

설문 기반 심리·행동 데이터를 활용한 투표 여부 예측 AI 알고리즘 개발

심리 성향 예측 AI 해커톤

4조: 사주가대세

김민찬 · 서우희 · 양은서 · 윤희주

목차

1. 데이터 탐색 (EDA)

2. 데이터 전처리

3. Feature Selection & Engineering

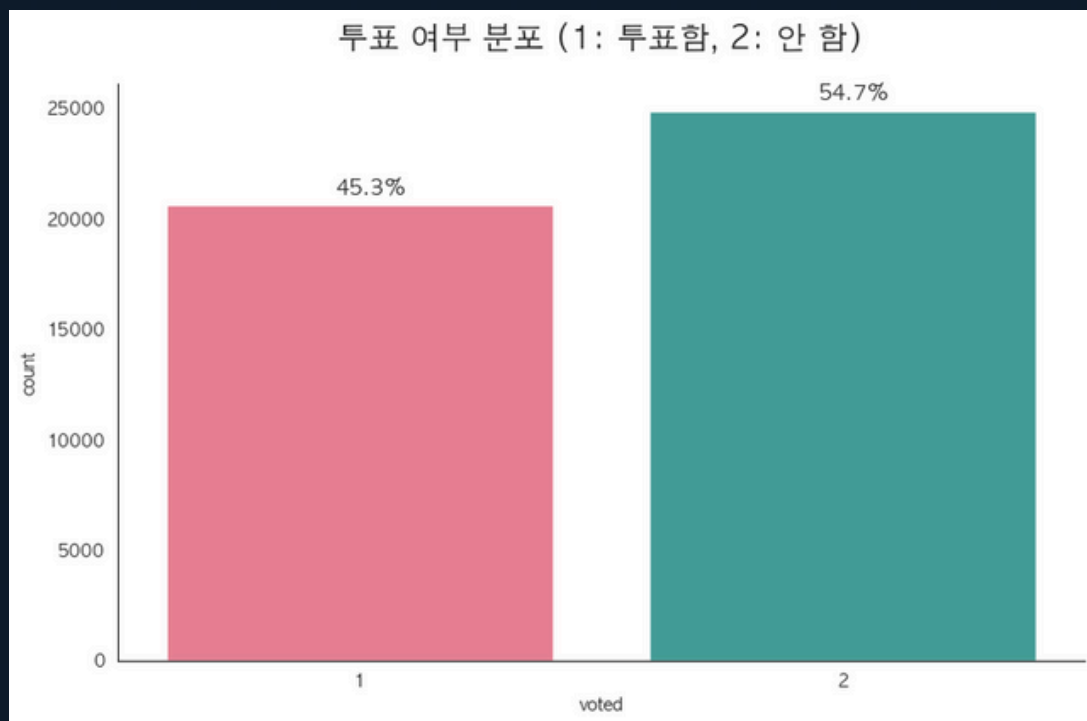
4. Deep Learning Modeling

5. Blending

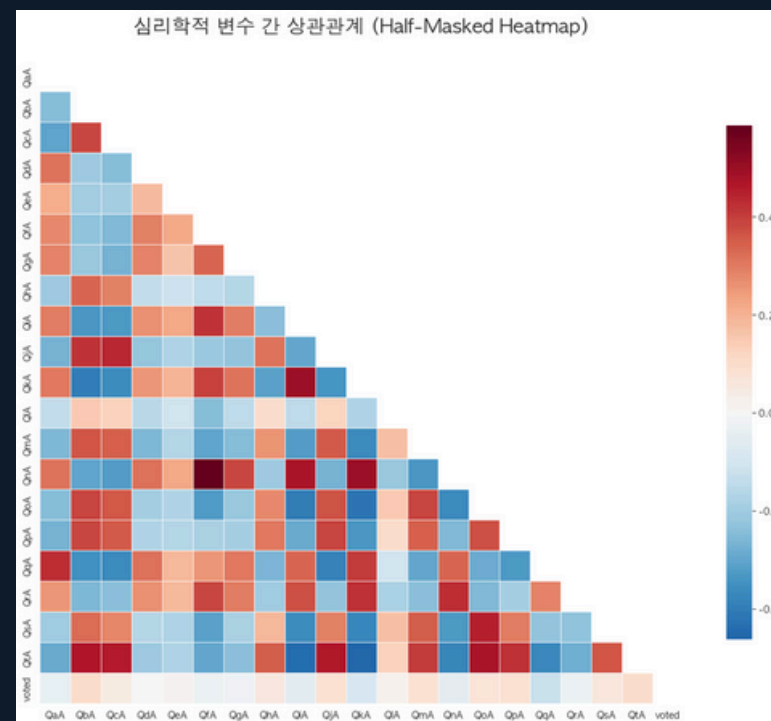
6. 결론 및 인사이트

"단순 통계를 넘어서는 인간 행동의 패턴 포착"

투표 여부 분포

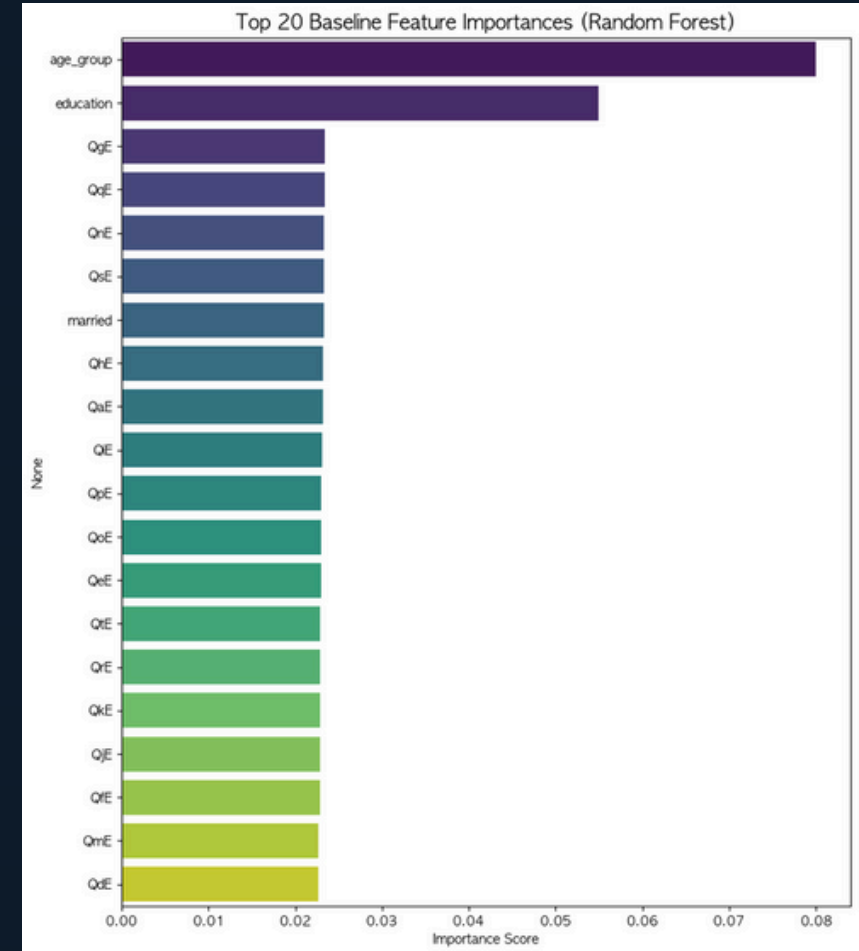


심리학적 변수 간 상관관계



"단순 통계를 넘어서는 인간 행동의 패턴 포착"

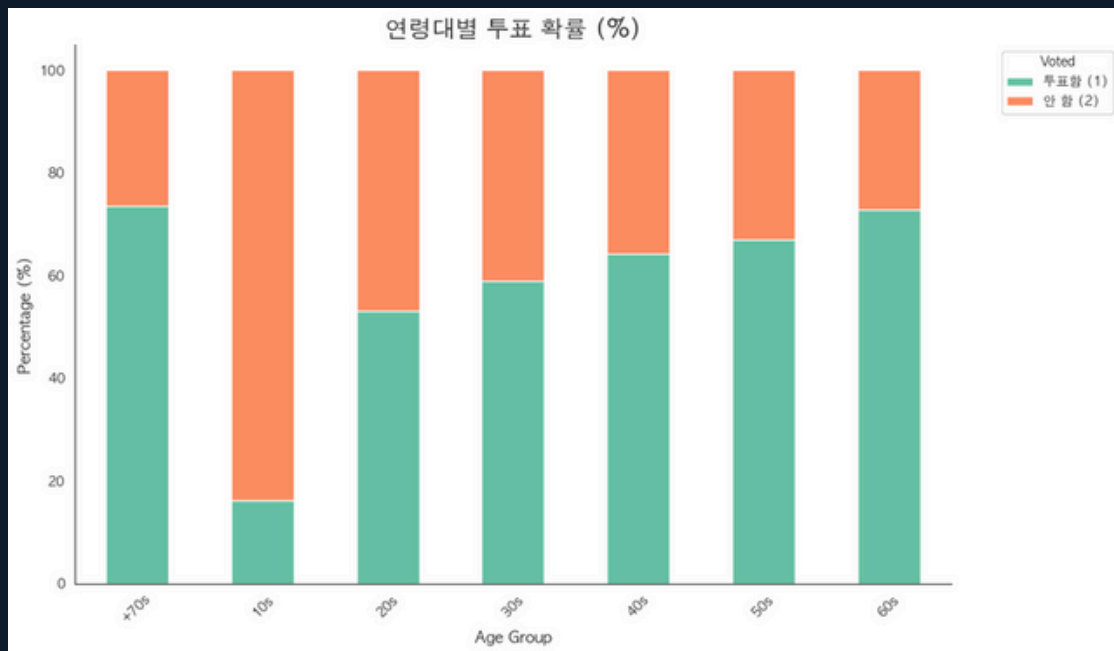
Top 20 변수 중요도



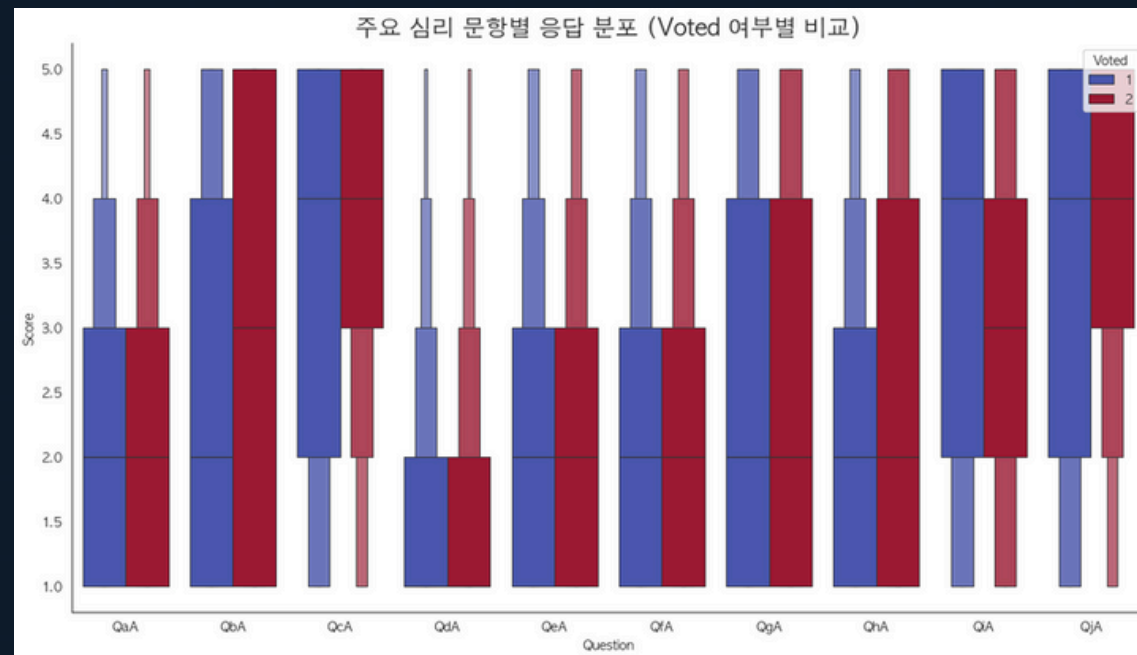
데이터 탐색

"단순 통계를 넘어선 인간 행동의 패턴 포착"

연령대별 투표 확률



TIPi 성격 지표



핵심 데이터 탐색

"단순 통계를 넘어선 인간 행동의 패턴 포착"



상관관계 분석



심리 질문 Qa, Qt가
투표 여부와 직결



연령대별 투표 확률



연령 ↑
투표 확률 비례 상승



TIPI 성격 지표

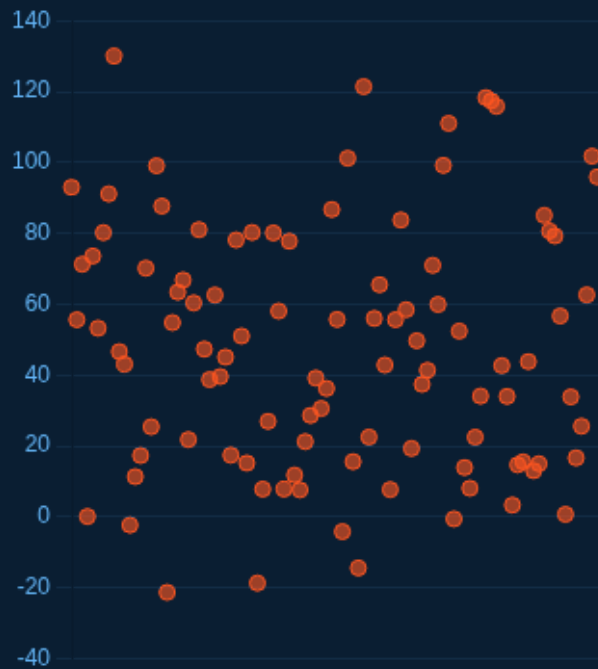


투표 '적극성'을
좌우하는 핵심 변수

데이터 전처리

"이상치 처리와 데이터 정제"

Raw Data



들쭉날쭉한 노이즈 분포 극단값과 이상치 포함

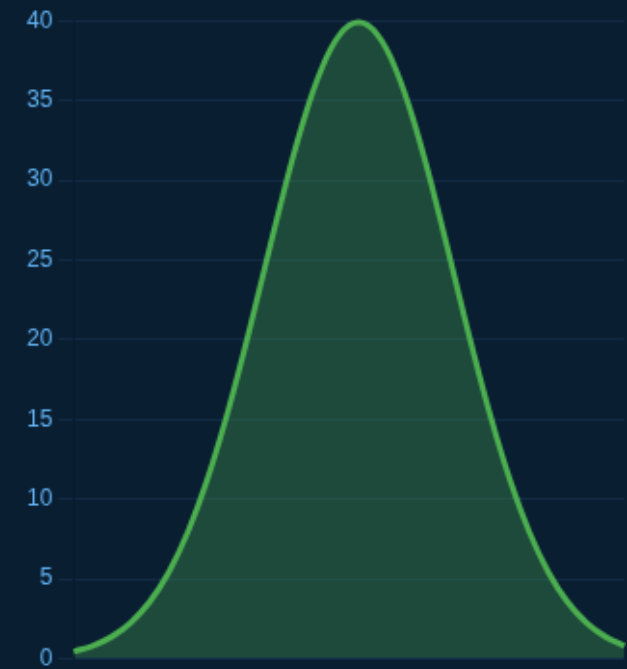


도메인 기반 필터링
도메인 지식을 바탕으로
극단값 제거



응답 시간 스크리닝
 $Q_E < 0.1초$
불성실 응답 제거

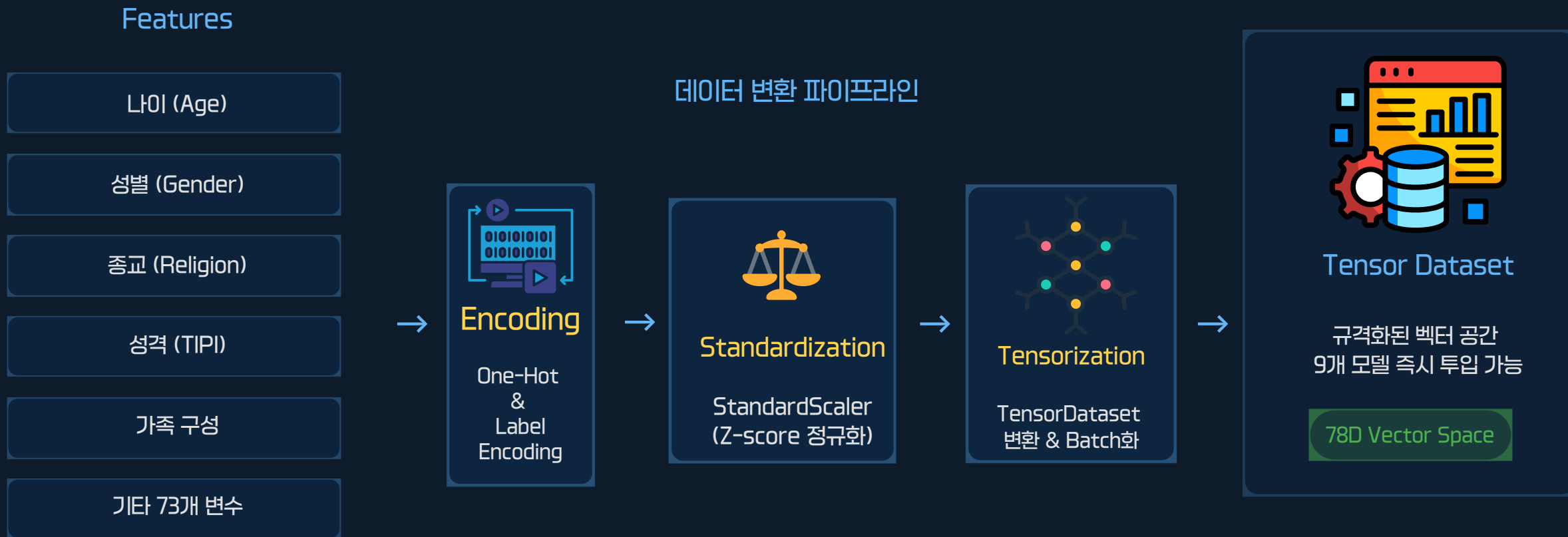
Clean Data



정규분포 형태의 정제된 데이터 신뢰도
95%+ 확보

피처 엔지니어링 9개 딥러닝 모델을 위한 최적화 구조 형성

"78개의 복잡한 변수를 하나의 유기적 벡터 공간으로 통합"

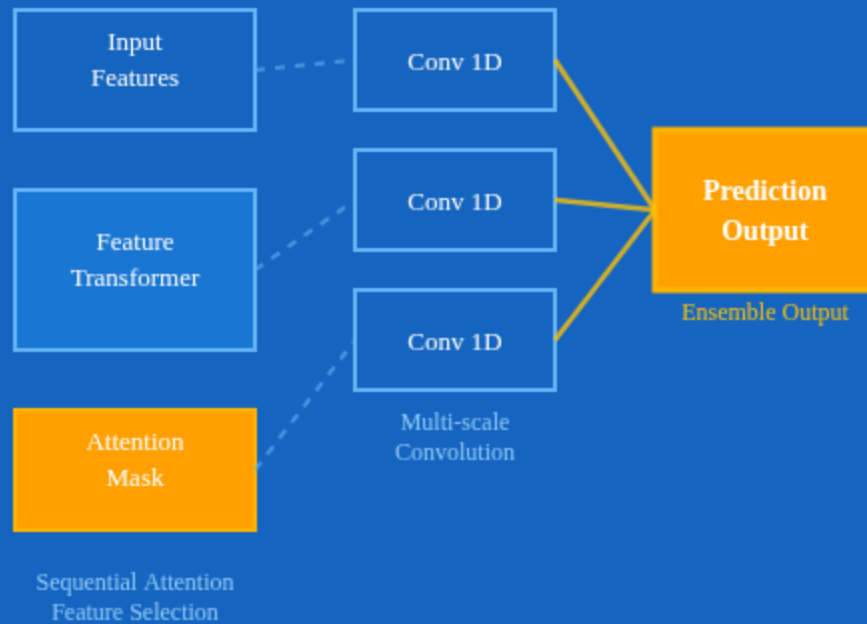


25%

Convolution TabNet — 정형 데이터 딥러닝 모델

메인모델 · 압도적인 예측력 보유

TabNet + Convolution Architecture



작동원리

설문지의 데이터를 훑으며 어떤 질문이 중요한지 '눈'으로 확인하고, 질문 간의 입체적인 관계를 찾아냄

핵심기술

TabNet Attention + Convolution
중요한 데이터에만 집중 + 주변 문항과의 관계를 스캔

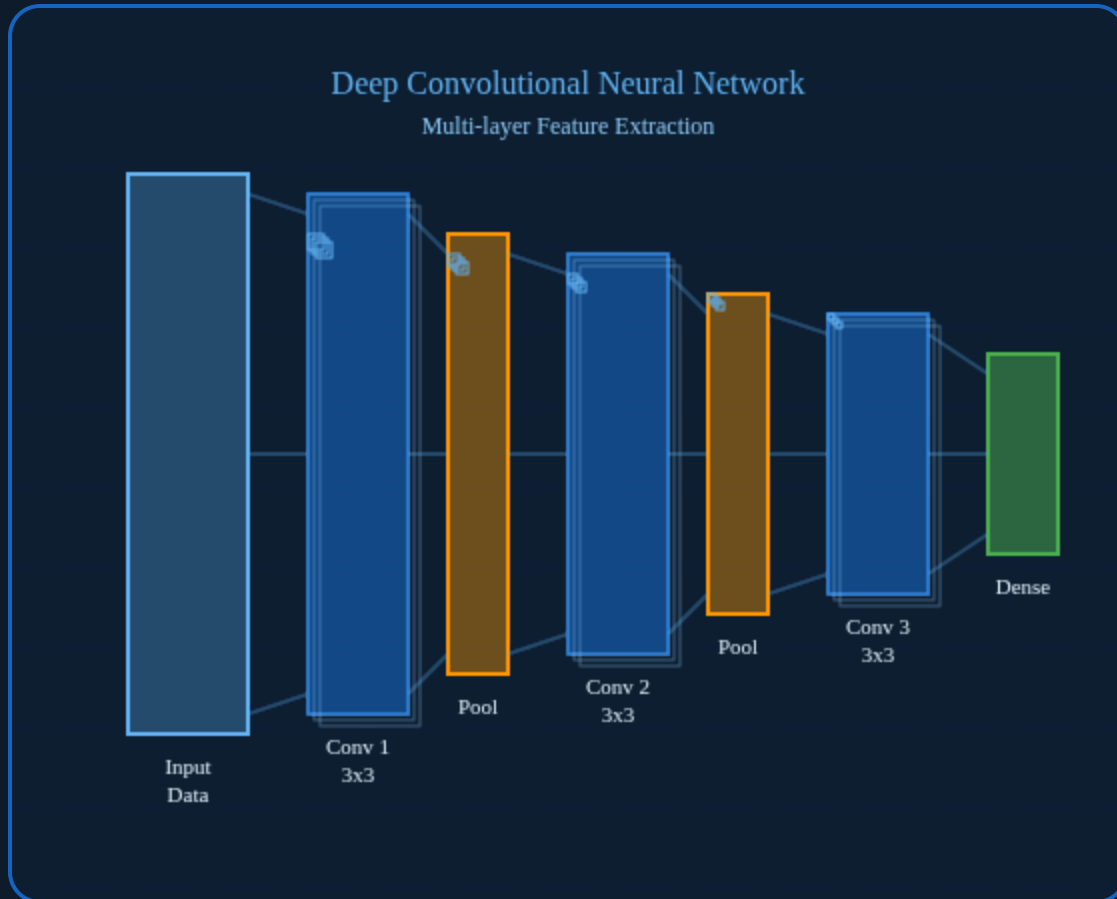
포인트

표 데이터의 강점과 이미지 패턴 분석의 장점을 결합하여,
가장 복잡한 심리 구조를 해석하도록 설계

Deep CNN

20%

질문 간의 복잡한 상관관계를 심층적으로 스캔



작동원리

수만 명의 응답을 하나의 '심리 지도'로 변환한 뒤, 특정 성향의 사람들이 공통적으로 보여주는 '답변의 형태'를 스캔

핵심기술

Deep 1D-CNN Layers 여러 층의 필터를 통해 데이터 속에 숨겨진 미세한 심리 패턴을 추출

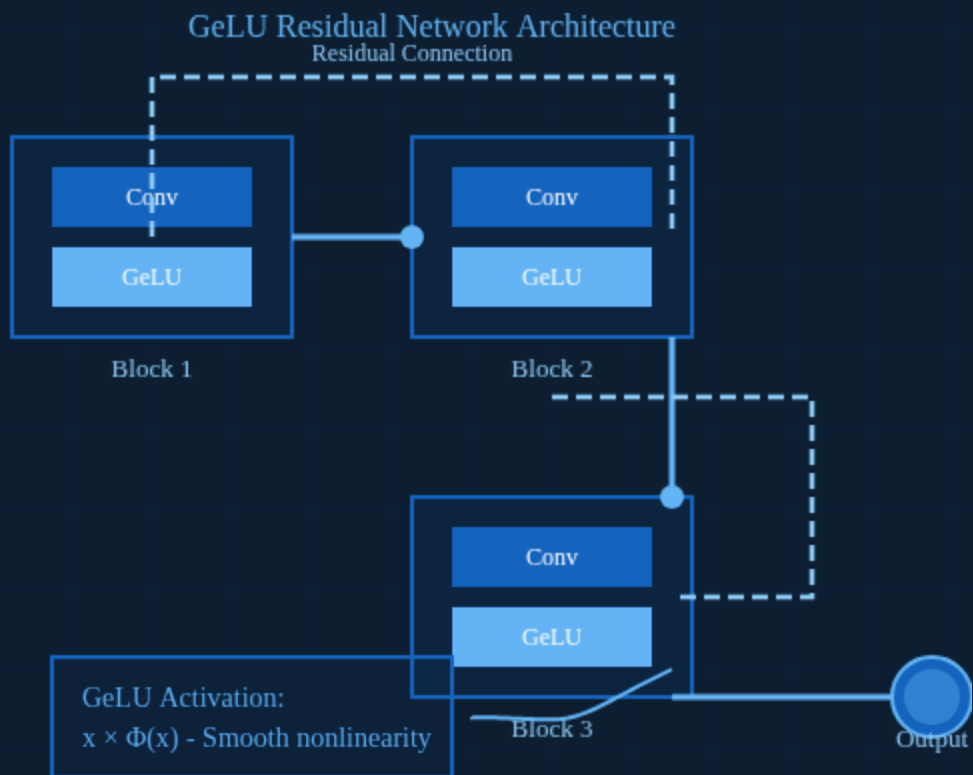
포인트

인간의 논리로는 설명하기 힘든 미세한 응답의 '리듬'과 '패턴'을 포착하여 예측의 정밀도를 높임

15%

GeLU ResNet

최신 활성화 함수와 잔차 연결로 학습 안정성 강화



작동원리

고민이 깊어질수록 중요한 정보를 잊지 않도록 지름길 (Skip Connection)을 만들고, 답변의 미묘한 차이를 부드럽게 연결해 해석

핵심기술

Residual Block 정보를 생생하게 전달하는 '지름길'과 GELU Activation 유연한 '비선형 활성화 함수'를 사용

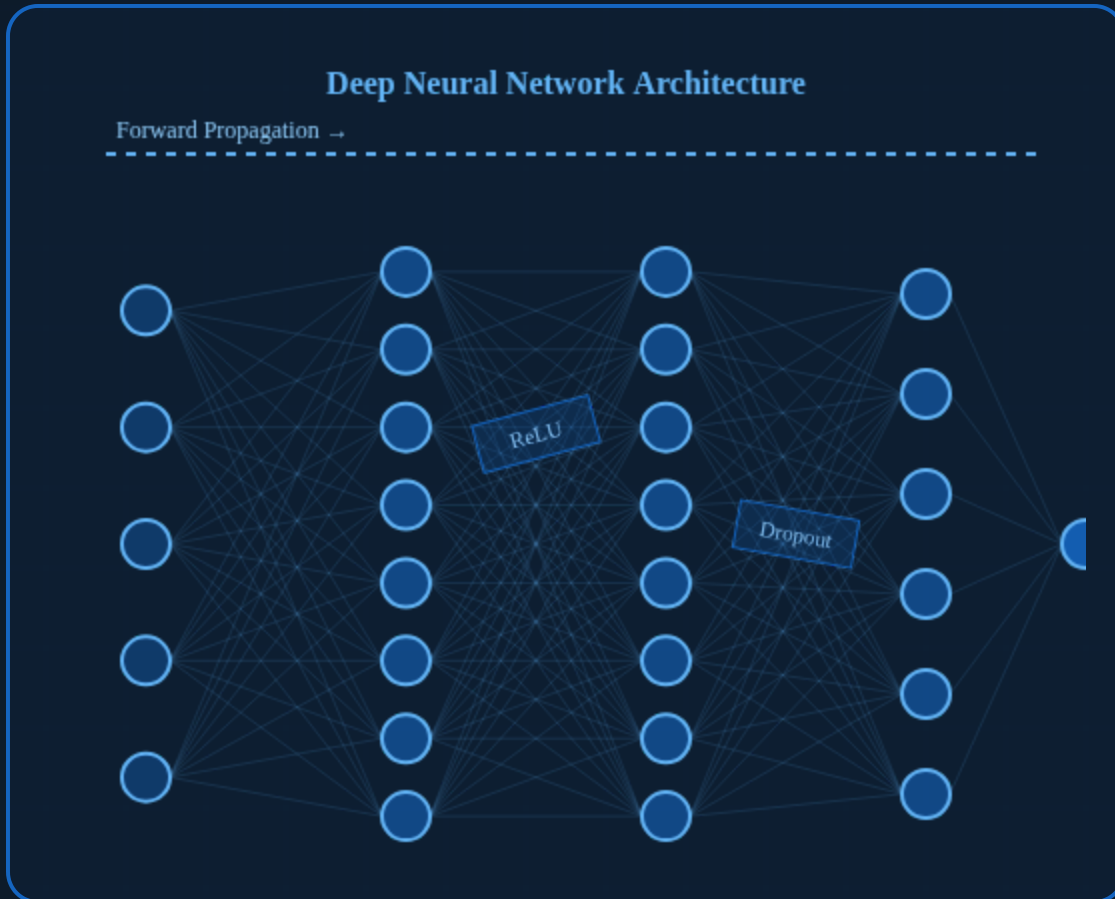
포인트

심리 상태의 비선형성(딱 떨어지지 않는 모호함)을 가장 유연하게 반영하여 모델의 안정성을 확보

Standard DNN

10%

데이터의 전반적인 베이스라인을 확립하는 표준 신경망



작동원리

복잡한 기교 없이 데이터의 가장 기본적이고 통계적인 흐름을 읽어냄



핵심기술

Multi-Layer Perceptron (MLP) 가장 표준적인 신경망 구조로 데이터의 전체적인 통계적 특성을 학습



포인트

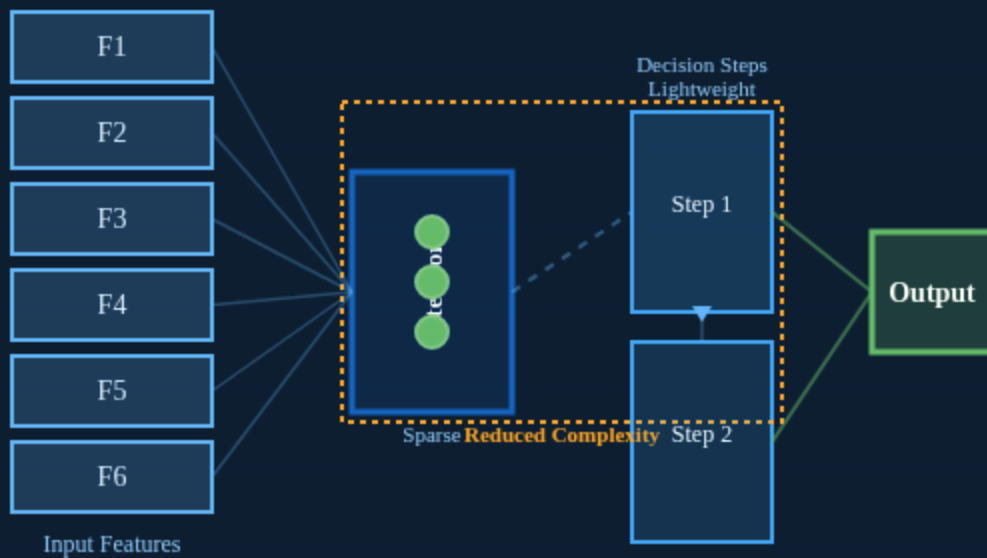
다른 모델들이 너무 복잡한 패턴에 매몰되지 않도록 중심 (Baseline)을 잡아주는 기준점 역할

8%

TabNet Lite

경량화된 TabNet으로 과적합 방지와 일반화 성능 향상

TabNet Lite Architecture



작동원리

수많은 정보 중 꼭 필요한 정보만 골라내는 '요약 달인'

핵심기술

경량화된 TabNet 구조로 연산량을 줄여 핵심적인 의사결정 경로만 추적

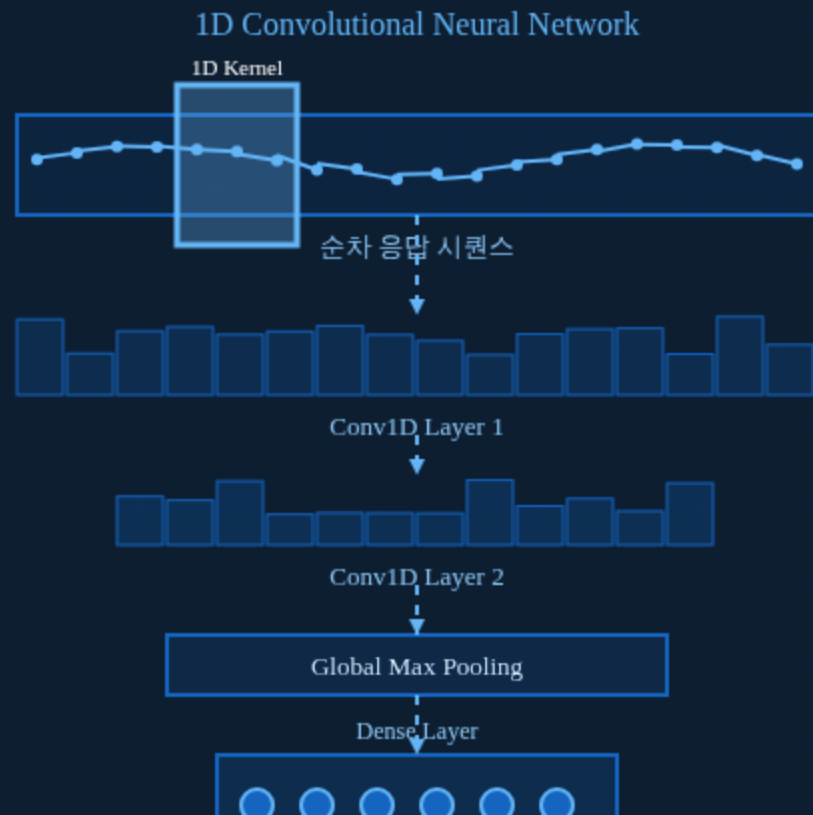
포인트

일반화 성능 향상 및 효율적 학습을 통해 모델이 불필요한 노이즈를 배우지 않게 방어하여 과적합 방지

1D-CNN

7%

Sequence Tracker 설문 응답의 순차적 흐름 속에서 미세한 심리적 변화를 감지



작동원리

설문 답변을 하나씩 보는 게 아니라, 앞뒤 답변의 연결 흐름 (리듬)을 읽음

핵심기술

Temporal Feature Mapping 질문 답변의 순서(시퀀스)에서 나타나는 특징을 시간적 흐름으로 파악

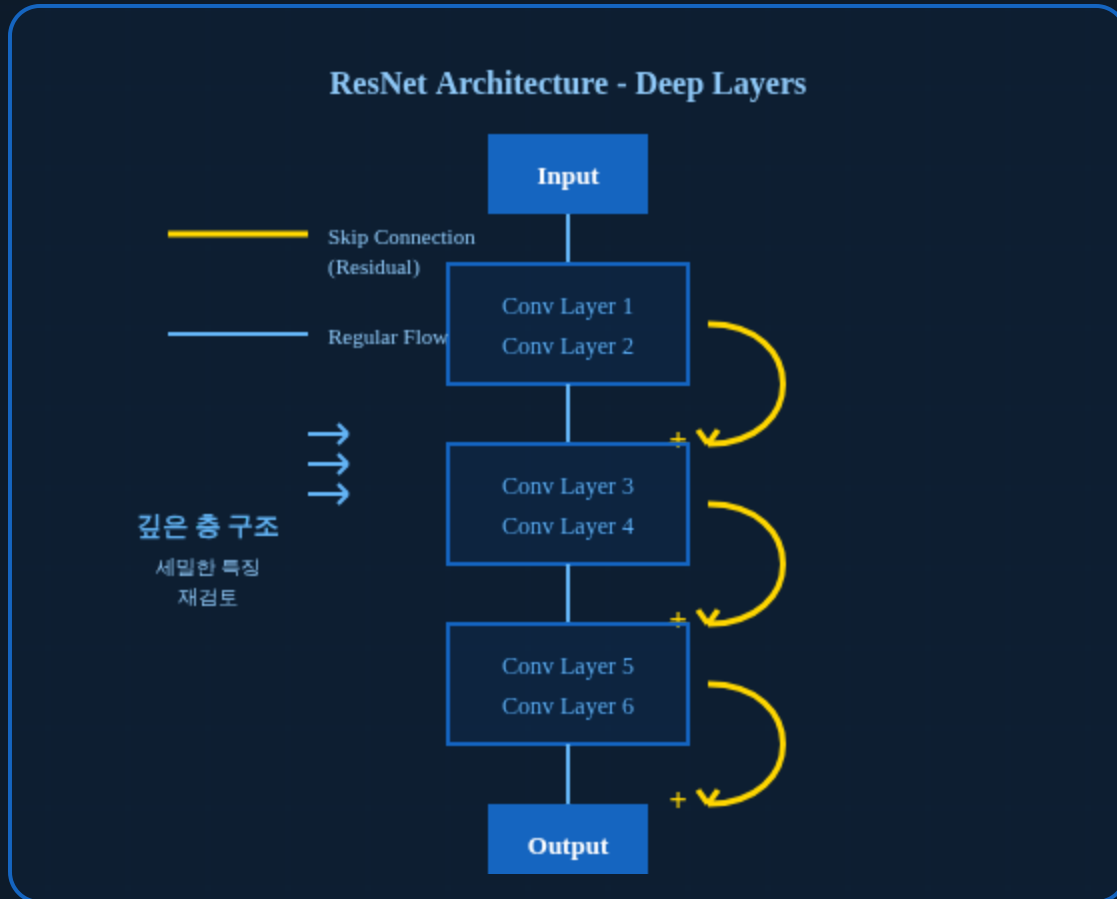
포인트

문항 간의 순차적인 연관성을 파악하여, 답변의 흐름 속에 숨겨진 심리적 일관성을 포착

ResNet

5%

층을 깊게 쌓아 모델이 놓칠 수 있는 아주 세밀한 데이터 특징 재검토



작동원리

한 번 배운 정보도 '지름길'을 통해 다시 확인하며 오차를 수정

핵심기술

잔차 연결로 층이 깊어져도 학습이 잘 되도록 잔차(오차)만을 따로 학습하여 정밀도를 높임

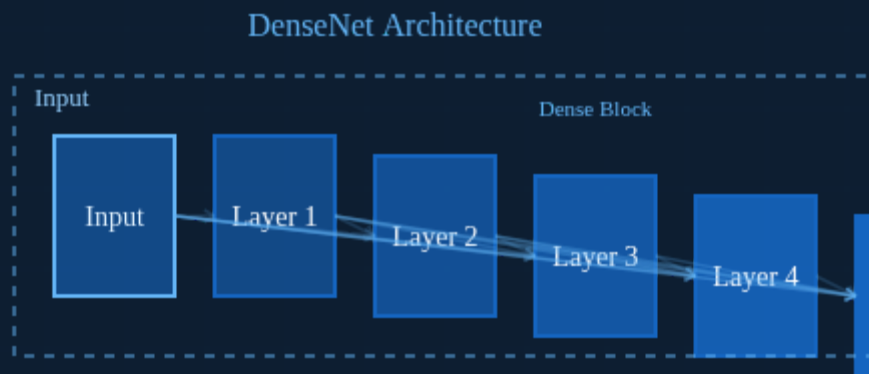
포인트

층을 깊게 쌓아도 정보가 왜곡되지 않게 하여, 아주 미세한 데이터의 특징까지 놓치지 않고 검토

DenseNet

5%

모든 층의 정보를 촘촘하게 연결하여 정보 손실 최소화



각 층은 이전의 모든 층으로부터 입력을 받아
정보 손실을 최소화하고 특징을 재사용합니다

작동원리

정보를 실시간으로 공유하듯, 앞 층의 모든 정보와 지식을
뒤 층으로 빠짐없이 전달

핵심기술

Feature Re-Use 앞 층의 정보를 뒤 층에 직접 연결하여
데이터 손실 없이 끝까지 활용

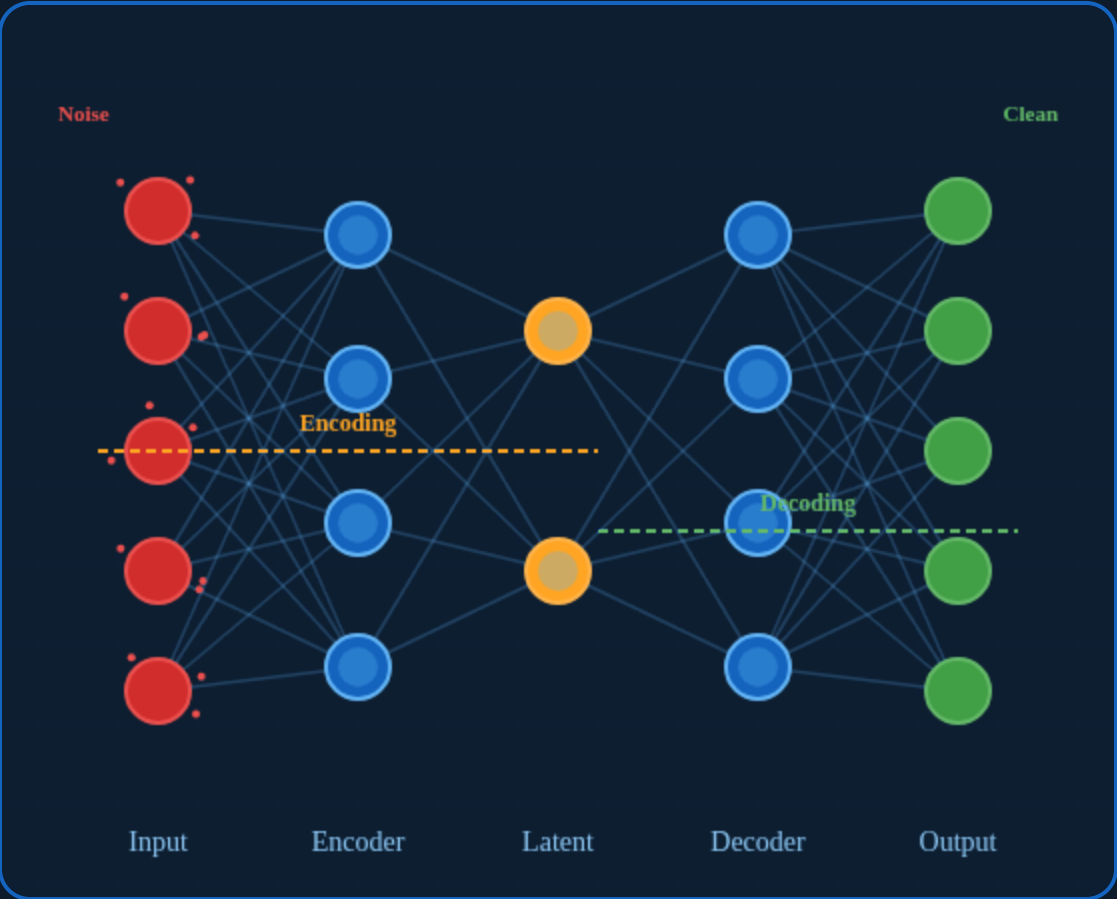
포인트

정보 손실을 원천 차단하여, 단 하나의 데이터 단서도 끝까지
결론에 반영하도록 설계

5%

DAE (Denoising AutoEncoder)

불완전 데이터 복원과 본질 학습



작동원리

일부러 훼손된 데이터 속에서도 원본의 의도를 추론하게 학습하여, 데이터의 '본질적인 핵심'을 꿰뚫어 봅니다.

핵심기술

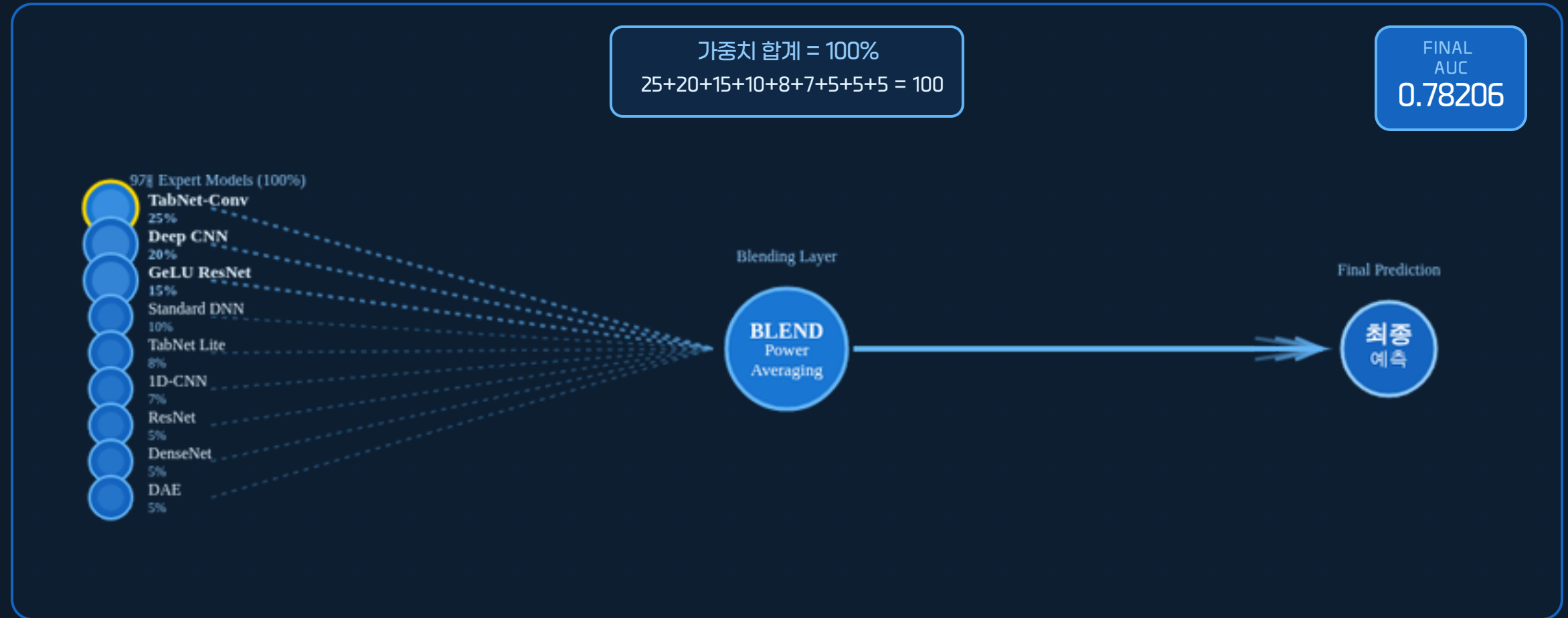
데이터에 일부러 노이즈를 넣어 본질적인 데이터의 형태를 스스로 복원하게 만듦

포인트

이상치(Outlier)나 무성의한 답변 속에서도 흔들리지 않는 모델의 '맷집'을 키우기 위해 배치

Blending — 집단지성의 완성

"9가지 시각이 하나로 모일 때, 예측은 확신이 된다"



Power Averaging

가중치 투표 방식으로 예측값에 지수(power)를 취해 점수의 변별력을 더 높임



Residual Calibration

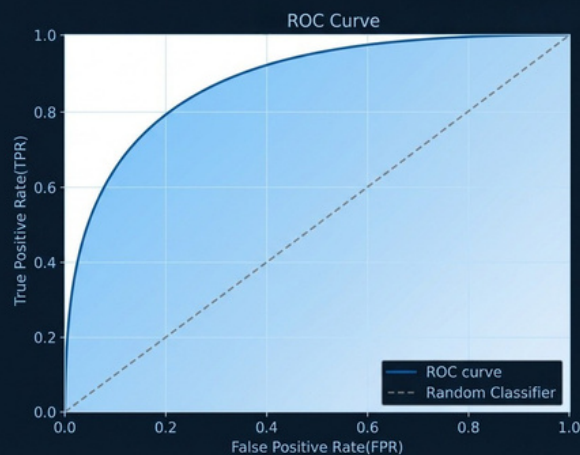
마지막 0.1%를 잡는 정밀 튜닝 기법으로 최종 오차 미세 보정

성능 검증 - 알고리즘 평가 지표 및 검증 전략

평가 지표 계산 및 모델 신뢰도 검증

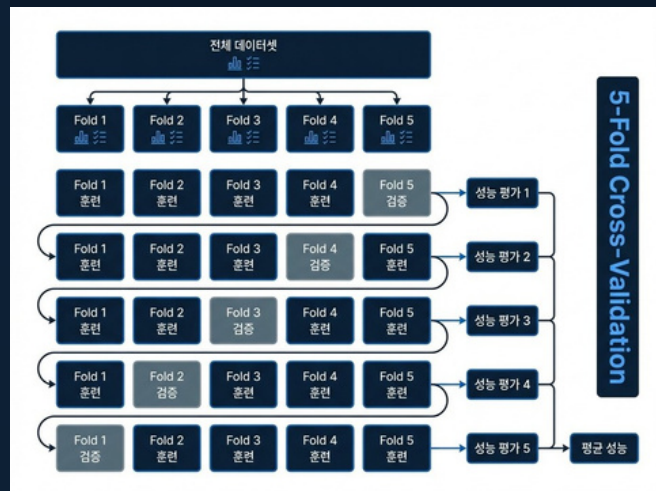
"다각도 검증을 통해 도출된 높은 변별력과 예측 안정성"

메인 지표: AUC-ROC



모델의 분류 변별력을 측정하는 핵심 지표. 리더보드 상위권 수준의 AUC 달성

검증 전략: K-Fold



Stratified 5-Fold 교차 검증으로
과적합 방지 및 일반화 성능 극대화

최종 성적

점수

0.7820682404

단일 모델 대비 블랜딩 후
AUC 향상

결론 및 인사이트

"데이터로 읽는 사람의 마음"

각기 다른 결정 경계(Decision Boundary)를 가진 9개 모델의 다양성을 확보
전략적 가중치 블렌딩을 통해 개별 모델의 편향을 상쇄
모델의 예측 성능과 AUC 점수의 향상

1. 지능형 여론 조사

핵심: 단순한 "누굴 뽑겠습니까?"라는 질문 없이도, 몇 가지 심리 문항 답변만으로 숨은 표심(사익 유권자)을 정교하게 예측 가능

활용: 여론 조사 비용 절감 및 결과의 정확도 개선. 특정 후보에 대한 거부감이 강한 층의 심리적 원인을 분석하여 맞춤형 선거 전략 수립에 활용 가능

2. 라이프스타일 기반 '인지 건강' 스크리닝

핵심: 투표와 같은 복잡한 사회적 판단은 뇌의 고위험 인지 기능을 사용하기에 답변의 일관성과 논리적 흐름이 무너지는 패턴을 추적

활용: 여론조사나 심리 테스트를 수행하는 과정에서 사용자의 인지 반응 속도와 논리적 일관성을 체크하여, '인지 기능 저하(경도인지장애) 징후'를 사전에 감지하는 스크리닝 보조 지표로 확장

3. 다각적 진단 데이터 통합

핵심: 설문 답변 과정에서 나타나는 논리적 일관성과 패턴의 변동성을 추적하면, 단순한 임상 수치를 넘어 뇌의 인지적 건강 상태를 정량화

활용: 임상 수치뿐만 아니라 환자의 심리 상태와 행동 패턴을 통합하여 질병을 진단하는 정밀 의료(Precision Medicine) 전략

사리웃!



김민찬 오전 1:46

목이 진짜 개아픔 진짜 레전드 고통 발생 이거 게비스콘같은거 강 부어버리고싶음



1



윤희주 오전 1:47

근데 자야되긴함
환자라,,



김민찬 오전 1:47

6시에 먹어서 지금 약효 꺼지고있음



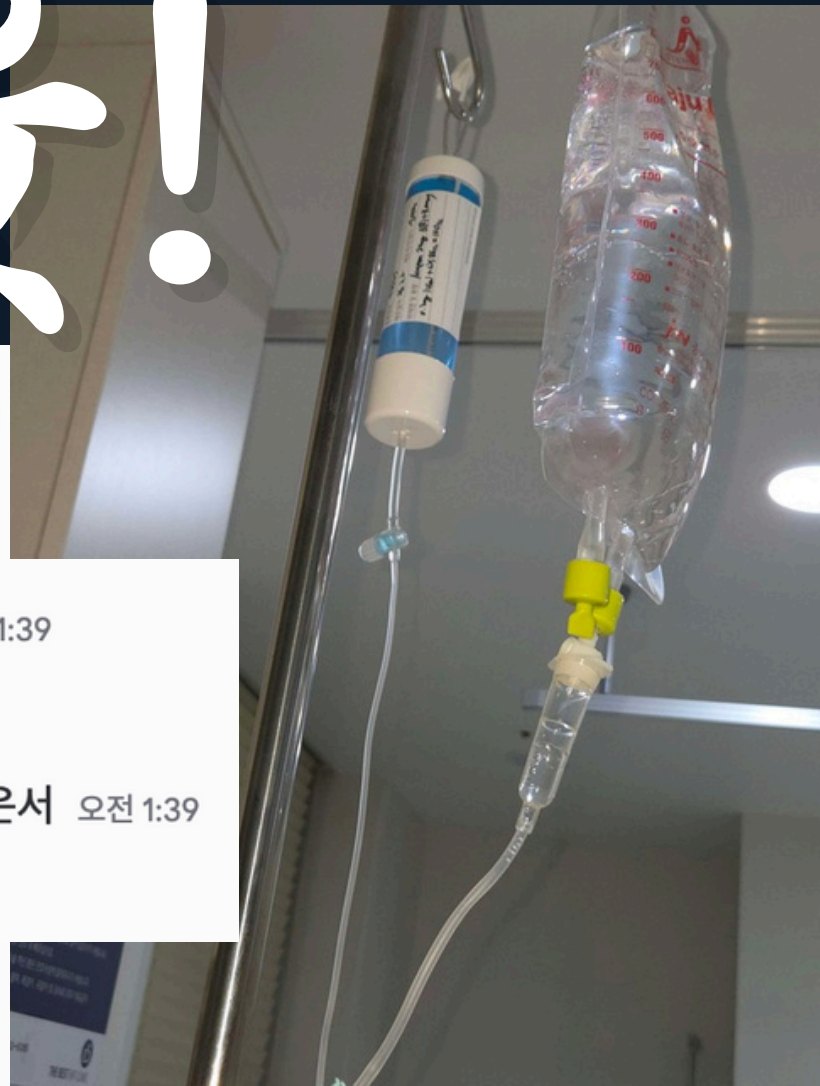
김민찬 오전 1:39

부상투혼



AH_01_양은서 오전 1:39

링거투혼



KINGSMEN

