Jev-Shield

결과 정리

통화 초반 문장만으로 보험·대출·피싱 전화를 분류하는 방식이 얼마나 맞고 어디서 틀리는지를 직접 측정한다.

모델 jev-1.13.0

분류 구성 · 66 TEST CALLS

INSURANCE	11
LOAN	11
PHISHING	11
NORMAL	33

정확도

89.4%

59/66 · category 일치

본 문서의 모든 문장과 번호는 가상이며,
수록된 수치는 직접 실행해 얻은 값입니다.

1. 개요

측정의 범위와 사용한 도구

통화 초반 한두 문장만으로 판정을 내리는 데모를 만들고,
그 자체를 같은 문장 세트로 다시 측정한 결과다.

목적

얼마나 맞고,
어디서 틀리는지를
직접 잰다.

통화 초반 문장만으로 보험·대출·피싱 전화를 분류하는 방식이 가진 한계를 숫자로 본다.

범위

가상 문장을 입력하는 웹 데모와 정확도 평가까지.

실제 통화 연동 · 음성 인식 · 모바일 앱은 포함하지 않는다.

판별 모델

TypeSafe의 분류 전용 모델 Jev.

문장을 길게 쓰지 않고, 정해 준 질문에 정해진 형식으로만 답한다.

2.1 · 2.2

두 가지 질문에 답하고,
기준을 넘으면 경고를 띄운다.

두 가지 질문

category
선택(choice)

INSURANCE, LOAN, PHISHING, NORMAL

네 가지 중 하나로 답한다.

risk
단계(score)



0 · 1 · 2 · 3 · 4 위험 없음 · 낮음 · 보통 · 높음 · 매우 높음

같은 문장을 두 번 호출하면 risk가 0.01 ~ 0.13 정도 달라진다. category는 모두 같았다.

경고 기준

category	risk	문구
INSURANCE	≥ 2.5	보험 가입 권유 전화로 의심됩니다.
LOAN	≥ 3.0	대출 권유 전화로 의심됩니다.
PHISHING	≥ 2.0	보이스피싱이 의심됩니다. 개인정보나 송금 요구에 응하지 마세요.

피싱은 피해가 커서 기준을 낮게 잡았다.
오탐이 있을 수 있어 문구는 “의심됩니다”로 단정하지 않는다.

2.3 · 2.4

응답 한 건과 데모 다섯 파일

2.3 응답 예시

입력

“안녕하세요, 고객님. 이번에 새로 나온 암보험 상품이 있어서 안내드리려고 연락드렸습니다.”

출력

category
INSURANCE confidence **1.0**

risk
2.67 confidence **0.55**

응답 시간
249 ms

2.4 웹 데모 구성

jev_client.py

Jev 호출,
응답에서
category와 risk 추출

evaluate.py

테스트 문장 전체 호출,
지표 계산,
results.csv 저장

app.py

웹 서버.
번호와 문장을 받아 호출,
경고 여부 판정

templates/index.html

전화 수신 화면 형태의
입력과 결과 표시

testset.csv

테스트 문장
66건

화면에는 발신번호, 경고 상자, 분류, 위험 단계(0 ~ 4 막대), 응답 시간(ms)이 표시된다. 경고가 없으면 “경고 기준에 해당하지 않습니다.”가 초록 상자로 표시된다.

3. 테스트 세트

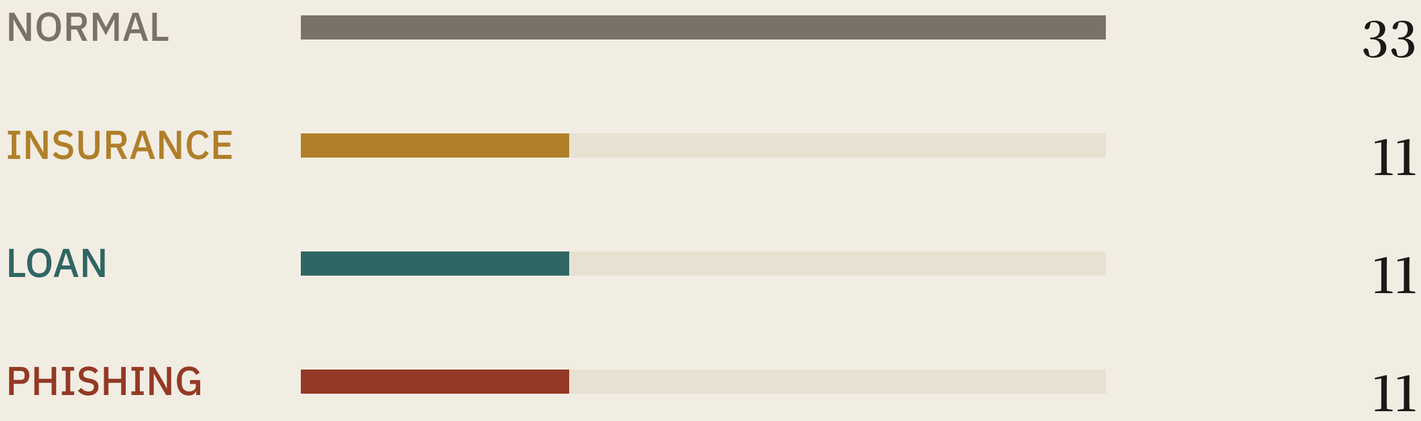
가상 문장 66건의 분포

총 테스트 문장

66

모두 호출에 성공했다. **실패 0건.**

분포



구성 메모

정상 33건 중 약 **18건은 스팸처럼 보이는 정상 문장**이다 — 보험금 지급 안내, 대출 만기 안내, 카드 재발급 안내, 도서관 책 대출 반납 등.

기준값을 정할 때 쓴 예시 문장 **10개는 테스트 세트에 넣지 않았다.** 실제 통화 내용은 쓰지 않았다.

4.1 핵심 지표

66건 중 59건이 맞았다.

category 일치 기준의 세 숫자다. 모두 같은 66건 호출에서 나온 값이다.

정확도

89.4%

59 / 66

category가 정답과 일치한 비율.

스팸 탐지율

100%

33 / 33

스팸 문장 중 NORMAL이 아닌 것으로 분류된 비율. 한 건도 빠뜨리지 않았다.

정상 전화 오탐률

18.2%

6 / 33

정상 문장 중 NORMAL이 아닌 것으로 분류된 비율. 여섯 건이 남아 있다.

4.2 혼동 행렬

행은 정답, 열은 예측

대각선이 맞은 곳, 나머지가 틀린 곳. 59건이 맞고, 7건이 빗나갔다.

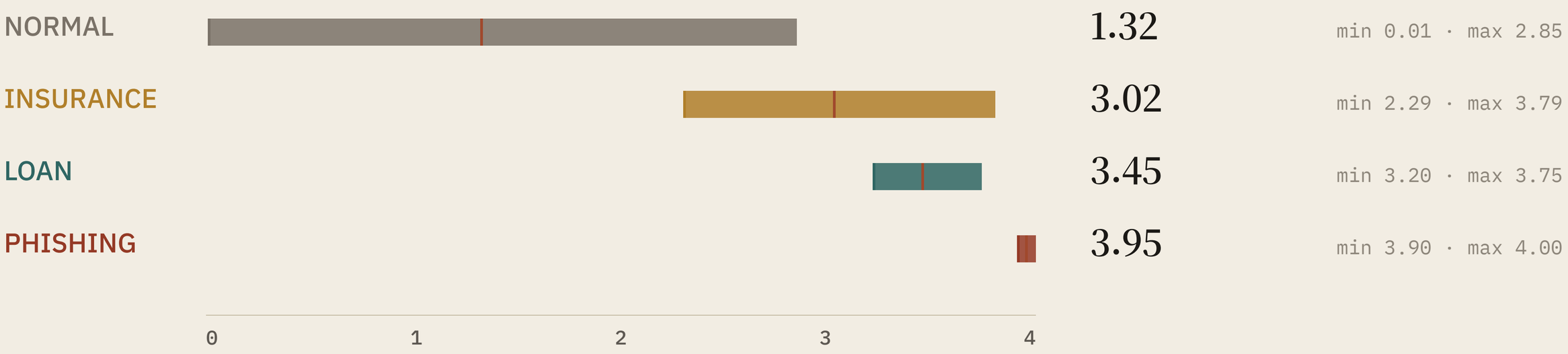
↓ 정답 / 예측 →	INSURANCE	LOAN	NORMAL	PHISHING
INSURANCE	10	0	0	1
LOAN	0	11	0	0
NORMAL	3	2	27	1
PHISHING	0	0	0	11

정답 일치 · 정답 아님, 셀 수만 표시 · 범주가 다른 오분류, 좌측 색이 정답

4.3 · 4.4 · 4.5

risk는 어디로 모이고, 경고는 어디에서 멈췄는가.

4.3 정답 종류별 risk



4.4 경고 결과

INSURANCE	LOAN	PHISHING	NORMAL
경고 9 / 11	경고 11 / 11	경고 11 / 11	경고 1 / 33

스팸 33건 중 31건에서 경고가 떴다. 응답 시간 중앙값 224 ms · 최소 198 · 최대 478

5. 틀린 사례

7건이 빗나갔고, 그중 6건은 보험사·은행이 건 정상 안내 전화였다.

5.1 category가 틀린 7건

01	INSURANCE → PHISHING	3.79	OO생명 이벤트에 당첨되셔서 연락드렸습니다. 사은품 받으시려면 상품 설명만 잠깐 들으시면 돼요.
02	NORMAL → INSURANCE	1.69	OO화재 보상팀입니다. 접수하신 사고 건 현장 확인 일정을 잡으려고 연락드렸어요.
03	NORMAL → INSURANCE	1.57	고객님, 보험료 자동이체가 잔액 부족으로 안 돼서 납입 방법을 확인하려고 전화드렸습니다.
04	NORMAL → INSURANCE	2.43	OO손해보험입니다. 계약자 정보 중 주소가 바뀌셨는지 확인차 연락드렸어요.
05	NORMAL → LOAN	2.54	OO은행입니다. 고객님 대출 만기가 다음 달이라 연장 여부를 여쭤보려고 연락드렸습니다.
06	NORMAL → LOAN	1.46	OO은행 지점입니다. 예약하신 대출 상담이 내일 오전 10시로 잡혀 있어서 확인 전화 드렸어요.
07	NORMAL → PHISHING	2.85	OO은행입니다. 방금 하신 해외 결제가 본인 사용이 맞는지만 확인 부탁드립니다.

5.2 · 5.3 · 5.4

놓친 보험 스팸 2건

2.29 · 요즘 치아보험 문의가 많아서 안내드리고 있어요

2.35 · 종신보험 하나로 사망보장이랑 연금까지 같이 준비

남아 있는 오탐 1건

PHISHING · risk 2.85

해외 결제 확인을 가장한 정상 안내 전화

보험 risk는 겹친다

기준을 2.25로 낮추면 더 잡지만,
정상 안내에도 경고가 뜬다. **조정하지 않았다.**

스팸은 모두 스팸 종류 중 하나로 분류됐다. 정상 7건 중 6건은 보험사·은행의 정상 안내 전화였다.

6. 기준값을 정한 과정

테스트 세트와 별개의 예시 문장 10개, 두 번씩 호출해 선을 그었다.

예시 10개 · 두 번씩 호출

의도	분류	risk 1차	risk 2차	요약
보험	INSURANCE	3.05	2.99	실손보험보다 보장 넓은 상품 안내
보험	INSURANCE	2.87	2.85	설계사, 무료 보장 분석
대출	LOAN	3.85	3.81	정부지원 저금리 대출 대상자 선정
대출	LOAN	3.84	3.85	신용등급 상관없이 당일 대출
피싱	PHISHING	4.00	3.99	검찰 수사관, 안전계좌로 이체
피싱	PHISHING	3.86	3.85	자녀 사칭, 돈 보내 달라는 문자형 요청
정상	INSURANCE (오분류)	1.88	2.00	자동차보험 만기 갱신 안내
정상	LOAN (오분류)	2.45	2.52	신청한 대출 서류 누락 안내
정상	NORMAL	1.11	0.99	병원 진료 예약 확인
정상	NORMAL	1.55	1.42	택배 기사

선택한 기준값

PHISHING

≥ 2.0

피싱은 피해가 커서
기준을 가장 낮게 잡았다.

INSURANCE

≥ 2.5

정상 misclass 위로,
스팸은 닿지 않게 잡았다.

LOAN

≥ 3.0

정한 기준값은 66건에 그대로
적용해 경고 결과를 얻었다.