



# SPARROW

## 이상 분석 탐지 프로젝트

SP 개발 1팀 임영석



**01 프로젝트의 개요 및 설계**

**02 데이터 수집**

**03 요인 파악**

**04 모델 설계**

**05 서비스 연동**

**06 개선점 및 느낀점**



“  
정상적으로 작동하지 않는 분석,  
자동으로 식별 할 수 있을까?”

분석 대상	분석 시작 시간	분석 종료 시간	분석 소요 시간
소스코드 A	2025-06-26 13:46:58	2025-06-26 13:52:13	5분 15초
소스코드 A	2025-06-09 08:01:56	2025-06-09 13:04:56	5시간 3분
소스코드 B	2025-06-02 09:58:38	2025-06-02 09:59:51	1분 13초
소스코드 B	2025-06-02 09:02:23	2025-06-02 12:17:23	3시간 15분

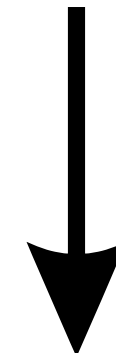
# 프로젝트 설계



소스코드



SPARROW  
SAST

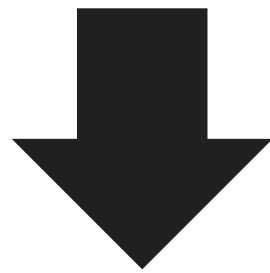


분석 소요 시간

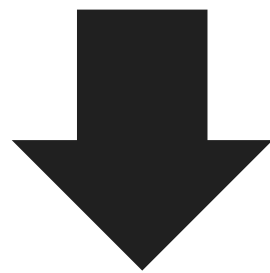
이상 탐지 프로젝트



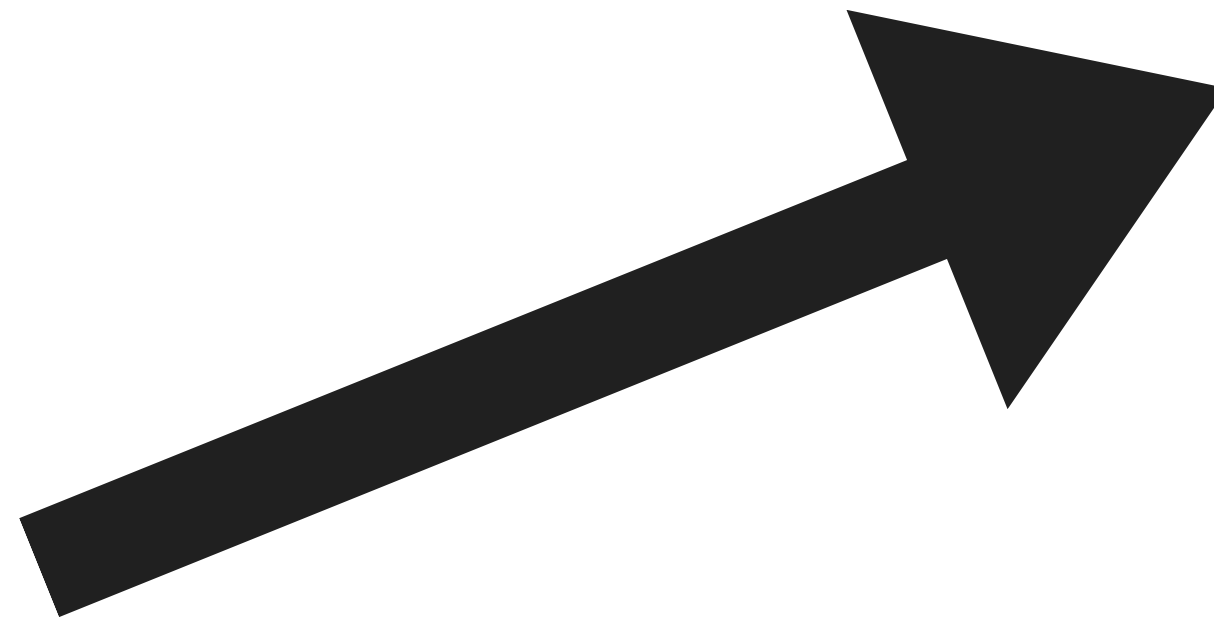
소스코드 700개 수집



분석 실행



소스코드의 복잡도, 용량,  
파일 수, 줄 수, 분석 소요 시간  
기록





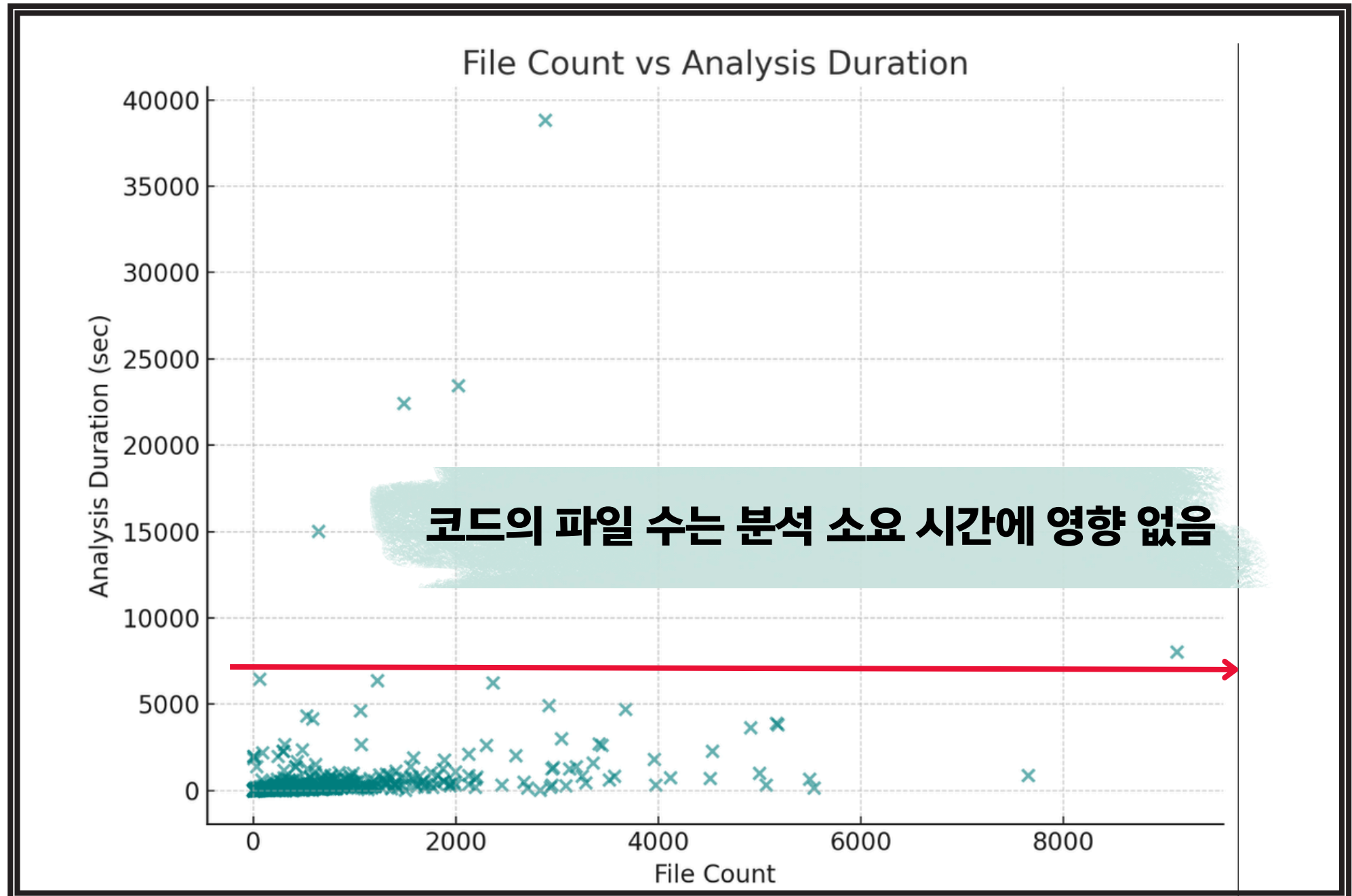
## 1. 코드의 파일 수와 분석 소요 시간의 관계



2. 코드의 복잡도와 분석 소요 시간의 관계

3. 코드의 용량과 분석 소요 시간의 관계

4. 코드의 줄 갯수와 분석 소요 시간의 관계





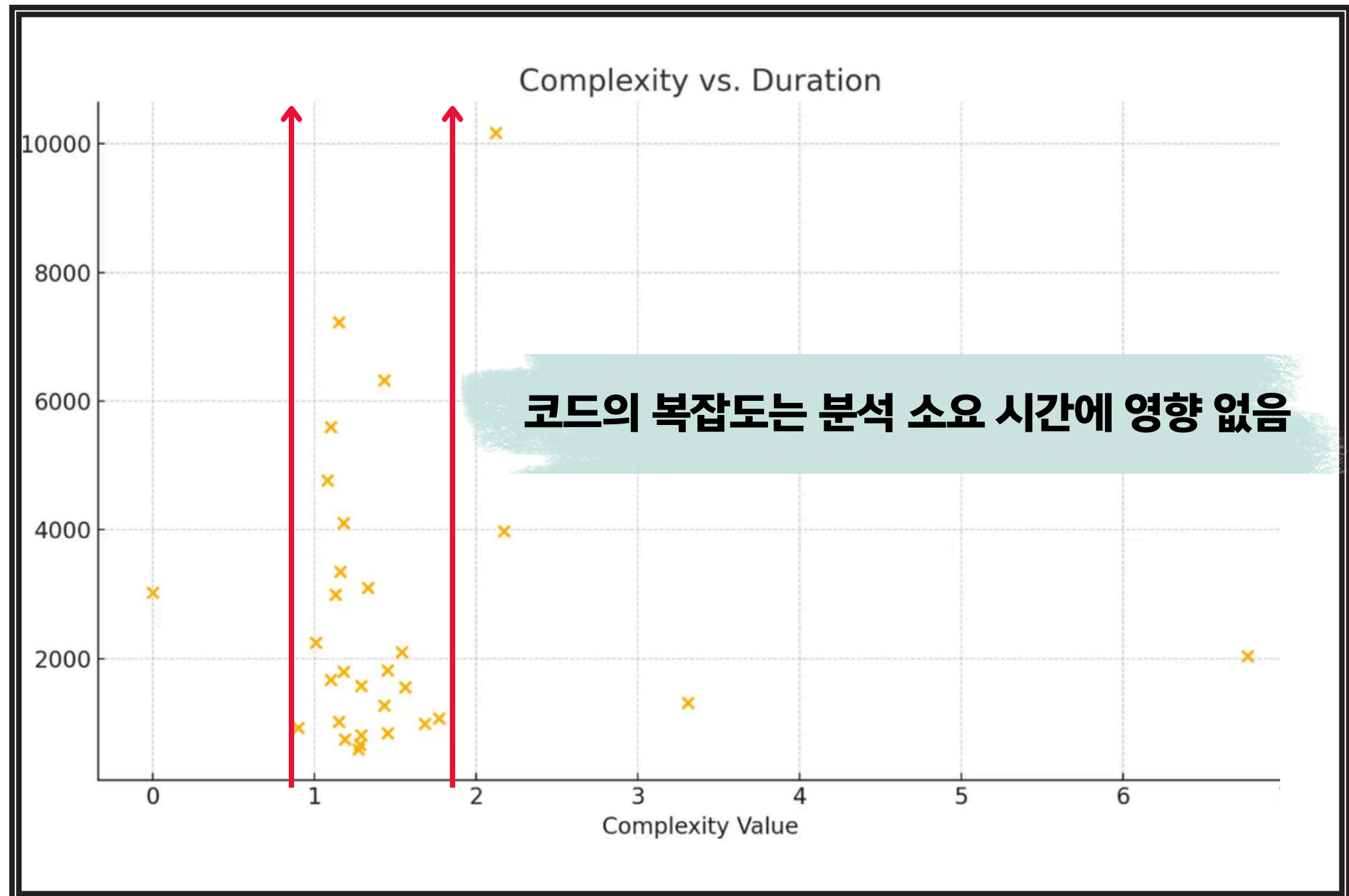
1.코드의 파일 수와 분석 소요 시간의 관계

**2.코드의 복잡도와 분석 소요 시간의 관계**



3.코드의 용량과 분석 소요 시간의 관계

4.코드의 줄 갯수와 분석 소요 시간의 관계



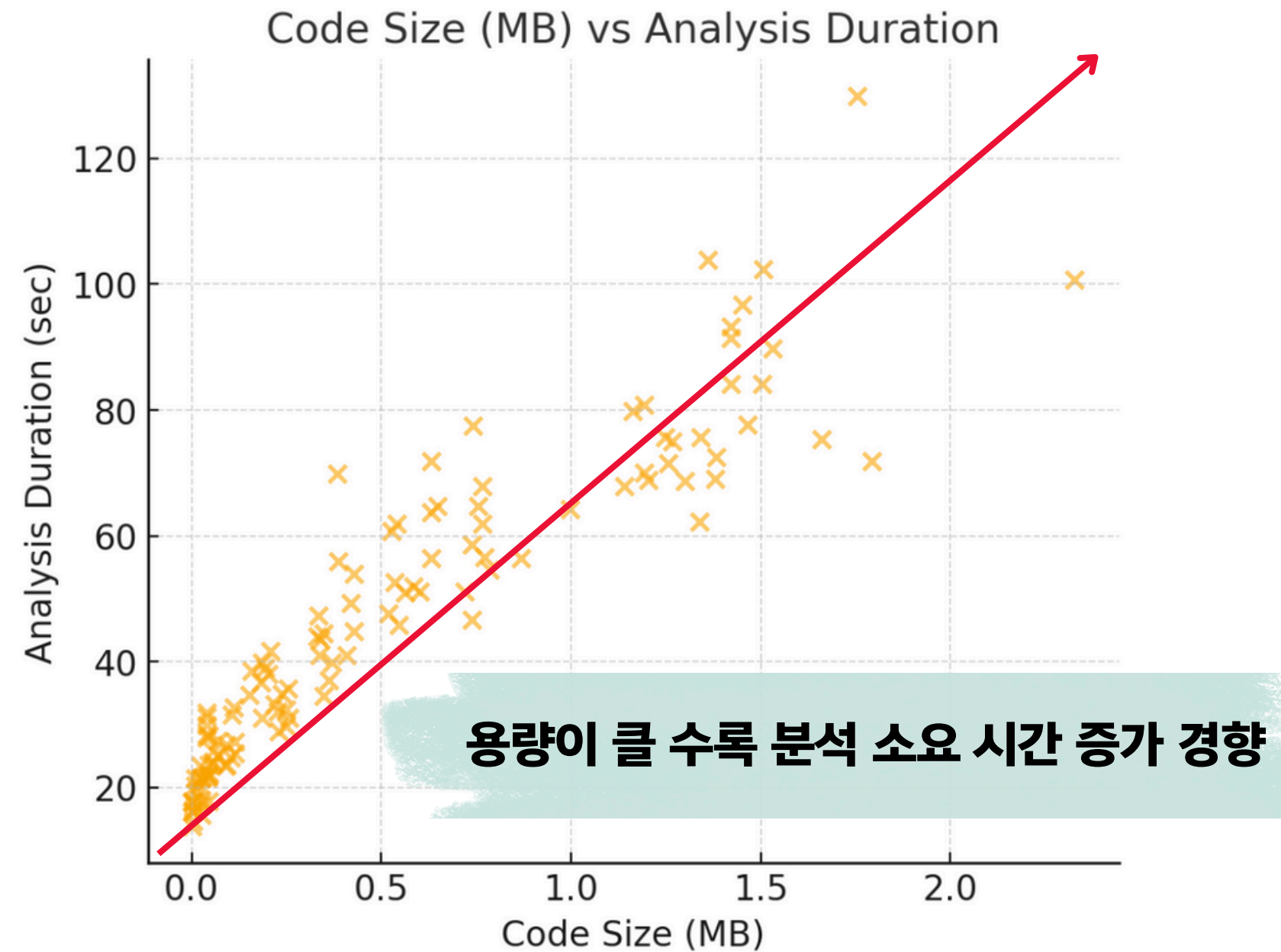


1.코드의 파일 수와 분석 소요 시간의 관계

2.코드의 복잡도와 분석 소요 시간의 관계

**3.코드의 용량과 분석 소요 시간의 관계**

4.코드의 줄 갯수와 분석 소요 시간의 관계



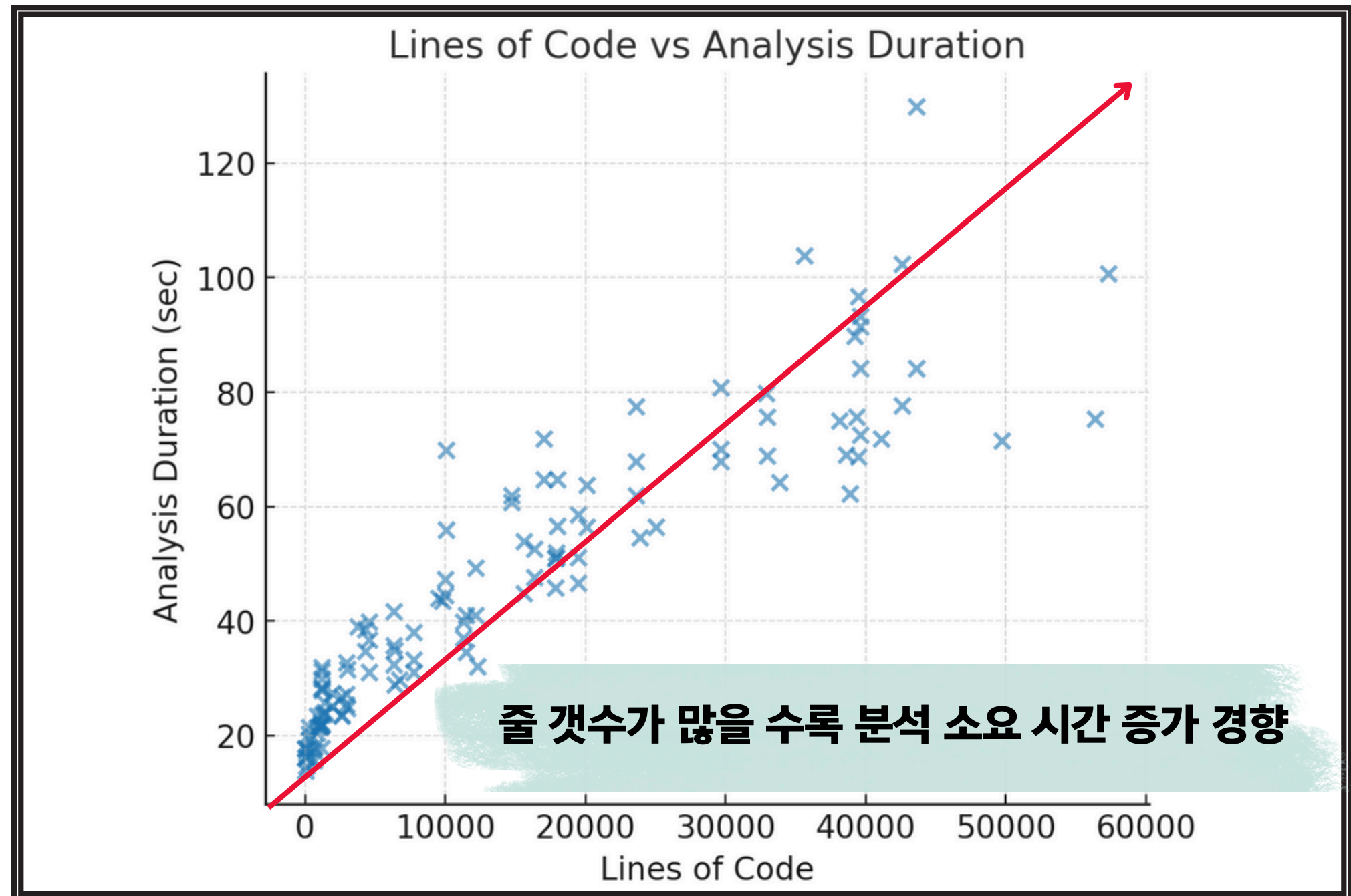


1.코드의 파일 수와 분석 소요 시간의 관계

2.코드의 복잡도와 분석 소요 시간의 관계

3.코드의 용량과 분석 소요 시간의 관계

**4.코드의 줄 갯수와 분석 소요 시간의 관계**





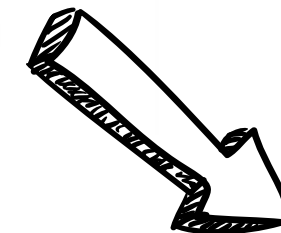
소스코드의 줄 갯수가 증가함에 따라 분석 소요 시간 선형적 증가

+

소스코드의 용량에 증가함에 따라 분석 소요 시간 선형적 증가

$$y = a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_{10} \cdot x_{10} + b$$

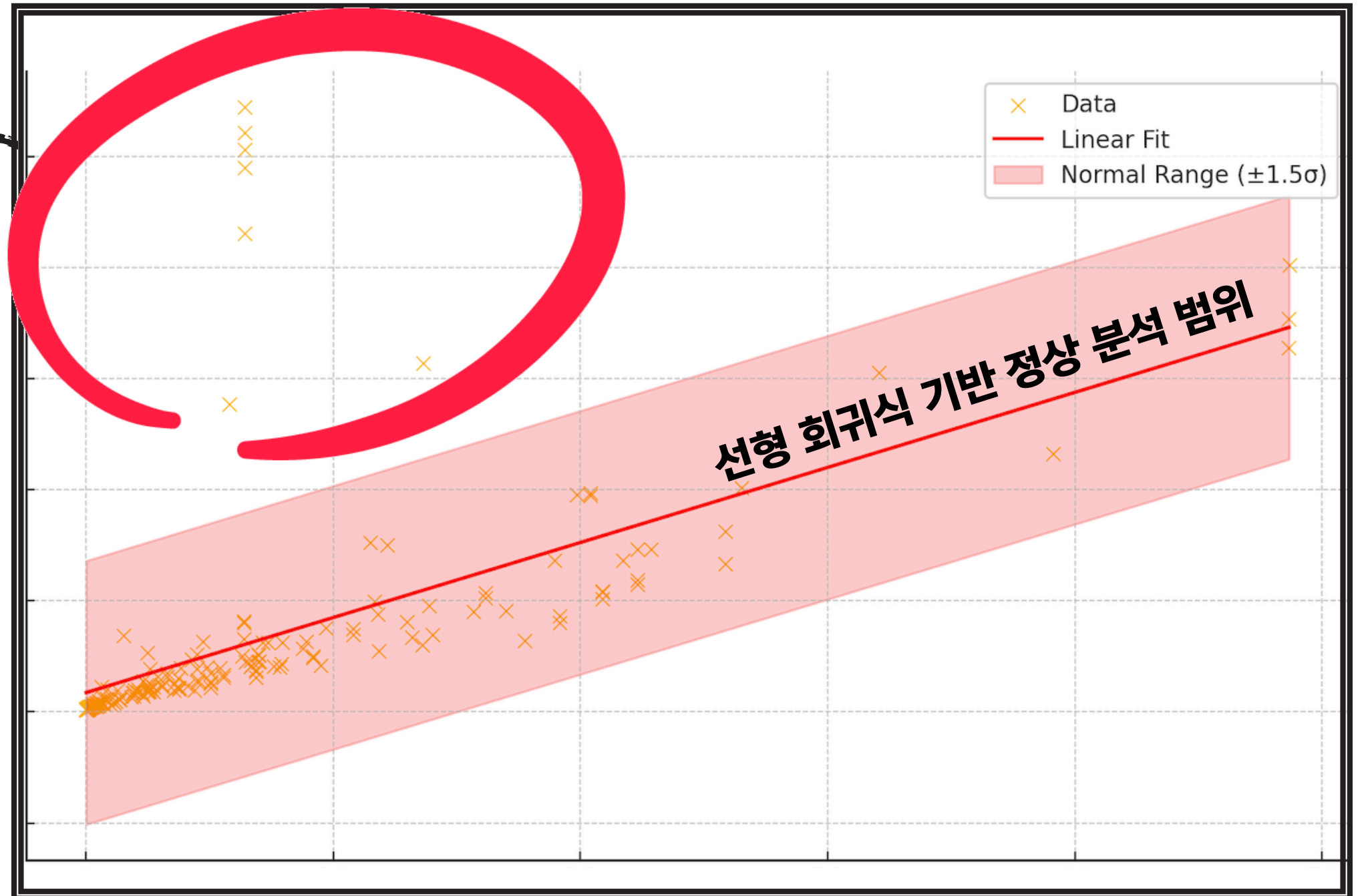
분석 소요 시간 예측식



- $a_1, a_2, \dots, a_{10}$  → 가중치
- $x_1, x_2, \dots, x_{10}$  → 입력 변수 (소스코드 줄 수, 소스코드 용량)
- $b$  → 상수항
- $y$  → 예상 분석 소요 시간



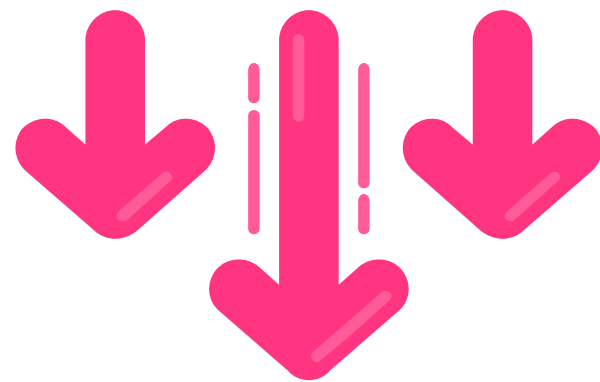
정상 분석 범위 초과한 이상 분석 탐지





선형 회귀 기반 이상 탐지

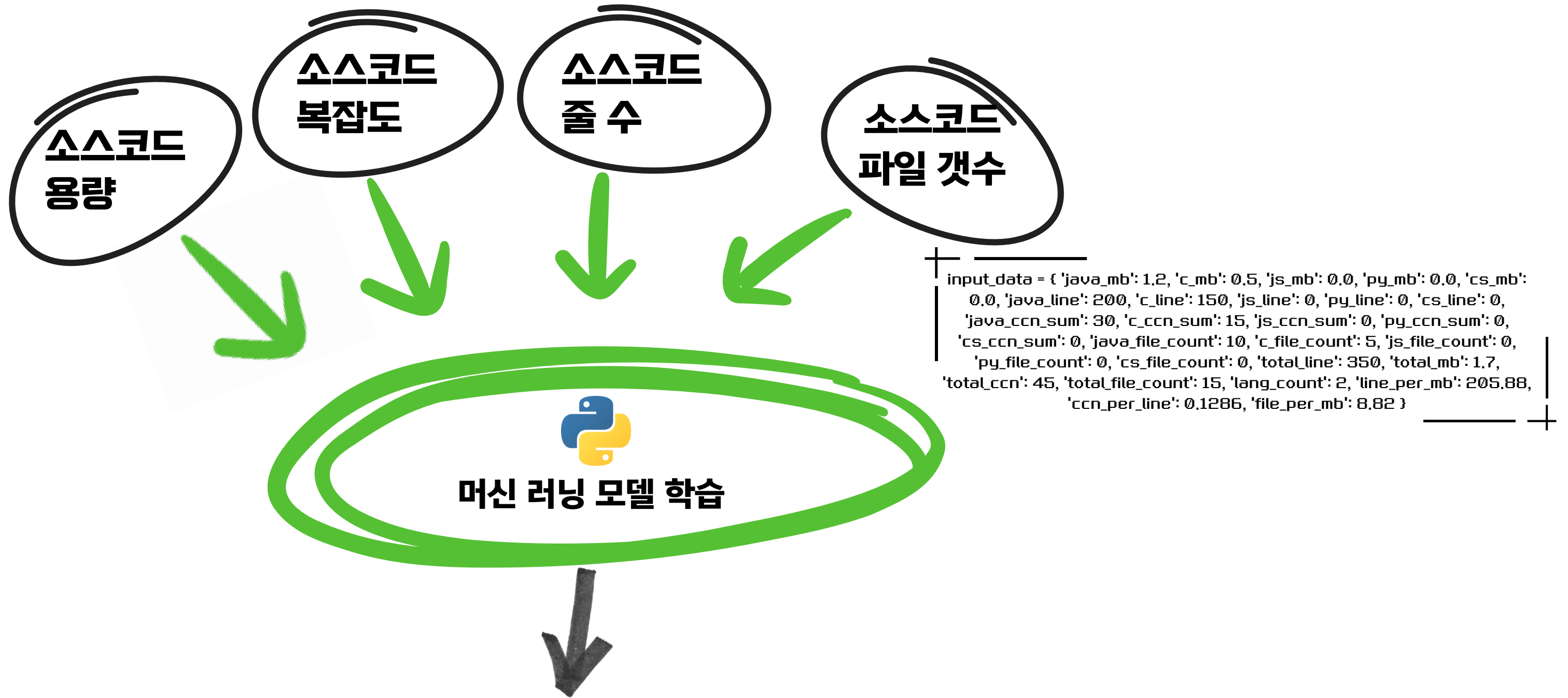
평균 예측 정확도
<b>30% 이하</b>



머신 러닝 기반 이상 탐지



눈으로 확인하기 힘든 요소들 간의 복잡한 상호 작용 파악해서 예측 모델 생성 가능



분석 소요 시간 예측 가능, 예측 시간에서 벗어날 시 이상 탐지



분석 아이디	실제 분석 소요 시간	모델 기반 예측 시간	정확도(%)
10039	84	86.33505	<b>97.22</b>
10038	76	76.01834	<b>99.98</b>
10037	62	66.66	<b>92.48</b>
10036	69	78.57419	<b>86.12</b>
10035	64	89.86089	<b>59.59</b>
10033	104	65.02225	<b>62.52</b>
10032	102	91.55056	<b>89.76</b>
10031	90	83.67673	<b>92.97</b>
10030	93	81.39842	<b>87.53</b>
10029	97	83.71684	<b>86.31</b>
10022	1017	940.31537	<b>92.46</b>
10021	764	727.8375	<b>95.27</b>

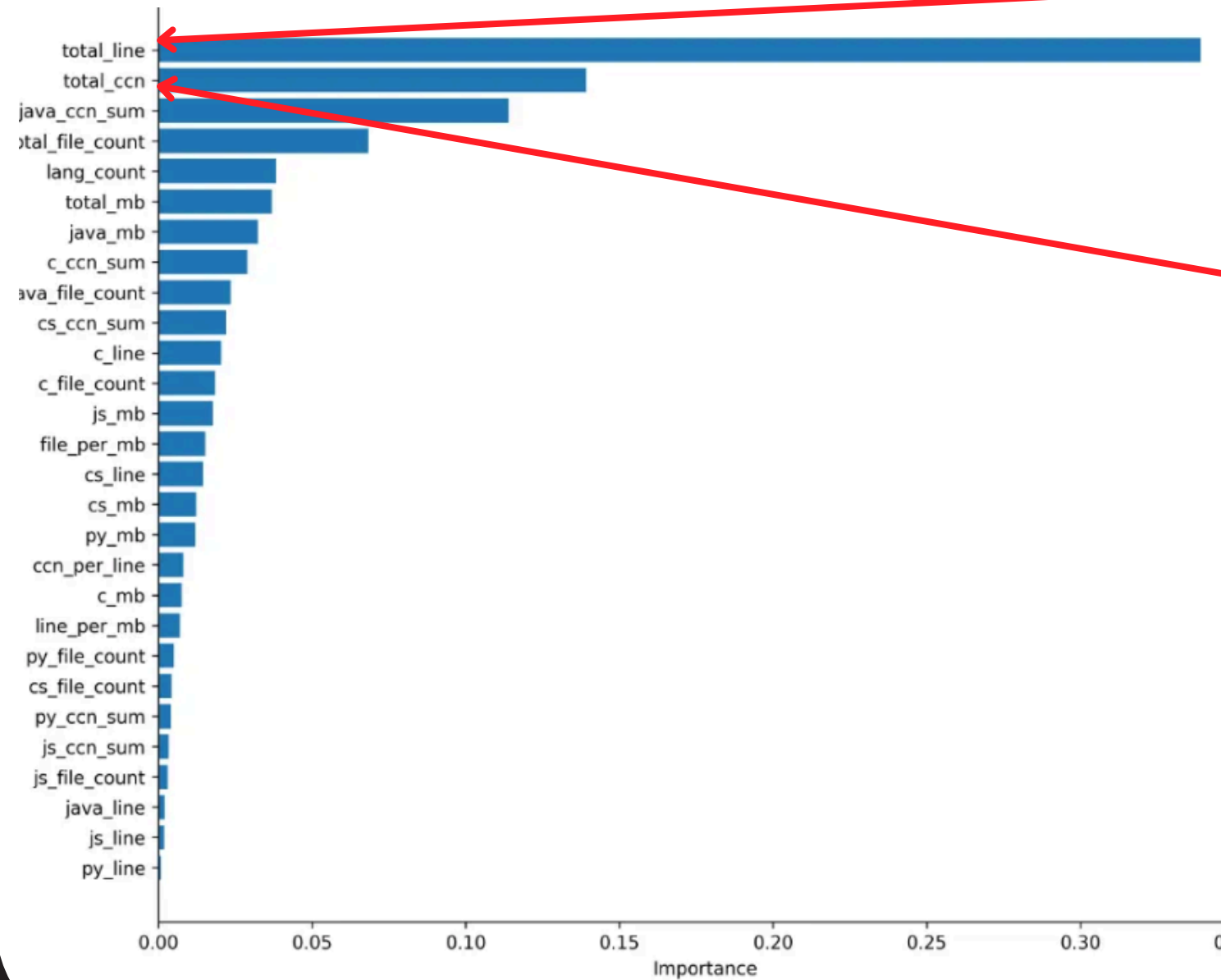
### 예측 신뢰도 지표 및 정확도 수치

모델	평균 예측 정확도 (%)
선형 회귀	29
머신러닝 (XGBoost)	93

선형 회귀 대비 예측 정확도 **64% 향상**



## 머신러닝 모델 학습 과정에서 요인별 중요도

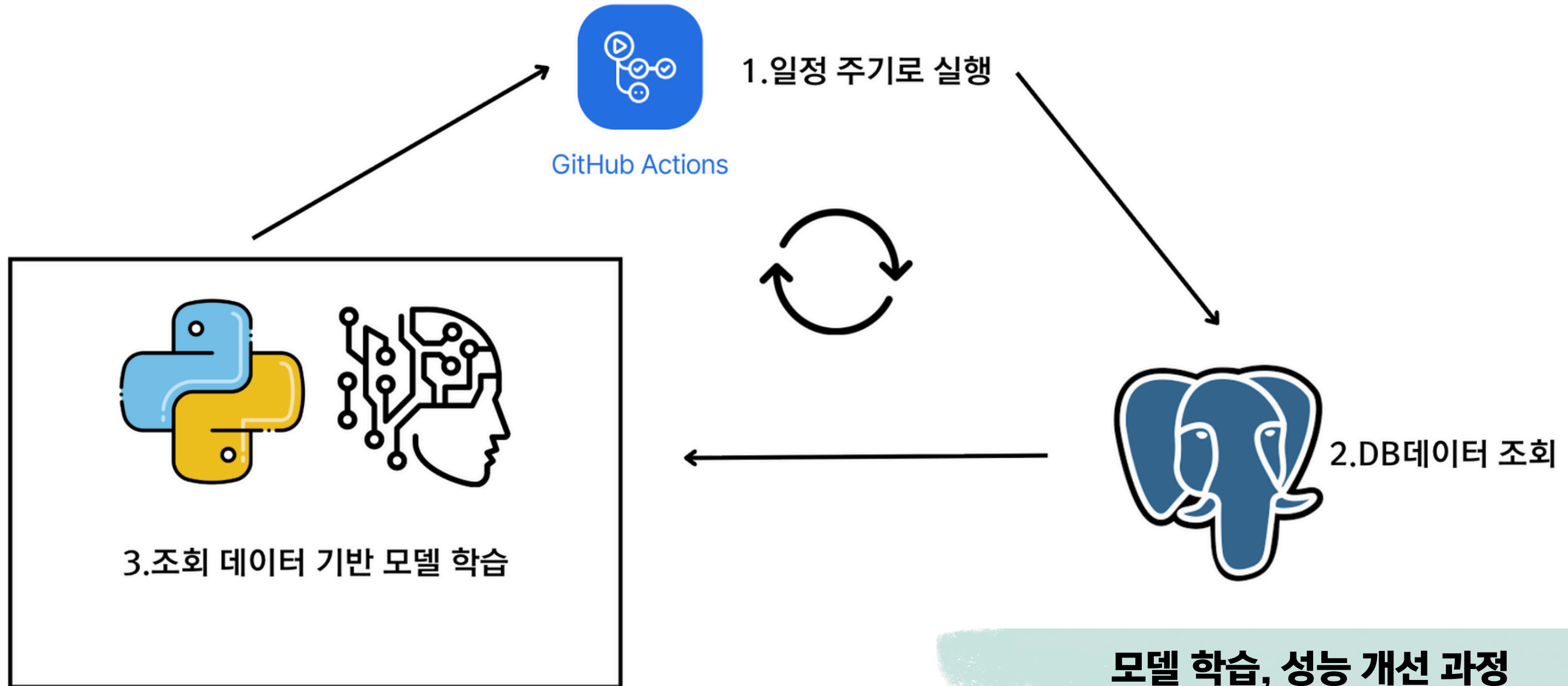


소스코드의 줄 갯수가 분석 소요 시간에 가장 영향력 높음

두번째로 소스코드의 복잡도가 분석 소요 시간에 영향력 높음

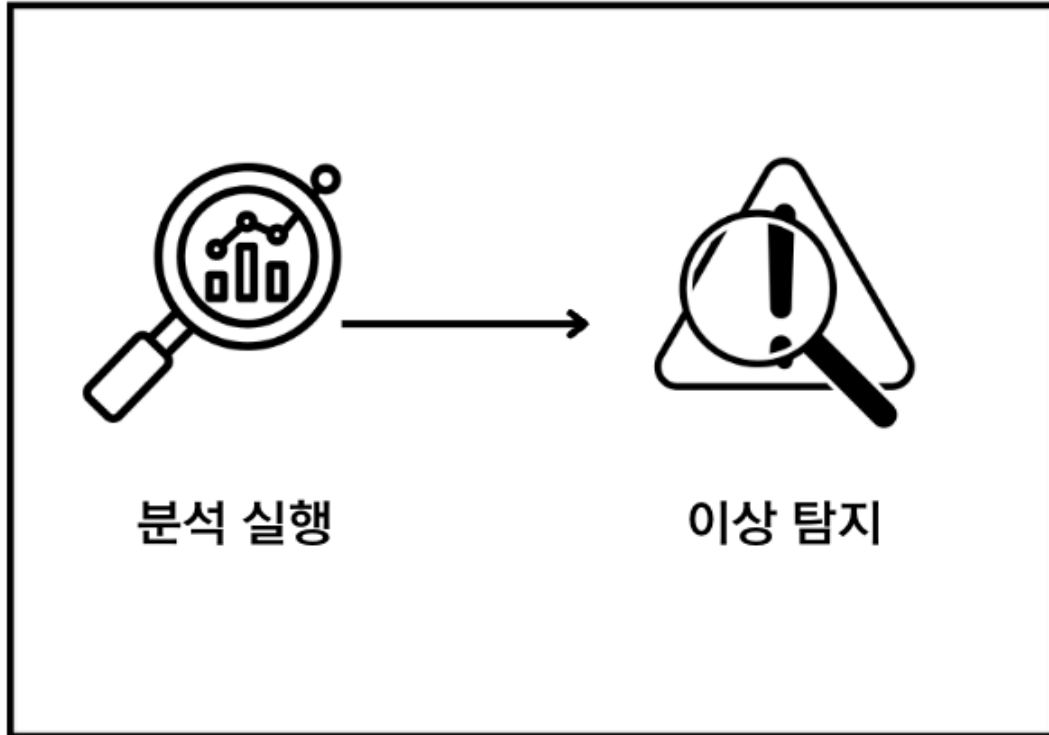
# WHY?

선형회귀 기반 예측 모델: 단순 선형적 관계만 반영  
 머신러닝 기반 예측 모델: 여러 요인들의 복합적인 상호작용 반영





## SPARROW 2508.1 릴리즈 적용 예정



모델 예측 시간(+/-20%): 55.80~83.70  
실제 분석 시간: 122.00



discord



Captain Hook 오픈 2025-08-05 오후 2:30

[Worker] 분석 시간 이상치 감지

<https://dev.ondemand.sparrowcloud.ai/requestList/request/110/analysis/67>

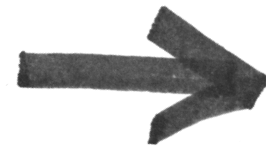
분석 아이디: 67

actual: 122.00, expected: 69.75 ,range: 55.80 ~ 83.70



#

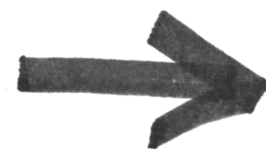
학습 데이터 부족



지속적인 분석 데이터 수집을 통한 예측 정확도 향상

#

이상 탐지 지원 언어 추가



현재 지원 가능한 다섯개 언어 외에도 지원 언어 추가



# 감사합니다

스파로우 개발 1팀 임영석