2024 미래에셋증권 AI·DataFestival

Hyper CLOVA X와 함께, AI로 만드는 금융투자의 새로운 경험

건강한 투자를 위한

고객언어 맞춤 투자 정보 제공 서비스

Team. DSlab

권성훈 · 윤민정 · 고영준

MIRAE ASSET × NAVER CLOUD PLATFORM

목차

01
제안 배경

- 배경
- 필요성
- 목적

02
아이디어 소개

- 제안 아이디어
- 프레임워크

03
데이터 분석

- 고객 세분화
- 고객군 선정

04
언어모델 활용

- 컨텐츠 재생성
- 상품 설명 재생성
- 효과성 측정 방안

05
서비스 구현

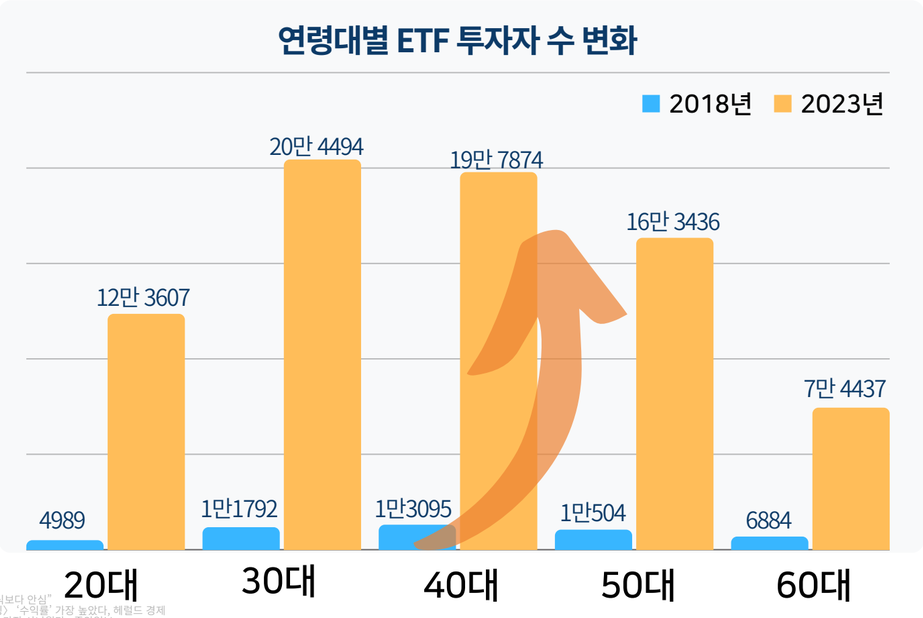
- AI 쉬운 설명
- AI 핵심 요약
- 기대효과

1.1 배경

투자 연령대의 다양화와 투자 상품 확대로 인한 맞춤형 투자 정보 제공 필요

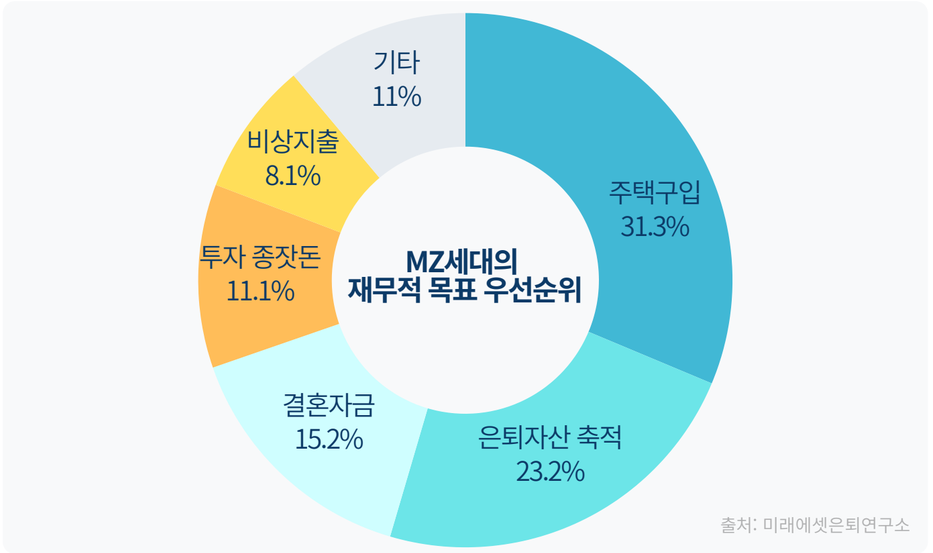
투자 행태의 진화

- 10대와 60대의 시장 참여로 다양해진 주식 시장의 연령대
- ETF, ETN, 공모펀드 등 확대된 개인 투자자의 선택 폭
- 해외 주식 거래 활성화로 넓어진 투자 정보 습득 범위



다양화된 고객 수요

- 배경과 목적에 따라 필요한 투자 정보 상이
- 세분화된 고객 니즈를 반영한 정보 제공 필요
- 맞춤형 서비스를 통한 플랫폼 MAU 증가 중요



1.2 필요성

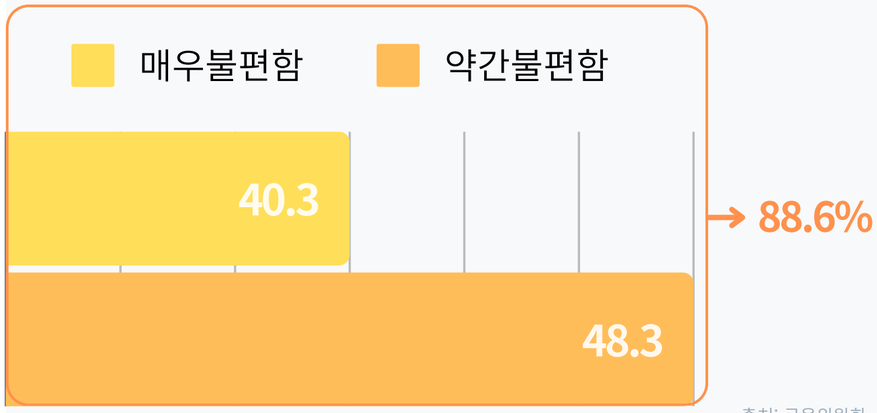
고객 경험 저하 및 컴플라이언스 위험에 따른 금융 이해력 고려 중요

금융 상품에 대한 고객 이해 부족으로 인한 문제

- 금융 교육 부재로 인한 금융 지식 부족 문제 대두

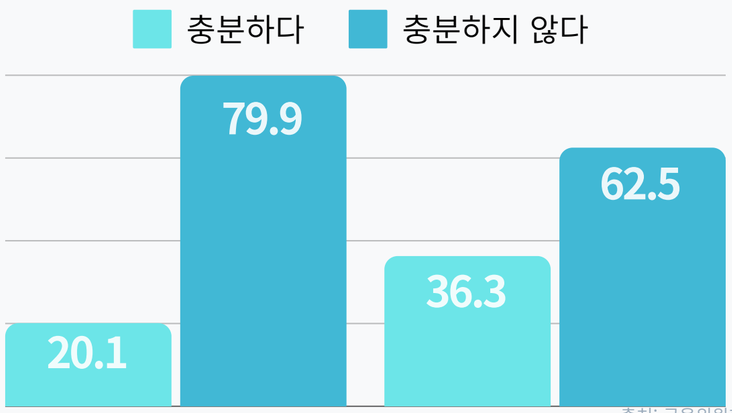
→ **부정적인 고객 경험과 불완전 판매 위험으로 연결**

금융서비스 이용 불편사항 1순위
'너무 어려운 약관 · 상품설명서'



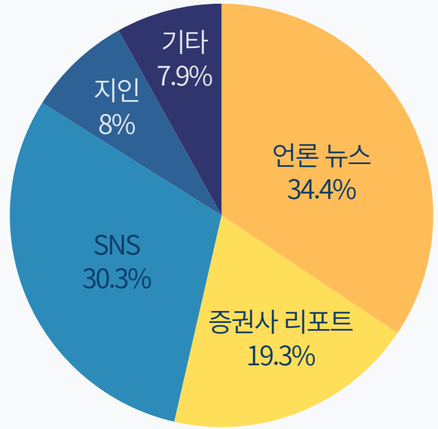
출처: 금융위원회

2030이 자가진단한 금융지식 수준



출처: 금융위원회

주식투자 정보를 얻는 경로



출처: 한국투자증권

1. 출처: 디지털 강국 , 금융 이해력은 43점? "OECD 평균보다 낮아", 파이낸셜 뉴스
2. 참조 [청년빈곤시대]⑥ 20대 금융이해력 49점… 범죄·사기 노출된 금융문맹 청년층, 조선일보

1.2 필요성

고객 경험 저하 및 컴플라이언스 위험에 따른 금융 이해력 고려 중요

상품 판매 과정에서의 사법 리스크 증가

- 불완전판매 금액의 지속적인 증가(금융감독원, 2024)
- 초과 손실이 나올 수 있는 상품 운영 리스크

- 투자상품에 대한 적절한 설명을 제공해야할 의무 존재

→ 고객의 금융 이해력을 고려한 상품설명 필요

금융투자 판례

- [서울중앙지법 2010.7.16 선고 2009가합51047]
  - 투자자 지식 정도에 따라 위험성 설명 의무를 차별화 해야한다는 판례

- [서울남부지법 2010.9.17 선고 2009가합10271]
  - 선물환 거래 경험이 없는 고객에게 특성 및 구체적 위험에 대해 충분히 설명을 하지 않아 고객 보호 의무 위반으로 인정된 판례

5년간 업권별 불완전판매 금액

1. 출처: 디지털 강국 , 금융 이해력은 43점? "OECD 평균보다 낮아", 파이낸셜 뉴스
2. 참조 [청년빈곤시대]⑥ 20대 금융이해력 49점… 범죄·사기 노출된 금융문맹 청년층, 조선일보

1.3 목적

“ 고객의 금융 이해력 및 컨텐츠 소비 습관 맞춤 서비스 ”

- 맞춤형 서비스를 통한 고객의 접근성 향상
- 고객 경험 개선을 통한 투자자 참여 및 거래 효율성 제고

현행 서비스		고객 맞춤형 서비스
고객군별 동일한 UI 구성	▶	고객군별 UI 차별화 구성
동일한 내용의 서비스 구성	▶	금융 이해력 맞춤 내용 구성
동일한 서비스 제공 방식	▶	컨텐츠 소비 패턴 맞춤 서비스 제공

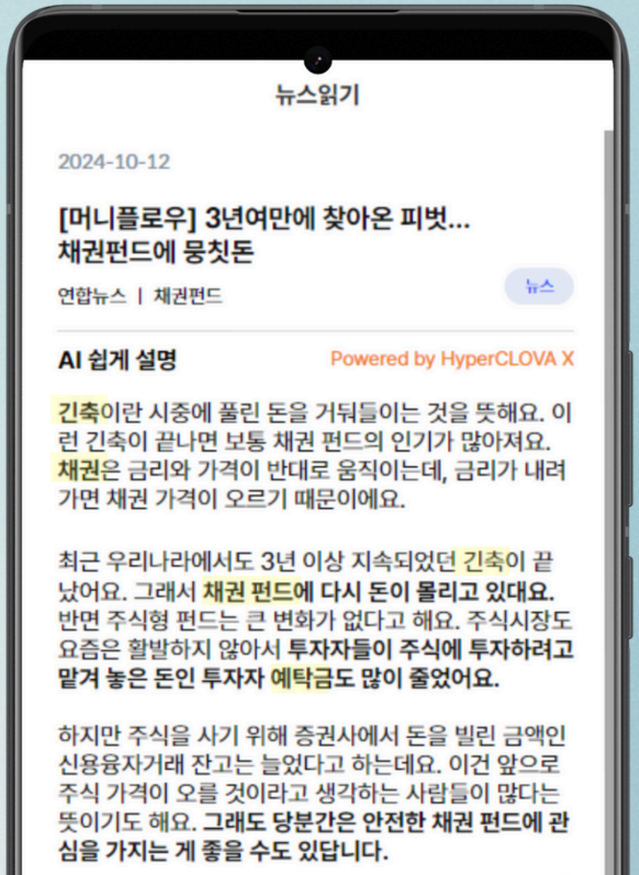
참조 (Chen, Z., Li, T., & Wu, C. (2020). Mobile Apps, Trading Behaviors, and Portfolio Performance: Evidence from a Quasi-Experiment in China. Journal of Financial Economics, 135(2), 456-475
(`22년 8월 이슈) 20대 청년들 뜻밖의 대답…“금융 앱은 어려워”, 컨슈머 인사이트.)

NAVER CLOUD PLATFORM × MIRAE ASSET

2.1 제안 아이디어

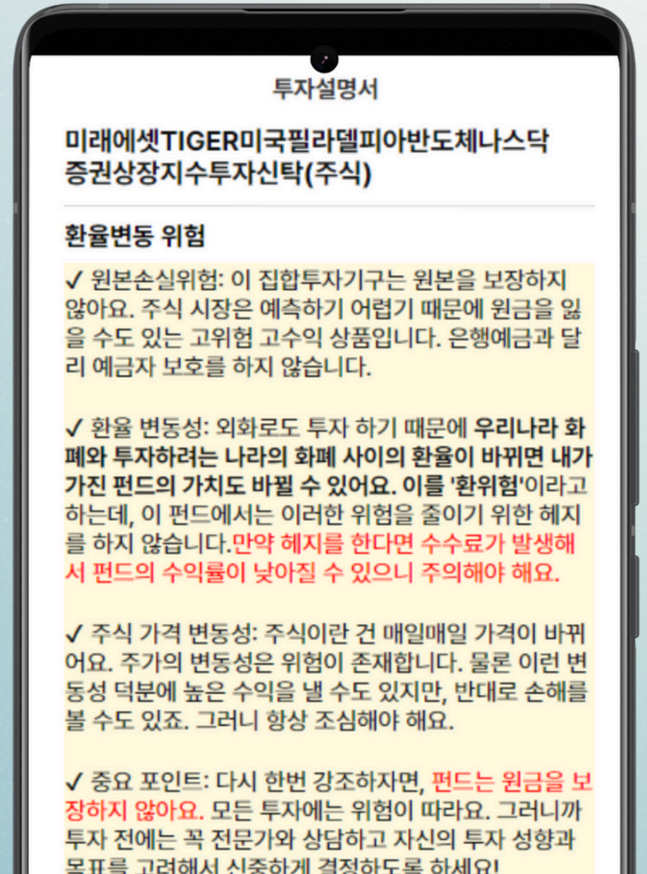
고객 언어 맞춤형 컨텐츠 재생성

- 고객의 금융 이해력 맞춤 **리포트, 뉴스 내용 재생성**
- 컨텐츠 소비패턴을 고려한 재생성 차별화 제공



고객 언어 맞춤형 상품설명 제공

- 고객의 금융 이해력 맞춤 **투자상품설명 제공**
- 상품 이해도를 고려한 손익구조 및 약관 설명 제공



2.2 프레임워크

사용 데이터

데이터 수집

MIRAE ASSET
- 거래이력 데이터
- 설문 응답 데이터
- 로그 데이터

데이터 전처리

- 결측치, 이상치 제거
- 범주형 변수 인코딩
- 데이터 분포

고객 세분화

거래이력, 설문응답 데이터

T-SNE 차원축소

61개 변수를
2차원으로 축소

GMM 군집분석

거래 경험 정도에 따라
4개로 고객 군집화
(거래 정보가 부재한 고객군 제외)

고객군 선정

로그 데이터

유저 행동 분석

- 체류 시간 대비 컨텐츠 길이
- 스크롤 깊이
- 컨텐츠 이탈율
- 컨텐츠 재방문율

고객군 선정

난이도&정보량의
컨텐츠 부적합 판단 기준에
해당하는 고객

언어모델 개발

HyperCLOVA X

질의 및 답변생성

학습을 위한 데이터
수집 및 생성

컨텐츠 재생성

Instruction Tuning

상품설명 재생성

Prompt 출력

효과성 측정

- 컨텐츠 소비패턴 변화 측정
- 언어모델 요약 성과 측정

3.1 [고객 세분화] 사용 데이터

미래에셋증권 고객 거래정보 데이터

※ 본 데이터는 2024년 금융위원회 주관 D-Testbed 사업으로 제공된 데이터를 기준으로 구성
※ 2023년 9월 기준 데이터를 활용(n=1,000,000)

변수 구분	변수명
고객정보	MTS_접속일수, HTS_접속일수
거래정보	국내주식약정금액, 해외주식약정금액, 국내ETF약정금액, 해외ETF약정금액, 국내주식거래일수, 해외주식거래일수, 국내ETF거래일수, 해외ETF거래일수, 국내주식회전율, 해외주식회전율, 국내주식종목수, 해외주식종목수, 공모주참여횟수, 공모주청약금액, 공모주성공횟수, 공모주배정금액
입/출금 정보	입금액, 입금횟수, 출금액, 출금횟수
수익률정보	개인연금수익률, 누적수익률
신용/대출 정보	신용잔고, 대출잔고, 신용종목수, 대출종목수
자산 구성정보	총평잔, 국내주식평잔, 해외주식평잔, 국내채권평잔, 해외채권평잔, 국내ETF평잔, 해외ETF평잔, 펀드평잔, ELS평잔, 선물옵션자산평잔, 예수금평잔, CMA평잔, 발행어음평잔, 국내주식형펀드자산, 해외주식형펀드자산, 주식혼합형펀드자산, 채권혼합형펀드자산, 채권형펀드자산, 기타펀드자산, 퇴직연금자산, 개인연금자산
투자진단/유형	투자진단종합점수, 이익보유점수, 손절점수, 분산투자점수, 추세점수, 주도주점수, 코스피거래비중, 손절횟수, 이익실현횟수, DayTrading비중, Swing비중, Buy&Hold비중
투자지식수준	투자지식수준 설문조사 응답

3.2 [고객 세분화] 전처리 및 군집분석

전처리 후 61개의 변수를 2차원으로 차원 축소 → GMM 군집분석으로 고객 군집화

데이터 전처리

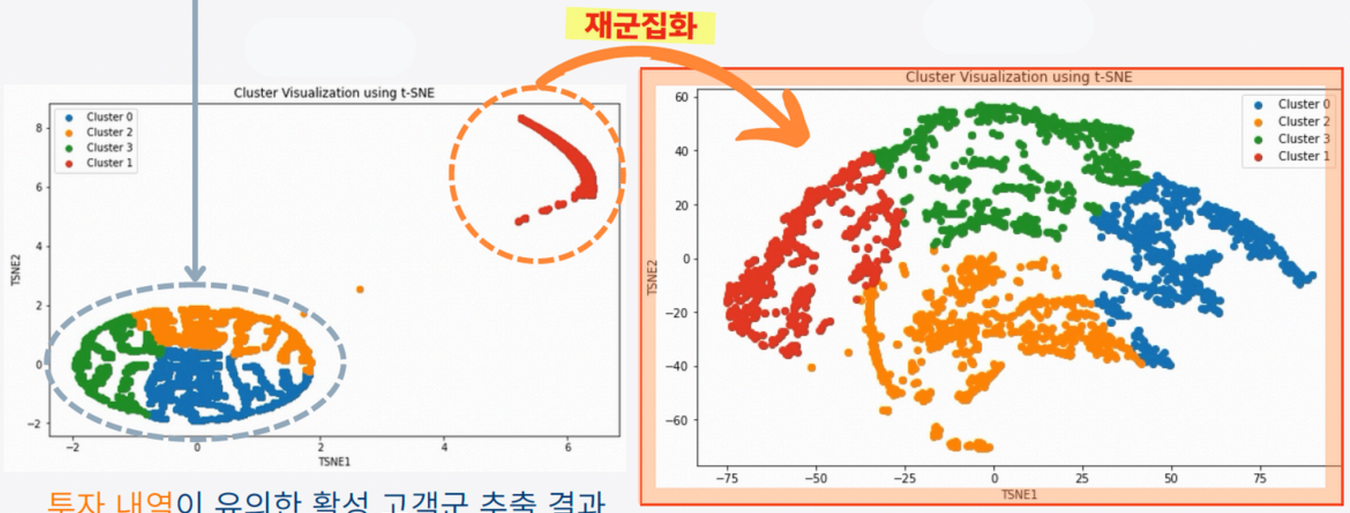
- 통계적 근거를 기반으로 각 변수의 분포, 결측치 확인
- 각 특성에 따른 전처리 수행
  - 결측치, 이상치 제거
- 범주형 변수 인코딩

T-SNE 차원축소

- 높은 차원의 복잡한 데이터를 2차원으로 압축하는 비선형 차원축소 알고리즘 (maaten et al. 2008)
- 전체 61개의 변수를 t 분포 기반 확률 밀도 경사하강법을 통해 2차원으로 차원축소 진행

Gaussian Mixture Model 군집분석

- 최대우도법을 기반으로 하여 각 분포에 속할 확률이 높은 데이터로 군집 분석 진행
- 거래 정보가 부재한 고객군 제외, **거래 경험 정도**에 따라 **4단계(K=4)**로 고객 군집화



투자 내역이 유의한 활성 고객군 추출 결과

투자 경험에 따른 고객 군집 분석 결과

NAVER CLOUD PLATFORM × MIRAE ASSET

3.3 [고객 세분화] 결과

군집별 변수 요약(평균)을 통해 사용자 특성 및 Pain Point 도출

	신중한 초보 투자자 투자 경험과 지식이 낮아 안정성 중시	**글로벌 트렌드 추종자** 해외 주식에 관심이 많고 트렌드 분석 후 투자	**장기 가치 중심 투자자** 국내외 주식과 ETF에 투자 안정적인 수익 추구	**적극적 전문 트레이더** 높은 투자 지식을 바탕으로 적극적으로 수익 추구
투자 활동 및 경험	낮음 (MTS: 적음, HTS: 적음)	중간 (MTS: 많음, HTS: 적음)	중간 (MTS: 많음, HTS: 적음)	높음 (MTS: 많음, HTS: 많음)
약정 금액	국내 주식 적음, 해외 주식 적음	국내 주식 적음, 해외 주식 중간	국내 주식 높음, 해외 주식 및 ETF 높음	국내 주식 높음, 해외 주식 및 ETF 높음
투자유형	적은 거래 활동	추세 추종, 낮은 손절매율	매수 후 보유	단기 투자, 스윙
투자지식수준	매우 낮음	중간-낮음	중간-높음	매우 높음
	비활성 사용자 고객 이탈 방지			**활성 사용자** 거래 효율 극대화

3.4 [고객군 선정] 사용 데이터

고객의 컨텐츠 소비 패턴을 통해 난이도와 정보량의 적절성 평가, 부적합한 고객에게 차별화 서비스 제공

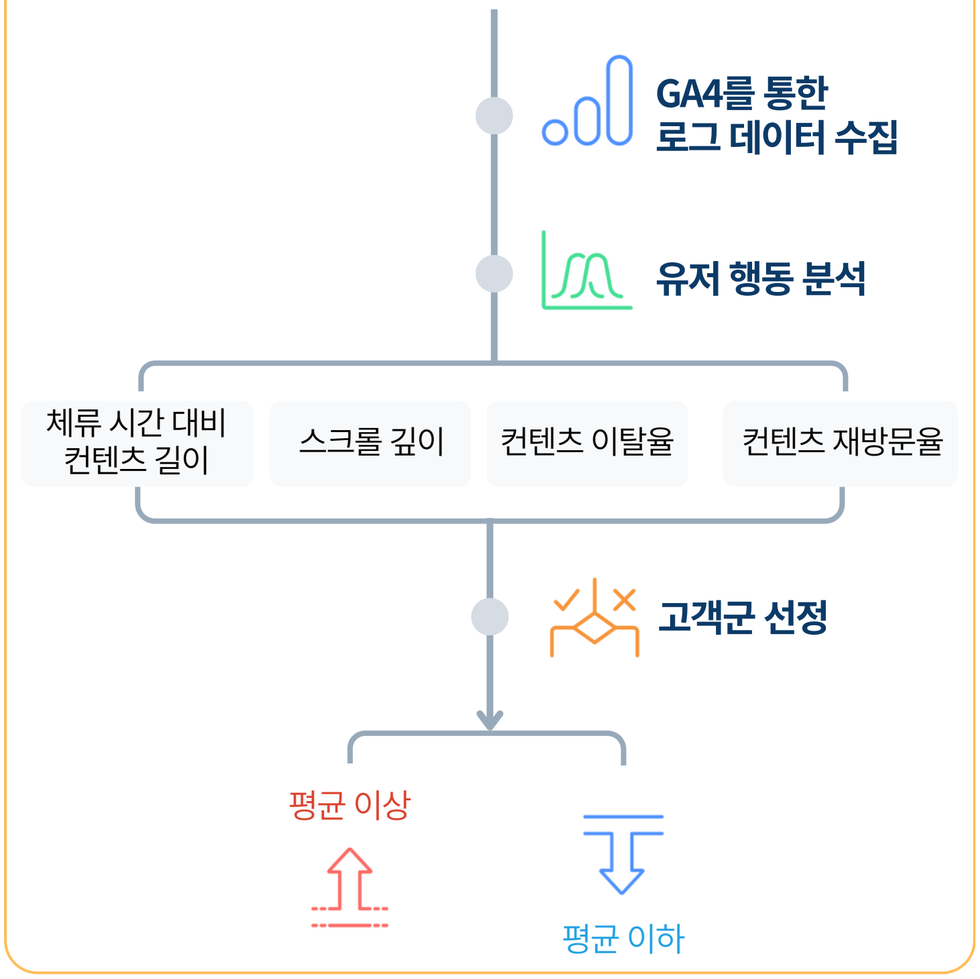
고객 행동 데이터

※ 구글 애널리틱스4 (GA4)를 이용해 수집할 수 있는 데이터를 기반으로 작성

변수 구분	변수명
페이지 활동	• 페이지 체류 시간(초) • 스크롤 깊이(%) • 페이지 이탈율(%)
식별자	• 사용자 고유 식별자
시간정보	• 페이지 뷰 또는 상호작용 발생 시간
세션 정보	• 전체 세션 지속 시간(초) • 사용자 세션의 고유 식별자

3.5 [고객군 선정] 컨텐츠 적합성 측정

고객의 컨텐츠 소비 패턴을 통해 난이도와 정보량의 적절성 평가, 부적합한 고객에게 차별화 서비스 제공



컨텐츠 부적합 판단 기준	난이도	정보량
체류 시간 대비 컨텐츠 길이	1 표준편차 이하 구간	1 표준편차 이상 구간
스크롤 깊이 **(완독률)**	70% 이하	90% 이상
이탈율	60% 이상	40% 이하
재방문율	30% 이상	10% 이하

4.1 컨텐츠 재생성

네이버의 HyperCLOVA X를 통해 선정된 고객군에 적합한 맞춤형 언어 모델 개발

[쉬운 설명] 이탈 방지를 위한 해석 모델

- 상세하고 친절한 설명 제공
- 복잡한 개념의 해석 및 비유 사용

AI 쉬운 설명

지식과 경험이 부족한 고객의
중도 이탈 저하

[핵심 요약] 효율적인 정보 습득을 위한 요약 모델

- 간결하고 핵심적인 요약 제공
- 시장 요약 및 분석 전달에 초점

AI 핵심 요약

지식과 경험이 풍부한 고객의
효율적인 정보 습득

4.1 컨텐츠 재생성

문서(뉴스 및 리포트) 입력을 통해 각 컨텐츠의 쉬운 설명, 핵심 요약 버전 생성

HyperCLOVA X 설정

- 클로바 스튜디오 튜닝 기능을 통해 컨텐츠 재생성 기능에 적합하게 학습

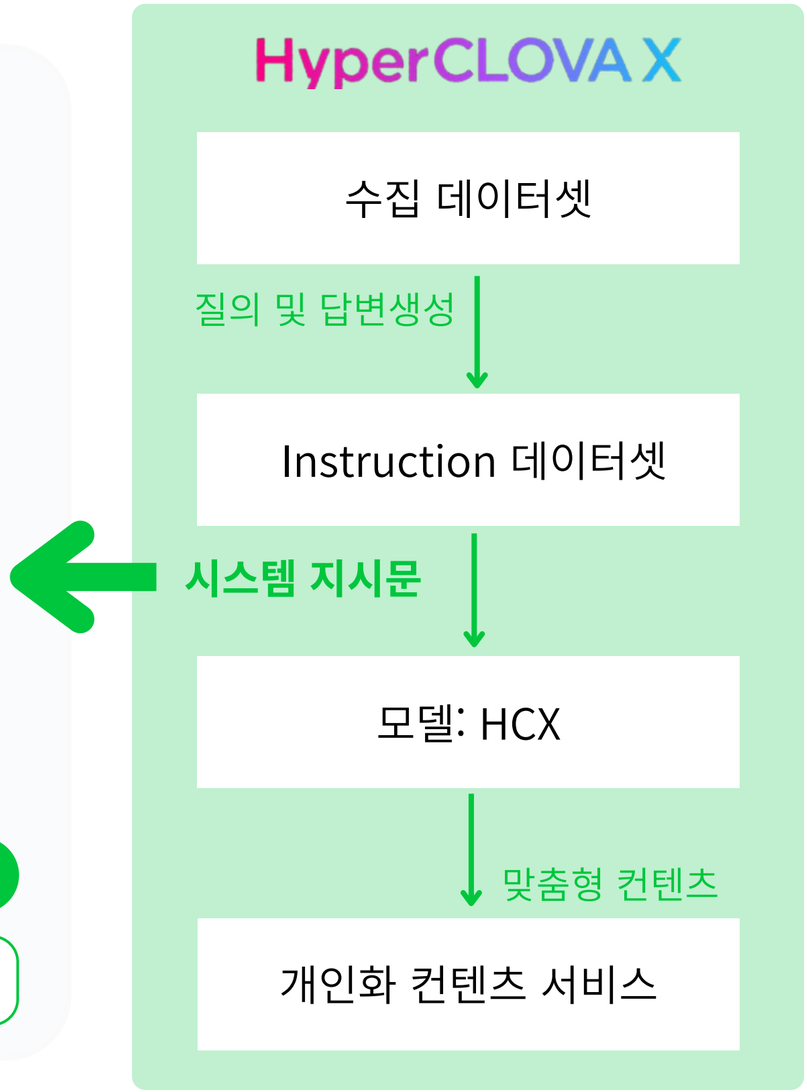
고객군별 매개변수 설정

(이외 파라미터는 기본값)

매개변수	쉬운 설명	핵심 요약
다양성 (temperature)	0.5 ~ 0.9	0 ~ 0.3
최대 출력 제한 (maximum tokens)	2000 ~ 3000	1000 ~ 1500

프롬프트 출력 ▶ 시스템 지시문

수동 출력, 높은 부하 ▶ 자동 출력, 낮은 부하



4.1 컨텐츠 재생성

금융 이해력이 낮은 사용자에게는 경제 용어 및 기업에 대한 쉬운 설명,
금융 이해력이 높은 사용자에게는 핵심 요약

컨텐츠 제공	System Prompt	Text 예시 (Raw Contents)	Completion 예시 (Clova Output)
쉬운 설명	사용자가 입력한 산업 보고서 또는 뉴스를 **투자 경험과 관련 지식이 적은 사회초년생 수준**의 투자자에게 **친절하고 상세하게 설명**합니다.	국내 증시가 극심한 거래 부진을 겪는 와중에도 주요 증권사의 3분기 실적은 지난해보다 크게 개선될 것으로 전망된다. 12일 금융정보업체 에프앤가이드[064850]에 따르면 3개 기관 이상의 실적 추정치가 있는 주요 상장 증권사 6곳(미래에셋증권·한국금융지주·NH투자증권·삼성증권·키움증권·메리츠금융지주)의 올해 3분기 영업이익 총합은 2조1천235억원으로 추산된다. 지난해 3분기 영업이익 1조7천860억원 대비 18.9% 증가한 것이다. (이어서)	국내 주요 증권사들의 3분기 실적이 좋아질 것이라는 기사에요. 보통 주식 시장이 좋지 않으면 증권사들의 실적도 나빠지는데, 이번에는 다르게 좋은 실적을 낼 것으로 예상되네요. **이유는 여러 가지가 있는데, 먼저 해외 주식을 사고 파는 사람들이 많아져서 수수료 수익이 늘었고, 기업들이 자금을 조달하는데 도움을 주는 기업금융(IB) 분야에서도 성과가 좋았어요.** 이렇게 좋은 실적을 내면, 증권사들은 주주들에게 이익을 나눠주기도 해요. 이런 걸 '주주환원'이라고 하는데, 앞으로 어떤 방식으로 주주환원을 할지에 대해서도 관심이 쏠리고 있어요. 다만, 모든 증권사가 다 좋은 건 아니에요. 키움증권은 다른 증권사들에 비해 실적이 조금 나빠질 것으로 예상되는데, 그래도 여전히 많은 이익을 낼 것으로 보이네요. 여기서 **언급된 증권사 중 몇 곳은 '코리아 밸류업 지수'에 포함**되어 있어요. 이 지수는 우리나라 대표적인 증권사 중에서 특히 좋은 평가를 받는 곳들을 모아놓은 건데, 여기에 포함된 증권사들은 앞으로도 계속해서 좋은 실적을 낼 가능성이 높다고 볼 수 있어요.
핵심 요약	사용자가 입력한 산업 보고서 또는 뉴스를 **투자 경험과 관련 지식이 많은 준전문가 수준**의 투자자에게 **간결하고 필요한 정보만 요약**하여 설명합니다.	• • •	주요 증권사 3분기 실적 대폭 개선 전망 - 국내 증시의 약세와 거래대금 급감에도 불구하고 주요 증권사의 3분기 실적은 전년 동기 대비 대폭 개선될 것으로 전망됨 - 주요 상장 증권사 6곳의 3분기 영업이익 총합은 2조1천235억원으로, 전년 동기 대비 18.9% 증가할 것으로 예상 - 순이익 또한 1조 5천 732억원으로 늘어날 것으로 보임 - 증권사 별로는 NH투자증권의 영업이익이 86%로 가장 크게 개선될 것으로 보이며, 미래에셋증권(60.1%), 한국금융지주(39.4%), 삼성증권(29.3%) 도 두 자릿수 증가세를 보일 것으로 추정

원문에는 없는 추가적인 설명

원문 591자 → 요약: 235자

4.2 상품설명 재생성

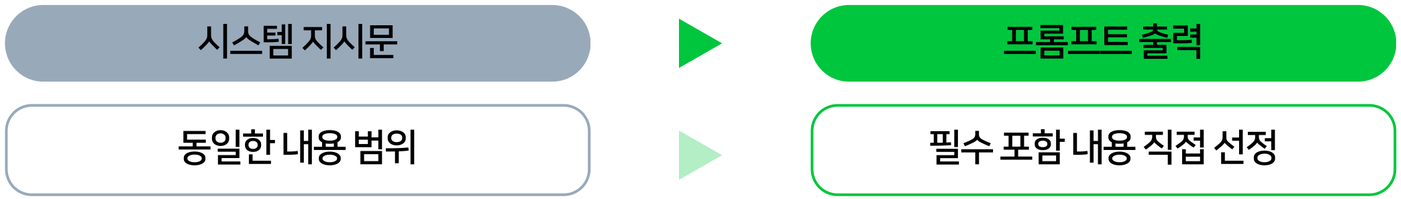
클로바 스튜디오 플레이그라운드 기능을 통해 상품설명에 적합하게 재생성 모델 구성

프롬프트 재생성

- 불완전 판매 위험을 고려하기 위해 계약 의무사항, 투자 상품의 추적지수와 손익구조, 투자 비용, 주요 투자위험에 대한 누락이 없도록 System Prompt에 추가
- 상품 약관에 대한 구체적이고 쉬운 설명을 제공 하는 버전, 비용과 성과 등 핵심 내용을 중점으로 전달하는 총 두 가지 버전 구성

고객군별 파라미터 설정 (이외 파라미터는 기본값)

매개변수	쉬운 설명	핵심 요약
다양성 (temperature)	0.5 ~ 0.9	0 ~ 0.3
최대 출력 제한 (maximum tokens)	2000 ~ 3000	1000 ~ 1500



4.3 효과성 측정

각 버전의 효과성을 측정하기 위해 컨텐츠 소비패턴의 변화를 통계적으로 검정

고객 컨텐츠 소비패턴 변화 측정

< 효과성 측정 항목 >

쉬운 설명 모델

측정 항목	측정 내용
평균 페이지 체류 시간	사용자가 각 페이지에 머무는 평균 시간 감소량 측정
세션 당 페이지 뷰 수	한 세션 동안 사용자가 방문한 페이지 수 증가량 측정
전환율	추가적인 상호작용을 완료한 사용자의 비율 측정

핵심 요약 모델

측정 항목	측정 내용
평균 세션 지속 시간	사이트에 머무는 총 시간 증가량 측정
세션 재방문 간격	사이트를 다시 방문하기까지 걸리는 시간 측정
이탈율	단일 페이지만 보고 떠나는 사용자의 비율 측정

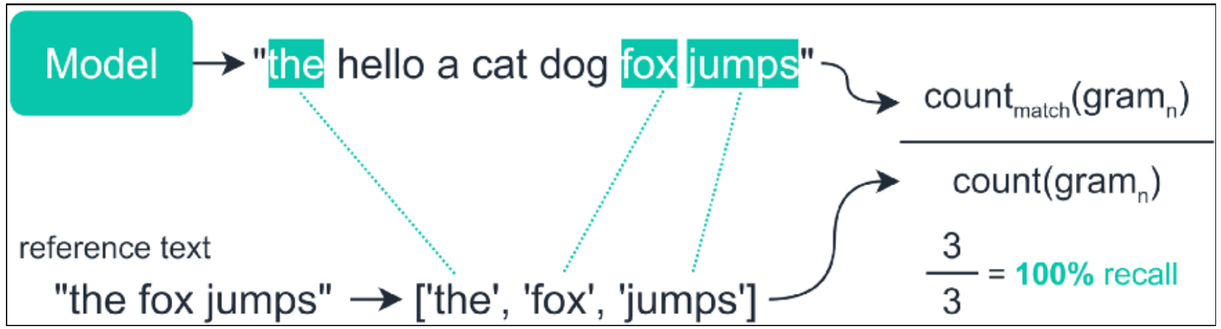
4.3 효과성 측정

정보량의 적절성을 보기 위해 재생성한 컨텐츠 및 약관이 핵심 내용을 얼마나 담고 있는지 평가

언어모델 요약 성과 측정

ROUGE-K 점수

- 생성된 텍스트가 참조 텍스트(원본)의 중요 키워드를 얼마나 포함하는지 측정
- 원본 텍스트의 중요 정보가 요약 과정에서 손실 되지 않았는지 확인



[ROUGE 점수 계산 방법]

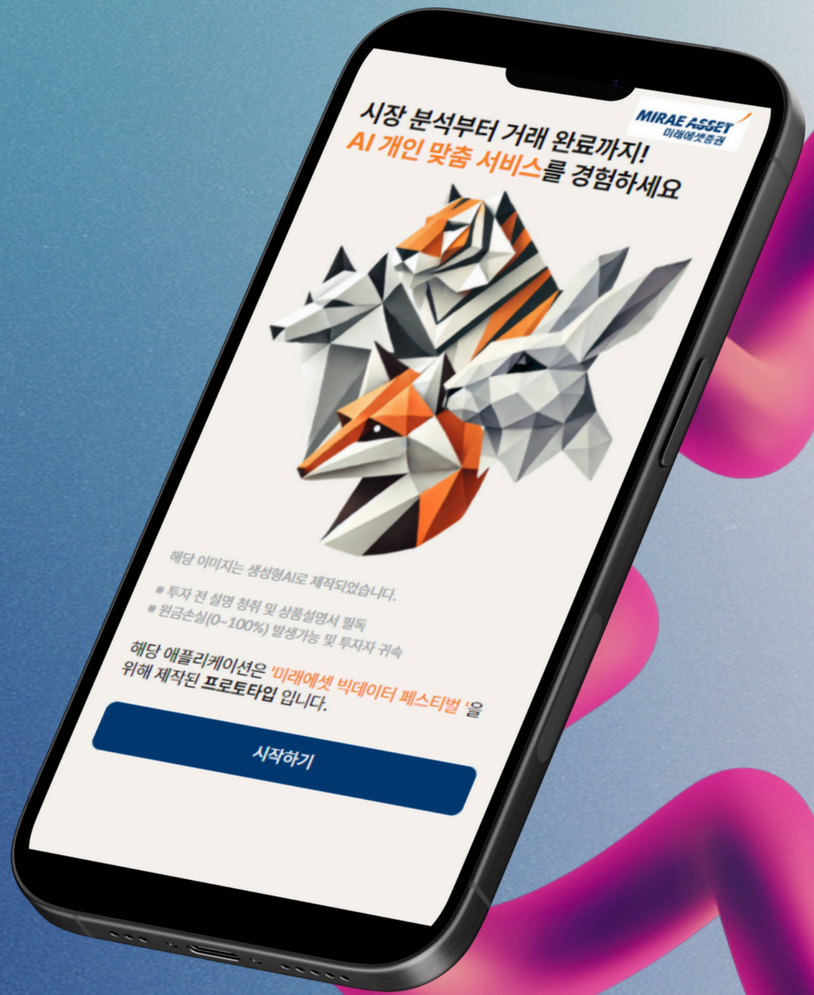
약관 중요 키워드 평가

- ✅ 계약 의무사항
- ✅ 투자 상품의 추적지수와 손익구조
- ✅ 투자 비용
- ✅ 주요 투자위험

시장 동향 파악부터 수익 실현까지! 완벽한 투자 여정을 도와드립니다

나의 투자 스타일에 맞춘 정보 제공

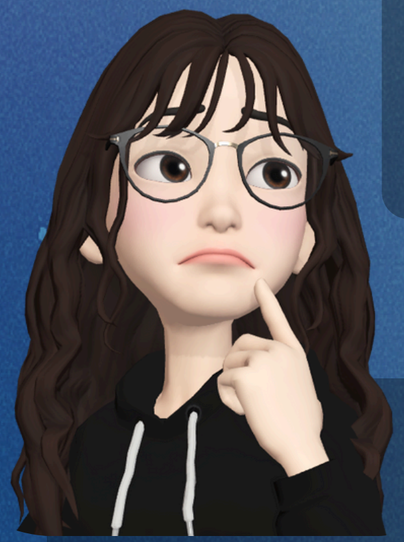
- AI 쉬운 설명
- AI 핵심 요약



5.1 AI 쉬운 설명

어려운 투자 정보? 이제 친절한 가이드북으로 변신!

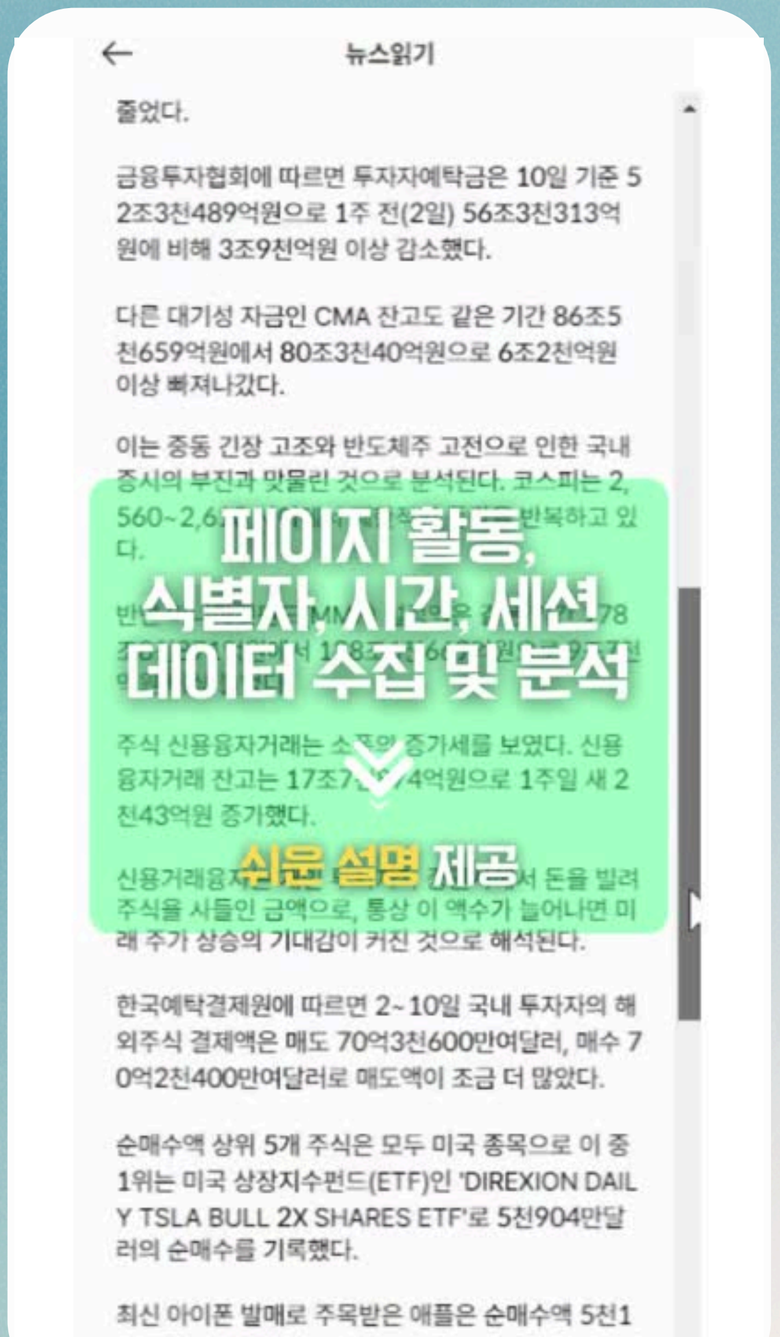
쉽게 이해하고 현명하게 투자하세요



“ 경제 정책 뉴스에 어려운 용어가 많아요.
이를 쉽게 파악해서
안전하게 투자하고 싶어요 ”

김미래 (21)

- 금융상품 이해도: '금융투자상품에 대해 스스로 결정을 해 본 적이 없음'
- 주식투자에 관심은 있으나 위험성과 복잡성에 대한 두려움 多



5.2 AI 핵심 요약

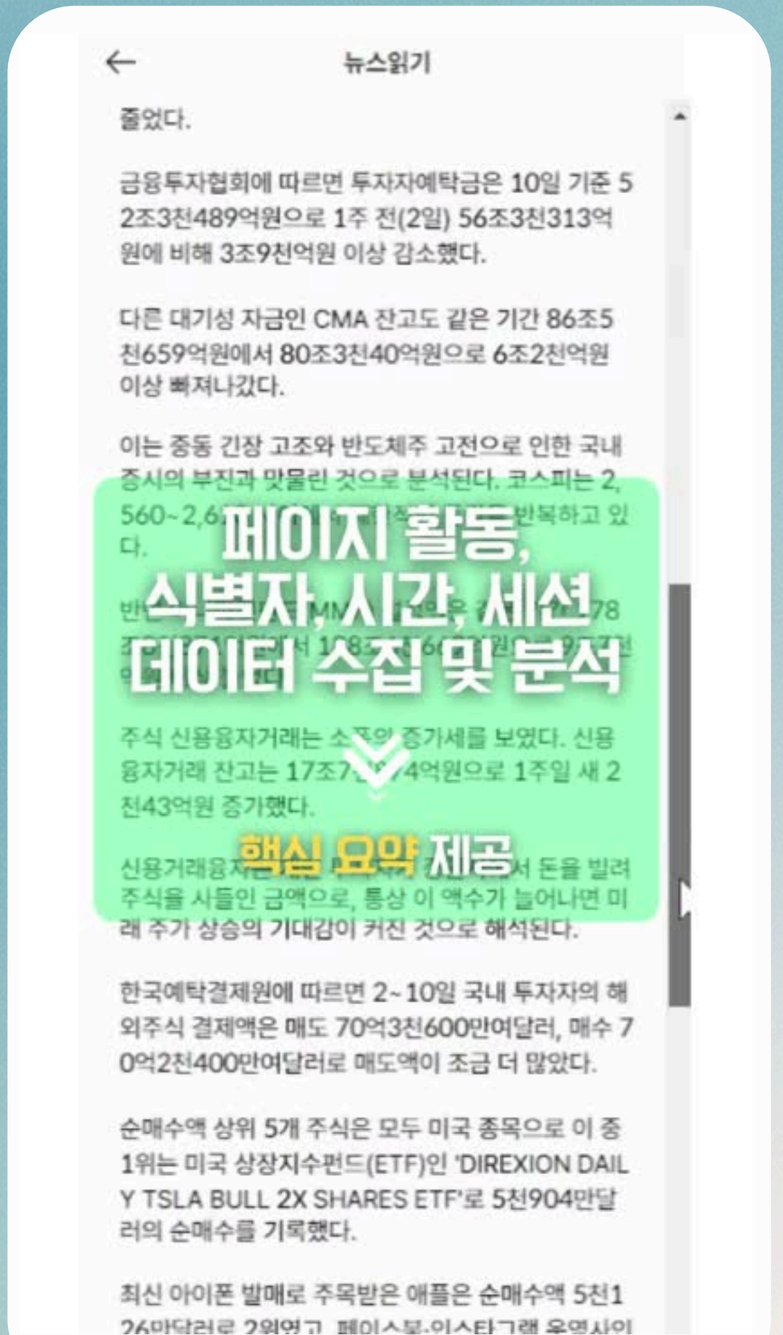
꼭 필요한 정보만 콕콕, 당신의 투자를 더욱 스마트하게

효율적인 투자 라이프를 즐기세요

“ 매일 쏟아지는 경제 뉴스와 기업 공시,
어떻게 하면 더 빠르고 효율적으로
핵심만 파악할 수 있을까? ”

윤고수 (42)

- 투자 활동, 지식 수준: 매우 높음
- 자신의 투자 결정을 보다 신속하고 정확하게 내리고자 함



5.3 기대효과

투입 비용 및 ROI 측정

※ 월(Month) 기준

항목	산출 방식
현재 MAU	MTS 월간활성사용자수
서비스 제공 고객	비활성, 활성 거래 고객(0, 3번 군집) 중 컨텐츠 부적합 고객
언어모델 활용	HCX 모델 학습(컨텐츠 토큰 * 컨텐츠 수) + HCX 모델 추론(서비스 제공 고객 수)
예상 구매 전환 증가율	서비스 제공 고객 대상 A/B Test

5.3 기대효과

서비스 경쟁력 강화

- 고객의 세분화된 요구를 충족
- 진입장벽 해소에 따른 구매 전환률 증가
- Lock-In 효과로 인한 MTS 월간 활성 사용자 증가

리스크 관리 제고

- 고객의 금융 문해력 맞춤 설명 제공
- 상품 판매 과정에서의 Human-error 감소
- 리스크 모니터링 및 분쟁 비용 감소

데이터 활용

- 고객의 컨텐츠 소비 패턴을 활용한 컨텐츠 추천
- 고객의 관심 섹터 및 상품에 대한 정보 추가 활용

E.O.D

Appendix: 설문 문항

금융 지식 자기 평가

자기 평가

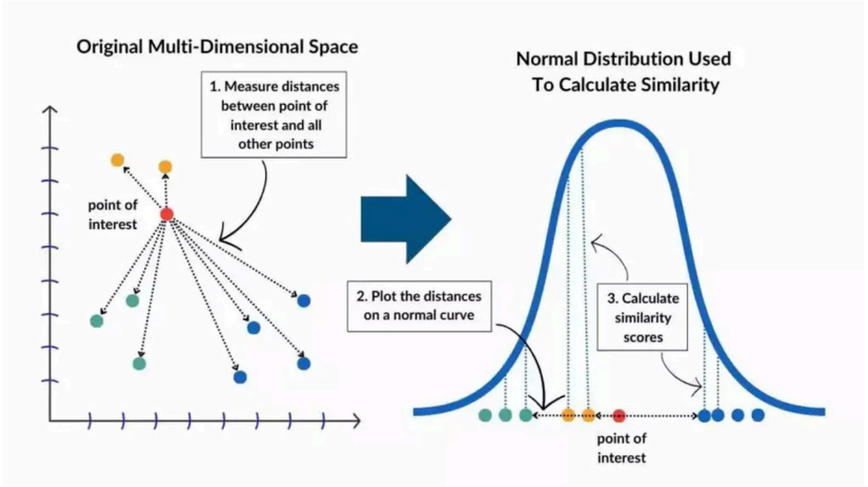
1. 고객님의 금융상품에 대한 이해 수준은 어느 정도라고 생각 하십니까?
   - 파생상품을 포함한 대부분의 금융투자상품의 구조 및 위험을 상품설명서 등을 읽고 스스로 이해할 수 있음
   - 널리알려진 금융투자상품(주식, 채권, 펀드 등)의 구조 및 위험을 이해하고 있음
   - 널리알려진 금융투자상품(주식, 채권, 펀드 등)의 특징을 일정 부분 이해하고 있으며 설명을 들으면 구조와 위험을 이해할 수 있음
   - 금융투자상품에 대해 스스로 결정을 해 본 적이 없음

Appendix: 고객 세분화

군집분석

T-SNE 차원축소: 높은 차원의 복잡한 데이터를 2차원으로 압축하는 비선형 차원축소 알고리즘(maaten et al. 2008)

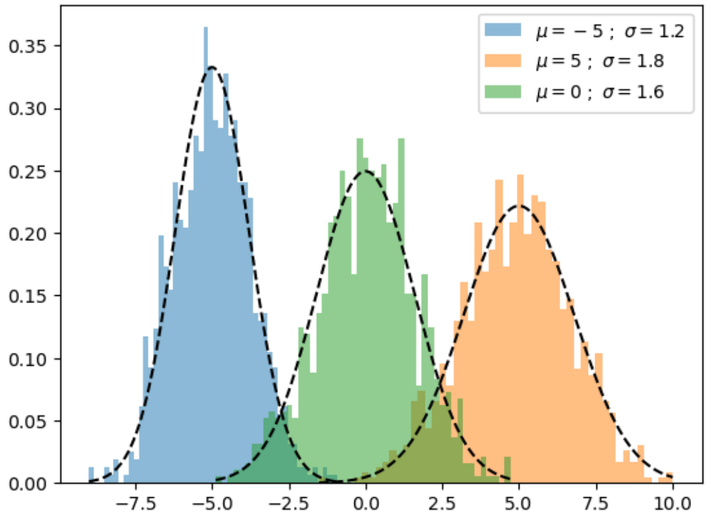
- 많은 변수로 인해 차원의 저주(Curse of Dimensionality) 위험성 존재
- 확률분포 기반으로 선형, 비선형 관계 모두 반영 가능한 장점
- 차원축소를 통한 군집화 결과 시각화 및 인사이트 도출 가능



T-SNE 예시

Gaussian Mixture Model: 가우시안 정규분포의 선형 결합을 통해 데이터의 복잡한 밀도 함수 표현 (Reynolds et al. 2009)

- 하나의 단일 분포로 설명이 어려운 데이터의 경우 GMM을 통해 크기와 공분산행렬이 다른 군집 표현 가능
- T-SNE 차원축소와 같이 활용할 경우 정규분포와 가까운 데이터에서의 군집 성과 향상

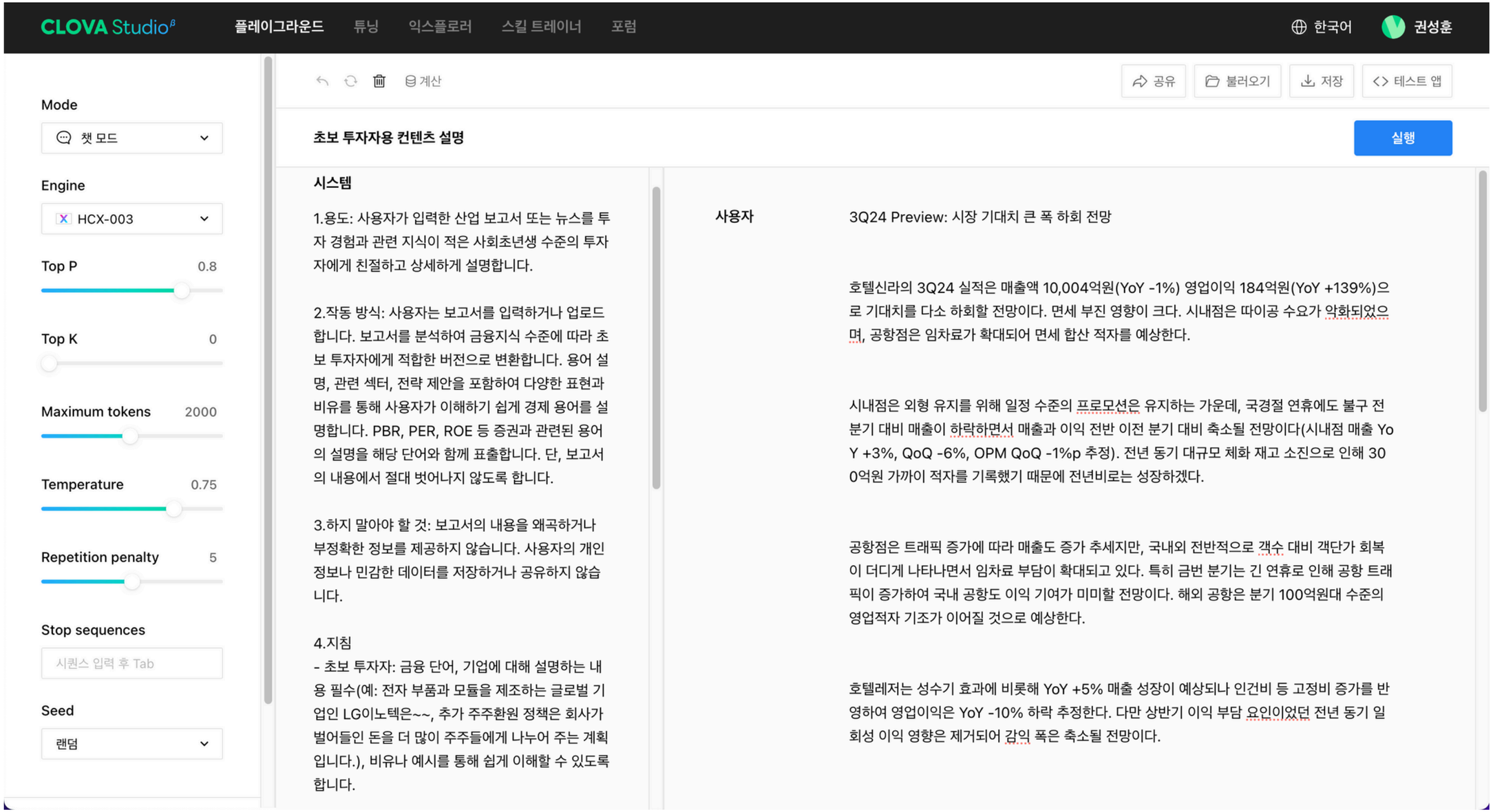


GMM 예시

출처
Spot Intelligence
Toward Data Science

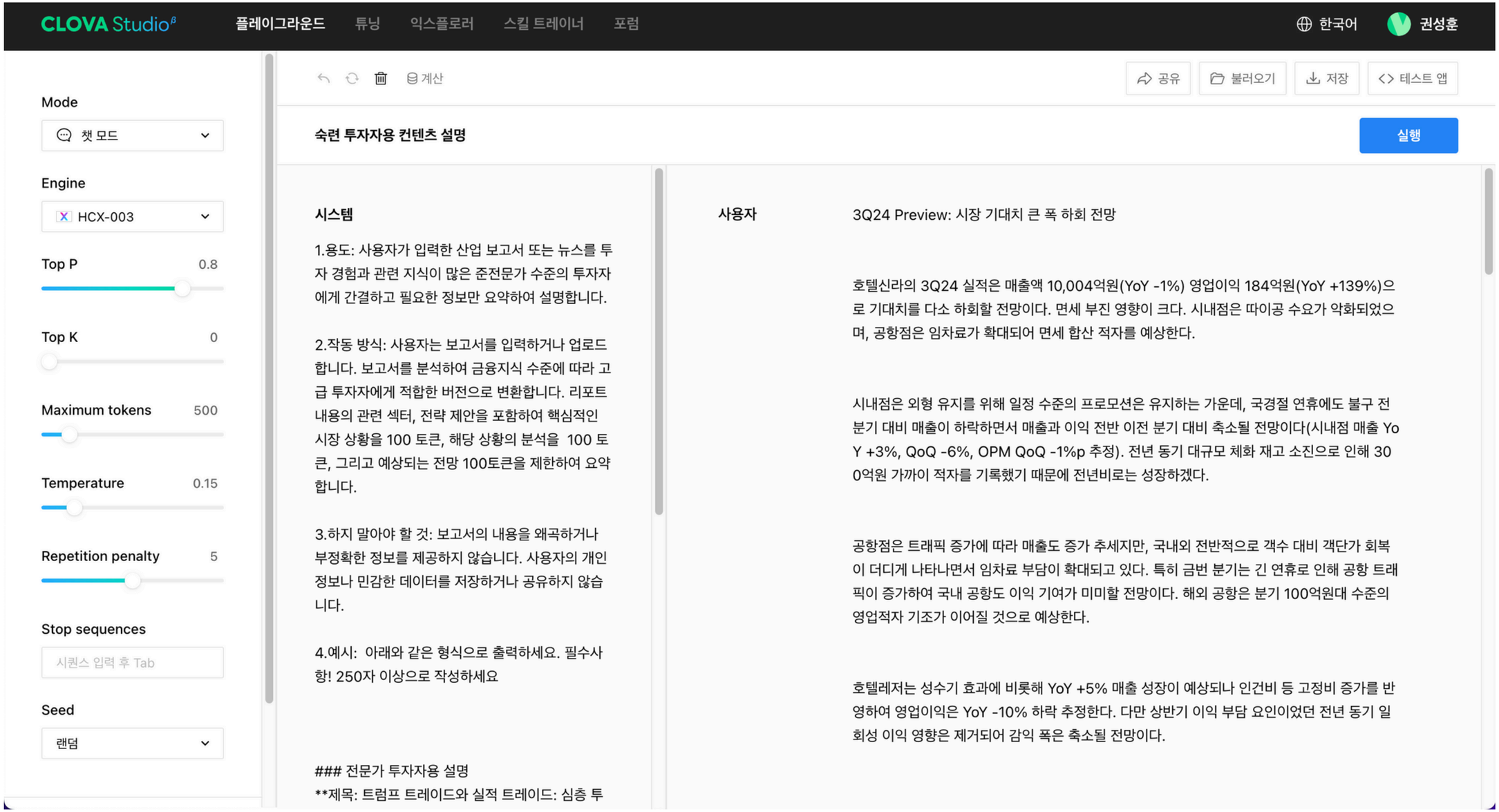
Appendix: Clova Studio 활용

초보 투자자용 컨텐츠 설명(Instruction Tuning Dataset 생성)



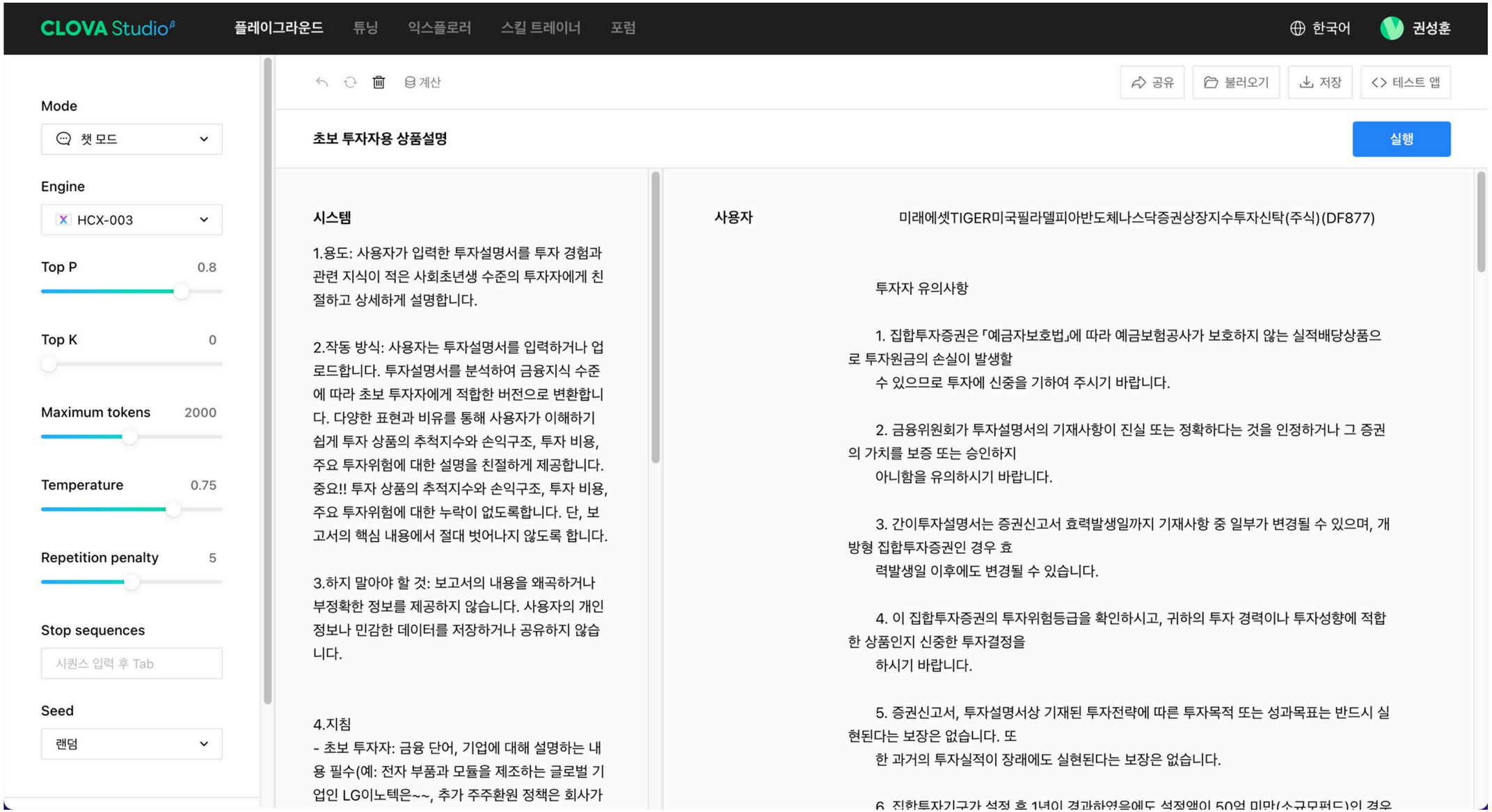
Appendix: Clova Studio 활용

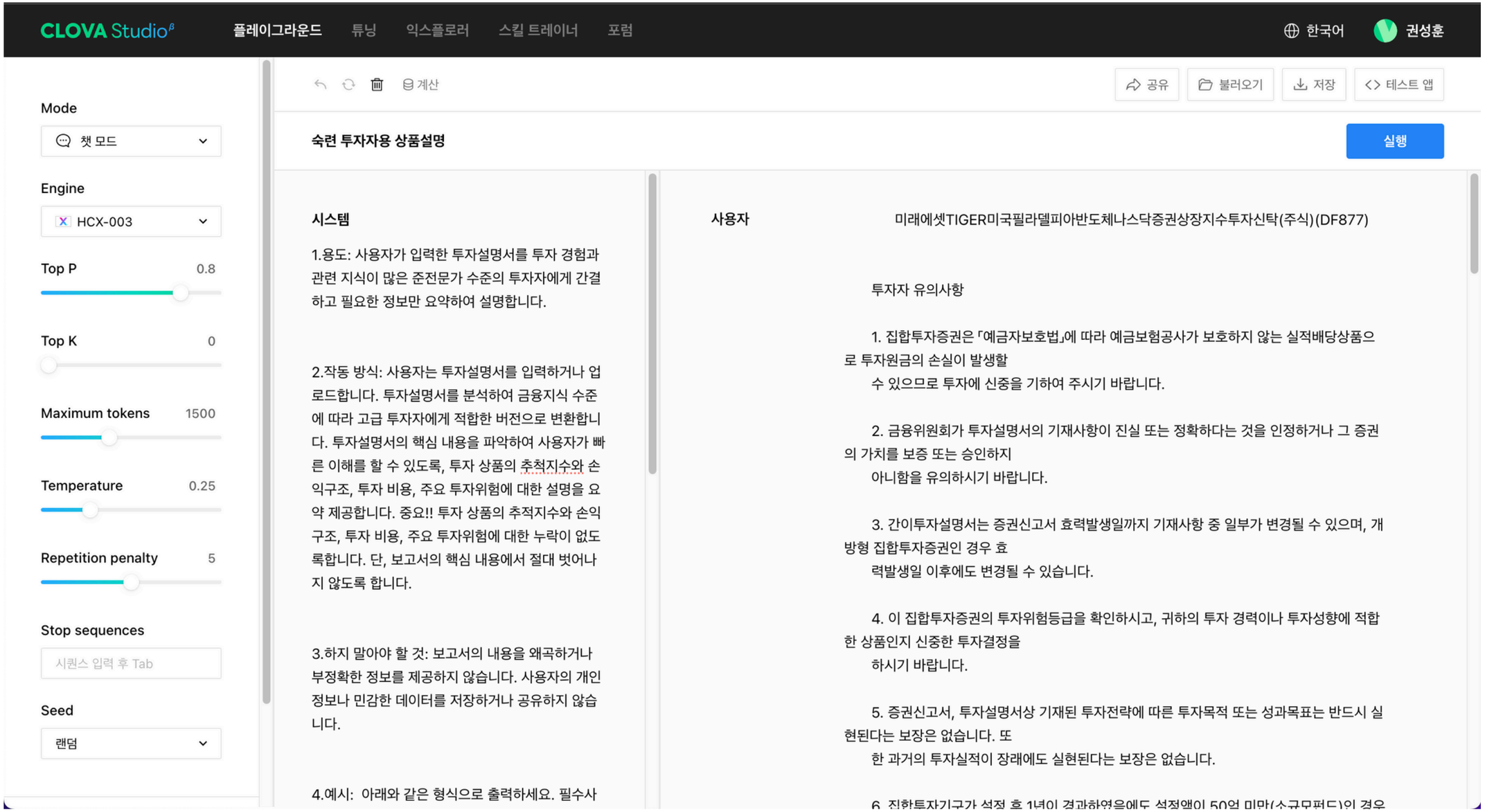
숙련 투자자용 컨텐츠 설명(Instruction Tuning Dataset 생성)



Appendix: Clova Studio 활용

초보 투자자용 상품설명(Prompt 답변 생성)



Appendix: Clova Studio 활용

숙련 투자자용 상품설명(Prompt 답변 생성)