

2024년도 공공기관 용역과제  
AI개발 수행내역서

과제명	AI기반 학생 우울증 예측모델 개발 및 시각화
담당자	박원빈

2025년 01 월 07일

AI개발 수행내용

1. 사업과제 : AI 기반 학생 우울증 예측모델 개발 및 시각화

2. 개요 및 현황

2.1 추진배경 및 목적

- 최근 정신건강 문제가 주요 사회적 이슈로 부각되며, 학생들의 심리적 상태를 조기에 파악하고 개입하는 시스템의 필요성이 증가하고 있음
- 기존 상담 및 설문 기반 접근법은 정확도와 실시간 모니터링 측면에서 한계가 존재
- 머신러닝 기술을 활용하여 주요 변수(성별, 수면 시간, 식습관 등)를 기반으로 학생들의 우울증 여부를 예측하는 모델을 개발하고, 이를 통해 조기 개입 및 지원 체계를 강화
- 우울증 관련 데이터를 분석하여 정확하고 신뢰성 있는 예측모델을 개발.
- 주요 특성을 시각화하여 데이터에 대한 통찰 제공
- 시스템을 통해 우울증 발생 가능성이 높은 학생들을 조기에 식별하여 필요한 지원 제공.

2.2 과제 범위

과제구분		내용
AI	AI기반 우울증 예측모델 구현	설문 데이터 수집 및 데이터셋 구축
		데이터 전처리 (결측치 처리, 원-핫 인코딩 등)
		랜덤 포레스트 기반 예측 모델 개발
		성능 평가(정확도, MSE, RMSE, MAE 등)
		주요 Feature 중요도 분석 및 시각화
		실제값 vs 예측값 비교 시각화
시각화	실시간 우울증 예측 데이터 시각화	주요 Feature의 분포 시각화(성별, 수면 시간, 식습관, 재정적 스트레스, 가족의 정신질환 여부 등)
		학습 곡선(Learning Curve) 시각화
		예측값과 실제값 비교를 통한 성능 분석

2.3 과제 추진 방법

1) 구축 대상 선정 기준

- 데이터 접근성 및 활용성
  - 설문 데이터를 기반으로 우울증 관련 변수를 수집하고 머신러닝 모델 학습에 적합하도록 전처리
  - 데이터셋 내 결측치 처리 및 주요 변수의 변환(원-핫 인코딩) 수행
- 예측모델 개발 효율성
  - 주요 변수(성별, 수면 시간, 식습관, 재정적 스트레스, 가족의 정신질환 여부 등)을 활용하여 모델 복잡도를 낮추고, 학습 및 평가를 효율적으로 진행.
- 환경문제 해결 기여도 및 경제성
  - 조기 개입을 통해 학생들의 심리적 건강 문제를 사전에 해결하고, 교육 현장에서의 긍정적 효과를 기대

2) AI 예측 분석모델 적용 대상

학생 데이터	수집 데이터	예측모델 입력 변수	AI예측 분석 대상
우울증 관련 설문 데이터	- 성별, 수면시간, 식습관, 재정적 스트레스, 가족의 정신질환 여부, 자살 생각 여부 등	- 성별, 수면시간, 식습관, 재정적 스트레스, 가족의 정신질환 여부, 자살 생각 여부	- 학생들의 우울증 발생 가능성 예측

3) AI 분석모델 구축 프로세스

## 연구개발 주요 결과물

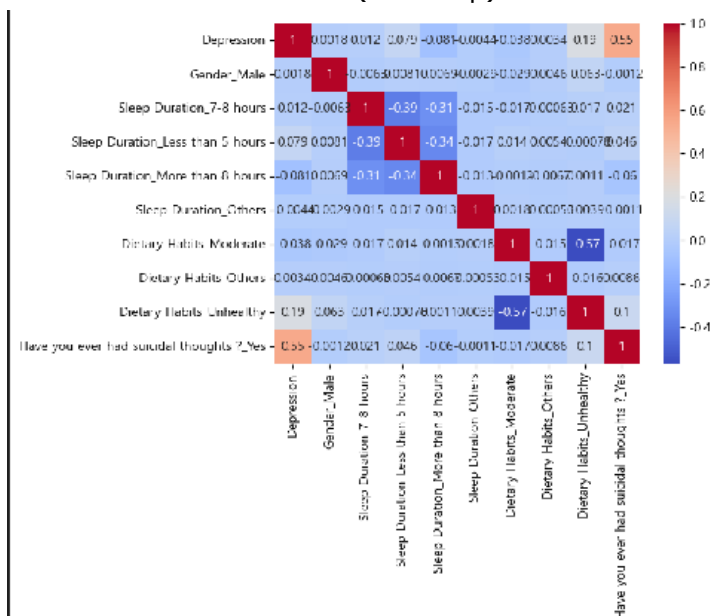
### 1. 데이터 수집

- KAGGLE STUDENT DEPRESSION DATASET (학생 우울증 설문 데이터)(엑셀) :

id	Gender	Age	City	Profession	Academic	Work Pres	CGPA	Study Sati	Job Satisf	Sleep Dur.	Dietary Hz	Degree	Have you ever	Work/Stu	Financial	Family His	Depression
2	Male	33	Visakhapatnam	Student	5	0	6.97	2	0	5-6 hours	Healthy	B.Pharm	Yes	3	1	No	1
0	Female	24	Bangalore	Student	2	0	5.9	5	0	5-6 hours	Moderate	B.Sc	No	3	2	Yes	0
26	Male	31	Srinagar	Student	3	0	7.03	5	0	Less than 5 hours	Healthy	BA	No	9	1	Yes	0
30	Female	28	Varanasi	Student	3	0	5.59	2	0	7-8 hours	Moderate	BCA	Yes	4	5	Yes	1
32	Female	25	Jalpur	Student	4	0	8.12	3	0	5-6 hours	Moderate	M.Tech	Yes	1	1	No	0
33	Male	29	Pune	Student	2	0	5.7	3	0	Less than 5 hours	Healthy	PhD	No	4	1	No	0
52	Male	30	Thane	Student	3	0	9.54	4	0	7-8 hours	Healthy	B.Sc	No	1	2	No	0
56	Female	30	Chennai	Student	2	0	8.04	4	0	Less than 5 hours	Unhealthy	Class 12	No	0	1	Yes	0
59	Male	28	Nagpur	Student	3	0	9.79	1	0	7-8 hours	Moderate	B.Ed	Yes	12	3	No	1
62	Male	31	Nashik	Student	2	0	8.38	3	0	Less than 5 hours	Moderate	LLB	Yes	2	5	No	1
83	Male	24	Nagpur	Student	3	0	6.1	3	0	5-6 hours	Moderate	Class 12	Yes	11	1	Yes	1
91	Male	33	Vadodara	Student	3	0	7.03	4	0	Less than 5 hours	Healthy	BF	Yes	10	2	Yes	0
94	Male	27	Kalyan	Student	5	0	7.01	1	0	Less than 5 hours	Moderate	M.Tech	No	10	1	Yes	1
100	Female	19	Rajkot	Student	2	0	6.52	4	0	Less than 5 hours	Unhealthy	Class 12	No	6	2	Yes	0
103	Female	19	Kalyan	Student	5	0	5.64	5	0	Less than 5 hours	Moderate	Class 12	Yes	4	5	Yes	1
106	Male	29	Srinagar	Student	3	0	6.58	3	0	More than 8 hours	Moderate	M.Tech	Yes	10	2	Yes	1
120	Male	25	Nashik	Student	5	0	6.51	2	0	Less than 5 hours	Unhealthy	M.Ed	Yes	2	5	Yes	1
132	Female	20	Ahmedabad	Student	5	0	7.25	3	0	5-6 hours	Healthy	Class 12	Yes	10	3	No	1
139	Male	19	Chennai	Student	2	0	7.82	2	0	7-8 hours	Unhealthy	Class 12	No	6	2	No	0
145	Male	25	Kalyan	Student	3	0	9.92	3	0	5-6 hours	Moderate	B.Ed	No	8	2	Yes	1
161	Male	29	Kolkata	Student	3	0	8.74	4	0	5-6 hours	Moderate	B.Ed	Yes	1	1	No	0
162	Male	29	Kolkata	Student	3	0	6.73	3	0	7-8 hours	Moderate	M.Tech	No	0	1	No	0
166	Female	25	Ahmedabad	Student	3	0	5.57	3	0	More than 8 hours	Unhealthy	M.Sc	Yes	10	5	No	1
172	Male	23	Thane	Student	1	0	8.08	4	0	7-8 hours	Healthy	BHM	No	11	3	No	0
173	Male	18	Bangalore	Student	4	0	7.1	3	0	More than 8 hours	Unhealthy	Class 12	Yes	11	5	Yes	1
176	Female	20	Mumbai	Student	5	0	8.58	5	0	7-8 hours	Moderate	Class 12	No	2	2	Yes	1
186	Male	31	Ahmedabad	Student	2	0	6.06	5	0	7-8 hours	Moderate	LLB	Yes	3	3	Yes	1
193	Male	25	Lucknow	Student	3	0	7.25	3	0	More than 8 hours	Unhealthy	M.Ed	Yes	10	5	No	1
200	Male	33	Indore	Student	5	0	5.74	2	0	Less than 5 hours	Moderate	M.Pharm	No	0	3	Yes	0
214	Male	28	Kalyan	Student	3	0	9.06	3	0	7-8 hours	Unhealthy	M.Pharm	Yes	11	2	No	1
222	Male	18	Surat	Student	4	0	6.7	5	0	Less than 5 hours	Moderate	Class 12	Yes	5	4	Yes	1
232	Male	18	Visakhapatnam	Student	2	0	6.21	3	0	5-6 hours	Unhealthy	Class 12	Yes	4	2	No	1
239	Male	21	Jalpur	Student	1	0	7.25	1	0	Less than 5 hours	Healthy	MCA	Yes	7	2	No	0

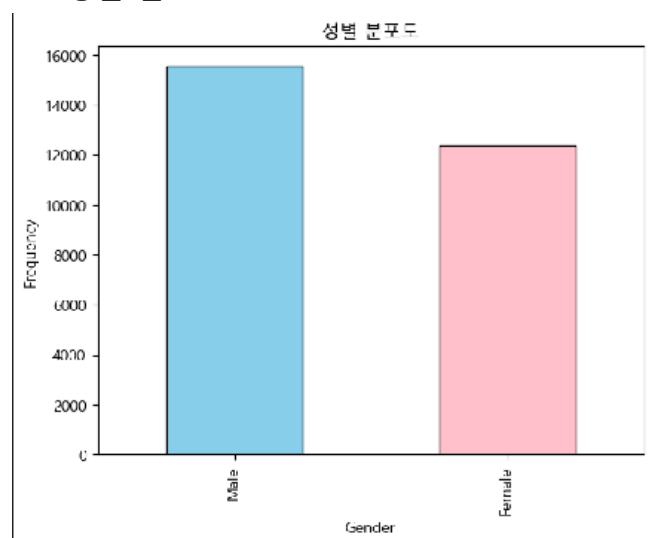
### 2. 데이터 분석

#### 2.1 변수 간 상관관계 시각화 (Heatmap)

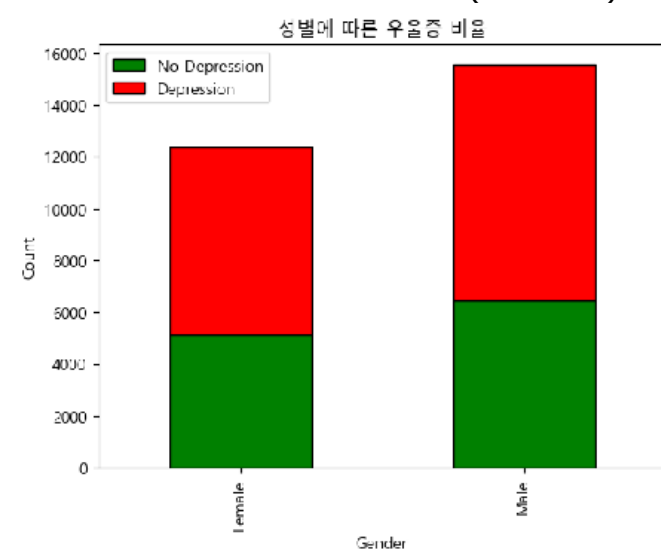


자살 생각 여부(Have you ever had suicidal thoughts?)가 0.55로 자살 생각 여부가 우울증 발생에 매우 큰 영향을 미치는 중요한 변수임을 의미합니다.

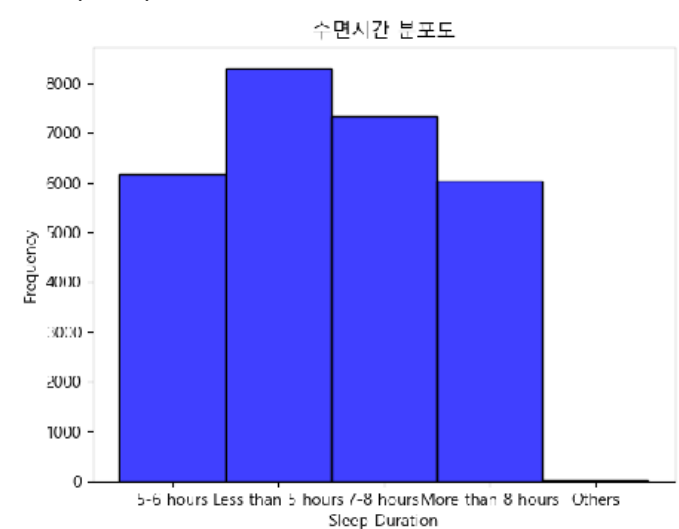
## 2.2 성별 분포도



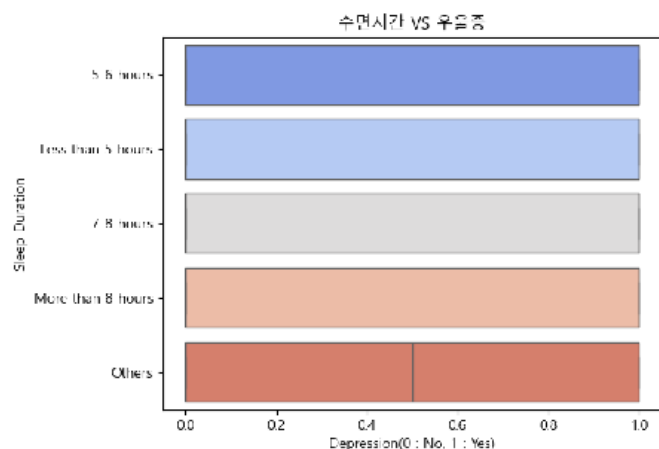
## 2.3 우울증 여부와 성별 간의 관계 (비율 비교)



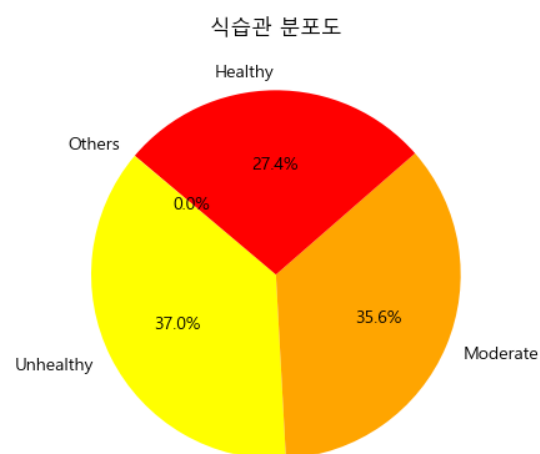
## 2.4 수면시간 분포도



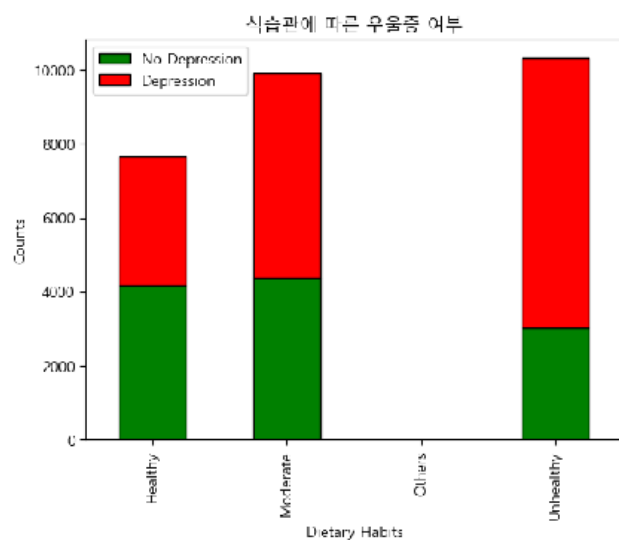
## 2.5 수면 시간에 따른 우울증 여부 시각화



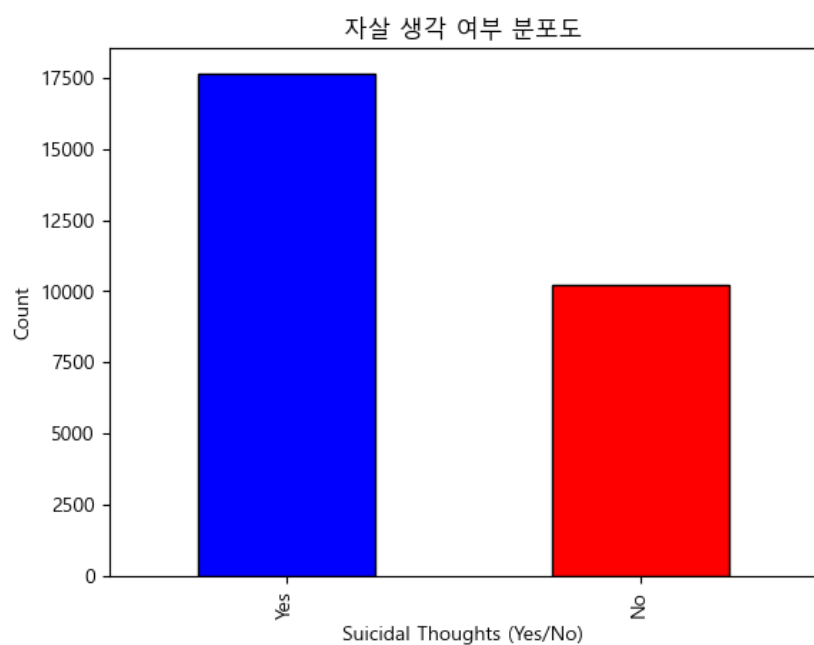
## 2.5 식습관 비율



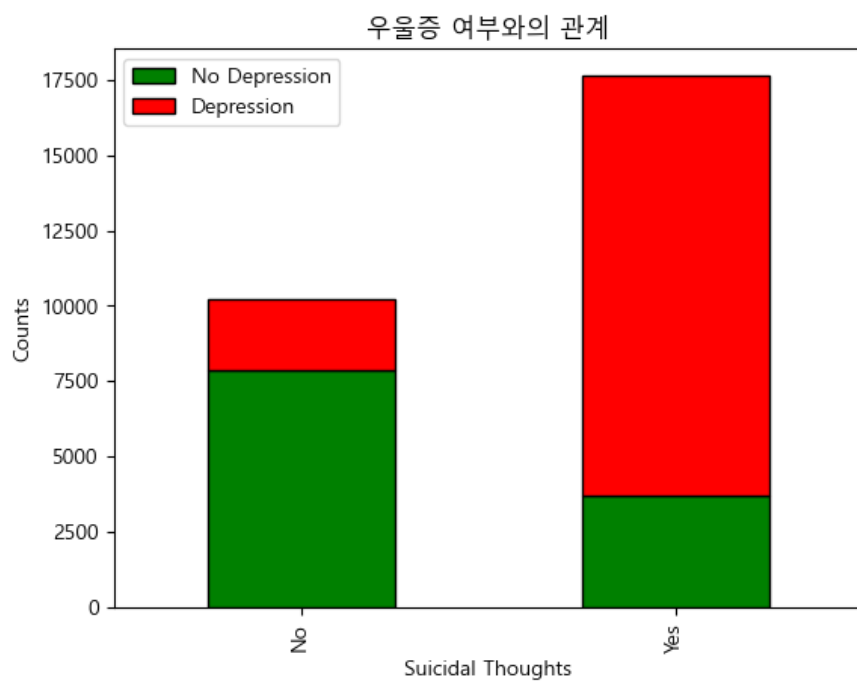
## 2.6 식습관에 따른 우울증 여부



## 2.7 자살 생각 여부



## 2.8 우울증 여부와의 관계



## 1. EDA(탐색적 데이터 분석)

### ○ 히스토그램 및 히트맵:

- 주요 변수(성별, 수면 시간, 식습관, 자살 생각, 재정적 스트레스, 가족들의 정신질환 여부)의 분포를 확인하기 위해 히스토그램을 작성.
- 히트맵을 통해 변수 간의 상관관계를 분석하여 모델 학습에 적합한 변수 선정.
- 데이터가 정규분포를 따르지 않는 경우 로그 변환 등의 데이터 변환 필요성 검토

## 2. 우울증 예측 모델링을 위한 대상 설정

### ○ 대상 변수:

- 예측 목표: 학생들의 우울증 여부(Depression)를 0(비우울) 또는 1(우울)로 분류.

### ○ 독립 변수(입력 변수):

- 성별(Gender)
- 수면 시간(Sleep Duration)
- 식습관(Dietary Habits)
- 재정적 스트레스(Financial Stress)
- 가족들의 정신질환 여부(Family Members with Mental Illness)
- 자살 생각 여부(Have you ever had suicidal thoughts?)

## 3. 주요 변수에 영향을 미치는 요인 분석

### ■ 심리적 변수:

- 자살 생각 여부(Yes/No)는 우울증 예측에 가장 중요한 변수로 확인.
- 수면 시간이 적을수록 우울증 발생 확률이 증가하는 경향 관찰.
- 재정적 스트레스가 많을 수록 우울증 발생 확률이 증가하는 경향 관찰

### ■ 생활습관 변수:

- 식습관(Healthy, Moderate, Unhealthy)은 우울증 발생 확률과 약한 상관관계를 보임.
- 성별(Gender)에 따라 우울증 발생 빈도의 차이가 존재.

### ■ 환경적 변수(추후 확장 가능):

- 학업 스트레스, 학업 만족도, 직업 만족도 등이 우울증 발생에 영향을 미칠 가능성이 있으나, 현재 데이터셋에는 포함되지 않음.

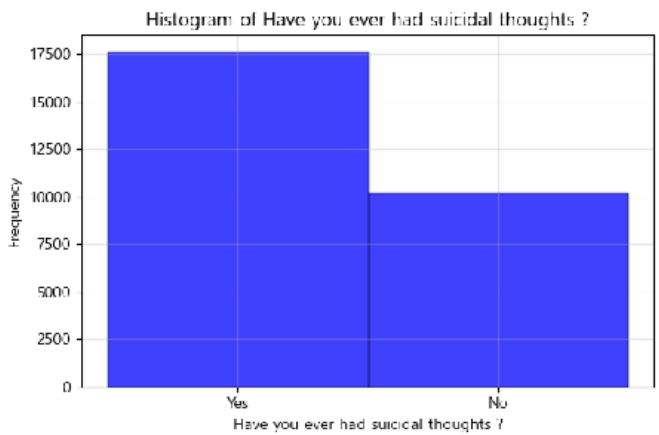
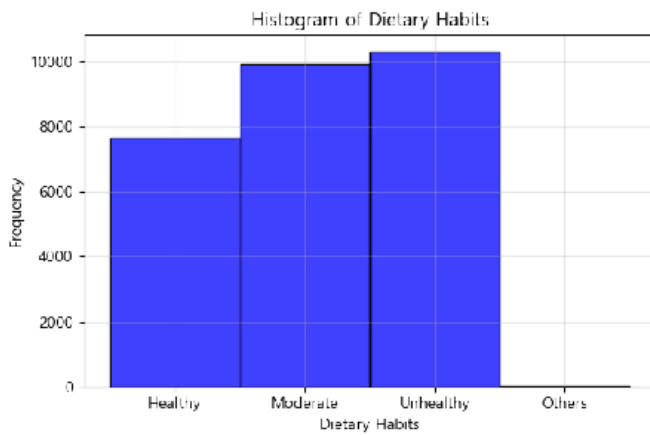
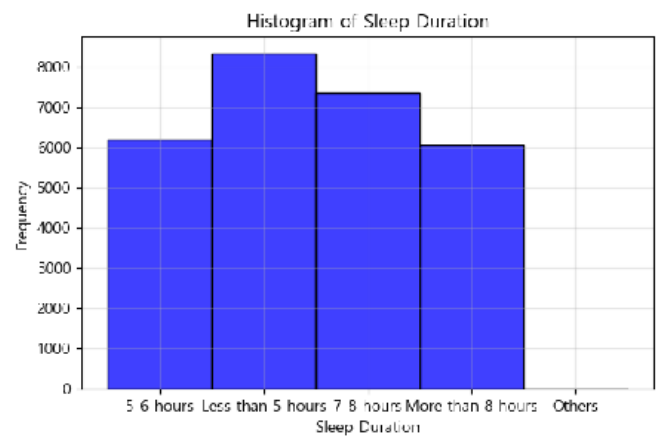
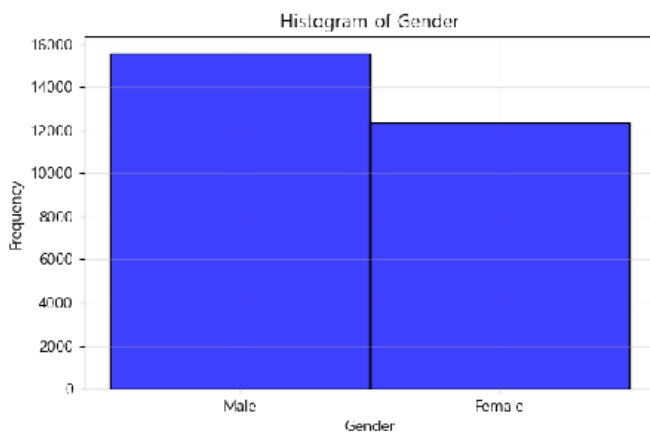
## 2.2 탐색적 데이터 분석

### ○ 결측치 및 중복값 통계

컬럼명	결측치	중복값
id	0	0.00
gender	0	0.00
age	0	0.00
city	0	0.00
education	0	0.00
financial_stress	0	0.00
family_members_with_mental_illness	0	0.00
have_you_ever_had_suicidal_thoughts?	0	0.00
sleep_duration	0	0.00
study_satisfaction	0	0.00
work_satisfaction	0	0.00
depression	0	0.00
number_of_duplicate_rows	0	0.00



○ 주요 변수별 데이터 분포(Histogram)



### ○ 데이터 전처리

원-핫 인코딩을 통해 범주형 데이터를 숫자형으로 변환.

열 정렬 및 기본값 추가: 학습 데이터에 없는 열을 추가하고 기본값(0)으로 설정

```
# 입력 데이터 전처리
def transform_user_input(input_data, X_columns):
    #사용자 입력 데이터 원-핫 인코딩
    input_data_encoded = pd.get_dummies(input_data)

    #학습 데이터의 열 순서에 맞춰 재정렬
    for col in X_columns:
        if col not in input_data_encoded:
            input_data_encoded[col] = 0
    input_data_encoded = input_data_encoded[X_columns]

    return input_data_encoded

input_data_encoded = transform_user_input(input_data, X.columns)
```

## 3. 데이터 학습 및 모델정의

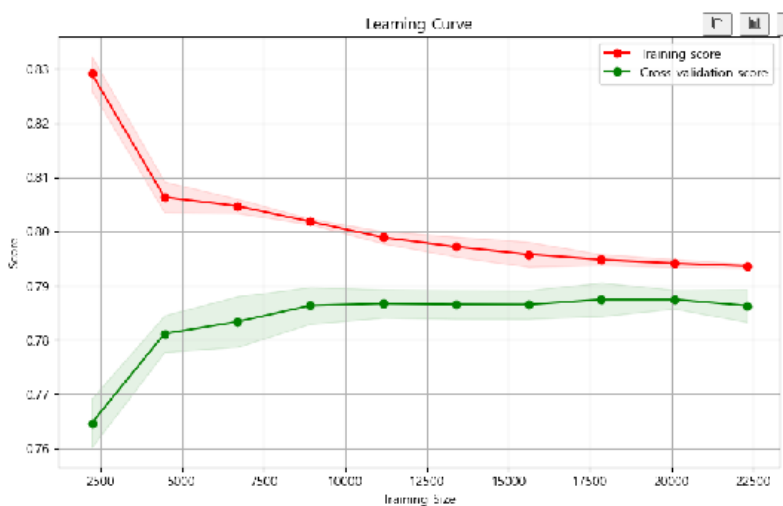
### 3.1 모델정의 및 컴파일

○ 의사결정 트리 모델 정의 : 랜덤포레스트

```
# 랜덤 포레스트 모델 학습
model = RandomForestClassifier(n_estimators=100, max_depth=10, random_state=42)
model.fit(X_train, y_train)
```

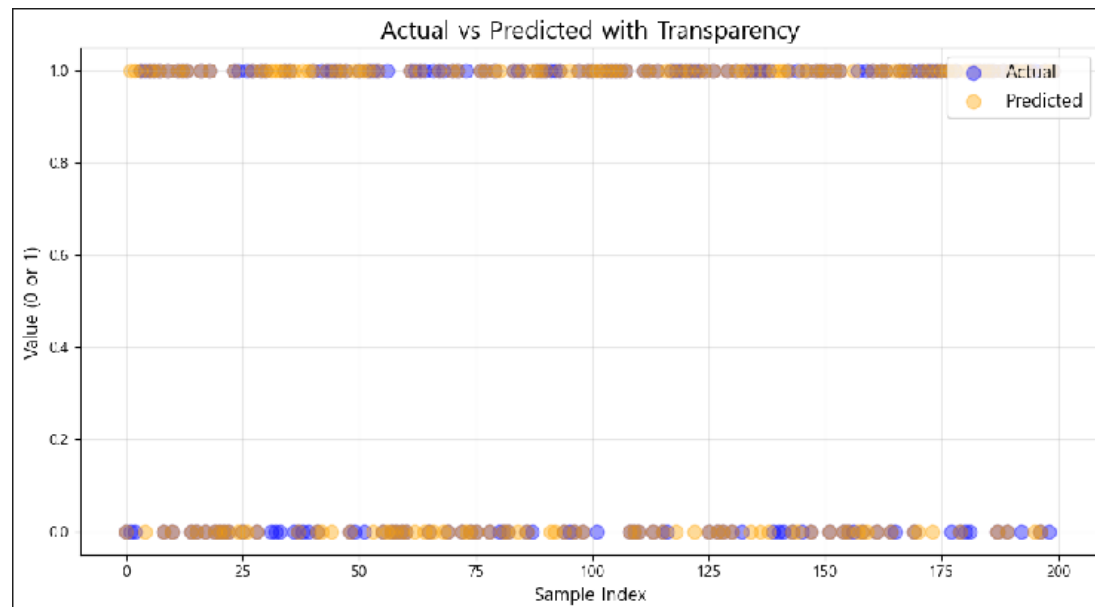
### 3.2 모델학습 및 학습 시각화

○ 학습과정 시각화

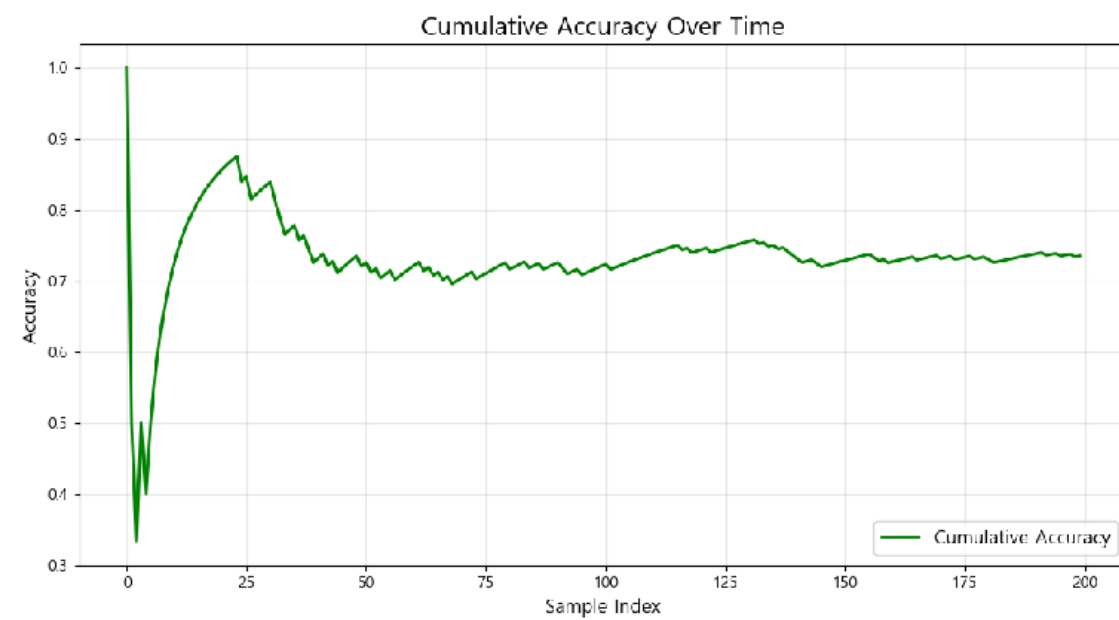


### 3.3 모델 예측

○ 예측값 vs 실제값 비교



### 3.4 누적 정확도



## 4. 프로토타이핑(화면)

### 4.1 모델 예측

#### ○ 학생 우울증 예측

## 학생 우울증 판독 모델

성별 (Gender)

Male

수면 시간 (Sleep Duration)

More than 8 hours

식습관 (Dietary Habits)

Healthy

자살 생각 여부 (Have suicidal thoughts?)

Yes

재정 상태 (Financial Stress)

Good

가족 정신병력 (Family History of Mental Illness)

No

우울증 여부 예측

#### ○ 우울증 여부 예측결과

성별 (Gender)

Male

수면 시간 (Sleep Duration)

More than 8 hours

식습관 (Dietary Habits)

Healthy

자살 생각 여부 (Have suicidal thoughts?)

Yes

재정 상태 (Financial Stress)

Good

가족 정신병력 (Family History of Mental Illness)

No

우울증 여부 예측

우울증이 의심되지 않습니다!

성별 (Gender)

Male

수면 시간 (Sleep Duration)

Less than 5 hours

식습관 (Dietary Habits)

Healthy

자살 생각 여부 (Have suicidal thoughts?)

Yes

재정 상태 (Financial Stress)

Poor

가족 정신병력 (Family History of Mental Illness)

Yes

우울증 여부 예측

우울증이 의심됩니다!