 효과적인 구조를 찾아내기에, 단순히 외부 서비스 피쳐만 추가하는 것 이상으로 각 피쳐별 기여도를 극대화하는 모델 조합을 찾을 수 있게 된다.

**----- 외부 서비스 설명 -----**

**■ Bitdefender TrafficLight**

글로벌 보안 회사인 Bitdefender에서 제공하는 실시간 웹사이트 평판 서비스 결과를 직접 활용함.

이 서비스는 수많은 사용자와 보안 센서, 크롤러, 위협 인텔리전스 네트워크를 통해 전 세계적으로 탐지된 최신 피싱/악성 사이트 목록을 관리하며, 위험도가 확인된 사이트는 수분 내로 DB에 반영됨.

단순히 URL 문자열이나 패턴만 보는 것이 아니라, 실제로 "피싱 시도", "악성 코드 배포", "위험한 행동 감지" 등을 종합적으로 분석해 해당 사이트가 피싱/위험으로 등록되어 있을 경우 바로 탐지할 수 있음.

즉, Bitdefender TrafficLight의 신호 하나만으로도 그 사이트가 보안 커뮤니티에서 이미 위험하다고 판정된 경우가 대부분이므로 분류 정확도가 매우 높아짐.

**■ Google Safe Browsing**

Google에서 전 세계적으로 운영하는 악성/피싱 URL 정보의 대표적인 집합.

크롬/파이어폭스 등 주요 브라우저, 안드로이드 스마트폰, 구글 검색엔진, 기타 파트너 사이트/앱에서 발생하는 보안 이벤트를 모두 집계하여 실시간으로 의심 사이트를 차단하고 DB에 등록함.

머신러닝, 휴리스틱, 사용자 리포트, 구글 검색 크롤러 등 다양한 소스를 통해 “자동 및 수동”으로 지속적으로 악성/피싱 사이트를 모니터링함.

**■ Tranco Rank**

Tranco는 전 세계에서 가장 많이 방문하는 사이트의 순위를 제공하는 랭킹 서비스로, 주로 "신뢰받는 정상 도메인"과 "의심스러운 신규/저품질 도메인"을 구분하는데 사용됨.

일반적으로 피싱/스팸 사이트는 방문량이 적거나 신규 도메인인 경우가 많으므로 Tranco 랭크가 낮거나 순위에 없는 도메인은 위험 신호로 간주할 수 있음.

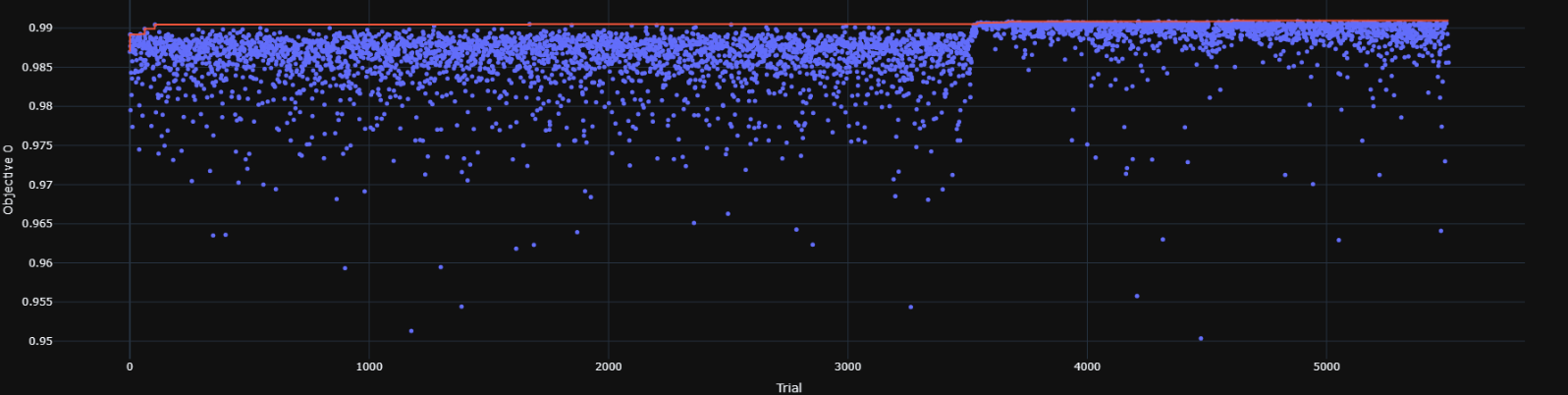
반대로, Tranco 순위에 오래 포함되어 있고 상위에 있는 사이트는 대부분 정상적인 사이트로 간주할 수 있으므로,

“정상/위험” 도메인 선별에 매우 효과적인 지표로 사용됨.

**■ Whois**

Whois 등록 여부, 도메인 생성일, DNS 블랙리스트 등도 이미 다양한 보안 커뮤니티에서 신뢰성 있는 위험 신호로 사용되는 데이터임.

도메인이 최근에 등록되었거나, Whois 정보가 부정확하거나 비공개일 경우 피싱 사이트일 가능성이 높음.



**- optuna 탐색 로직 간단한 설명**

n\_trial = 5500 // 탐색 횟수

n\_startup\_trial = 3500 // 랜덤 탐색

3500 횟수 동안 랜덤 모델 조합, 파라미터 값 선정 후 나오는 정확도 탐색 후

랜덤 탐색에서 나온 정확도를 이용해 trial 3500 이후부터 최적탐색 시작.

(정확도 분포를 보면 trial 3500까지는 분포가 퍼져있다가 3500부터 높은 정확도 구간에 밀집 후 지속적인 최고 정확도 갱신)

최적 탐색 중 계속 밀집된 분포를 보이는게 아니라 분포가 점점 넓어지거나 매우 낮은 정확도를 보이는 (튀는) 이유는 랜덤 탐색에서 산출한 모델 조합, 파라미터 값 최적 구간을 거의 사용해서 랜덤 탐색에서 낮은 정확도를 보였던 구간도 재 탐색하는 것.

(optuna 는 자체적으로 정확도에 따라 good/bad 후보군으로 분류해서 탐색함)