



Mental Health

Prediction for Depression

EDA / Feature Engineering / Machine Learning

Mental Health

Structure of Dataset

*** Data Dictionary** Columns: 20, Row: 140,700 (train.csv)

Target: Depression

Variable

- Name
- Gender
- Age
- City
- Working Professional
- or Student
- Profession
- Academic Pressure
- Work Pressure
- CGPA
- Study Satisfaction
- Job Satisfaction

Definition

- Name
- Gender
- Age in years
- City of residence
- Whether the person is a Professional or a Student
- Profession (Applicable for professionals)
- Level of academic stress
- Level of work stress
- Cumulative Grade Pont Average
- Satisfaction level with studies
- Satisfaction level with job
- Average hour of sleep

Key

- Male, Female
- Professional, Student

• Sleep Duration

Variable	Definition	Key
• Name	• Name	
• Gender	• Gender	• Male, Female
• Age	• Age in years	
• City	• City of residence	
• Working Professional • or Student	• Whether the person is a Professional or a Student	• Professional, Student
• Profession	• Profession (Applicable for professionals)	
• Academic Pressure	• Level of academic stress	
• Work Pressure	• Level of work stress	
• CGPA	• Cumulative Grade Pont Average	
• Study Satisfaction	• Satisfaction level with studies	
• Job Satisfaction	• Satisfaction level with job	
• Sleep Duration	• Average hour of sleep	
• Dietary Habits	• General dietary patterns	
• Degree	• Academic degree pursued or completed	
• Have you ever had suicidal thoughts ?	• Whether the person has had suicidal thoughts	
• Work/Study Hours	• Average hours spent on work or study	
• Financial Stress	• Level of financial stress	
• Family History of Mental Illness	• Family history of mental illness	
• Depression	• Depression status	• 1 = Depressed, 0 = Not

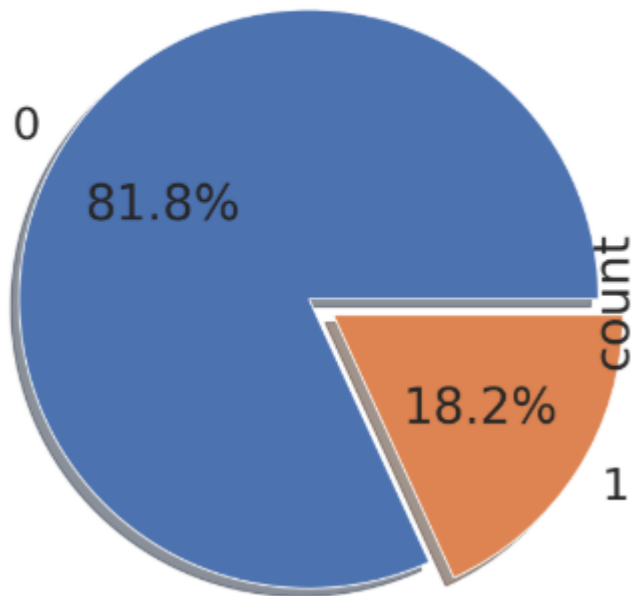
Mental Health

EDA

Depression

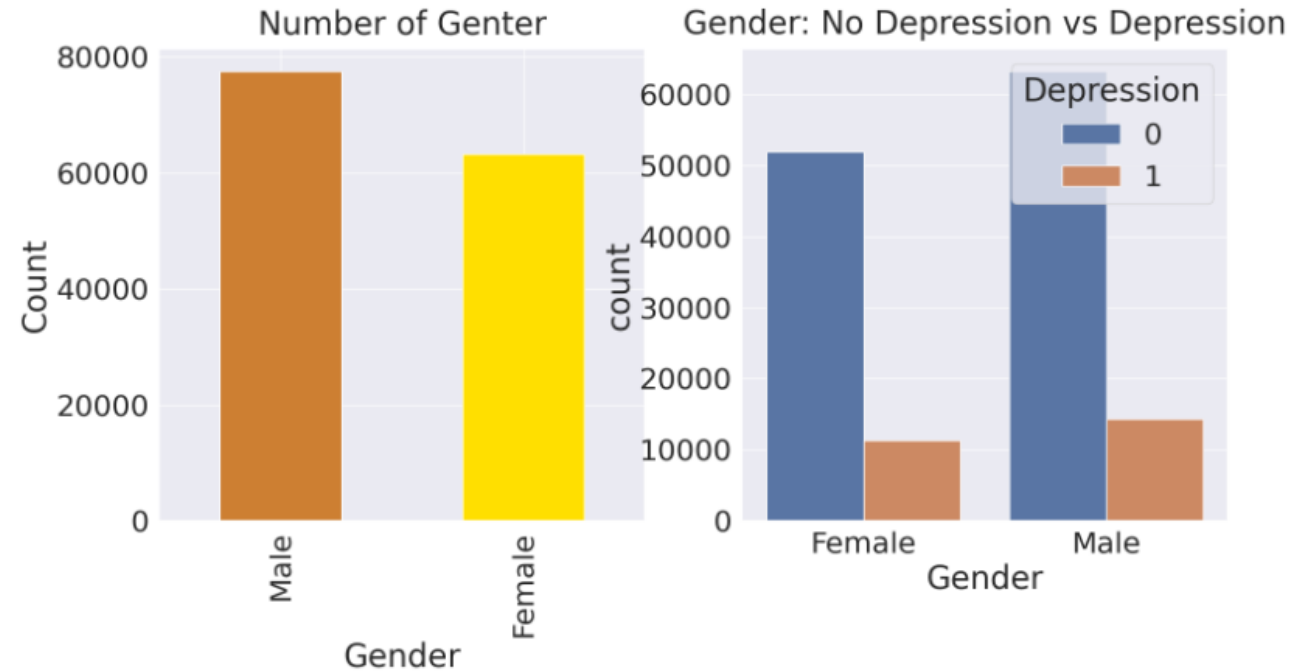
1: Depressed, 0: Not depressed

Pie plot - Depression



Gender with Depression

성별에 따른 Depression 비율

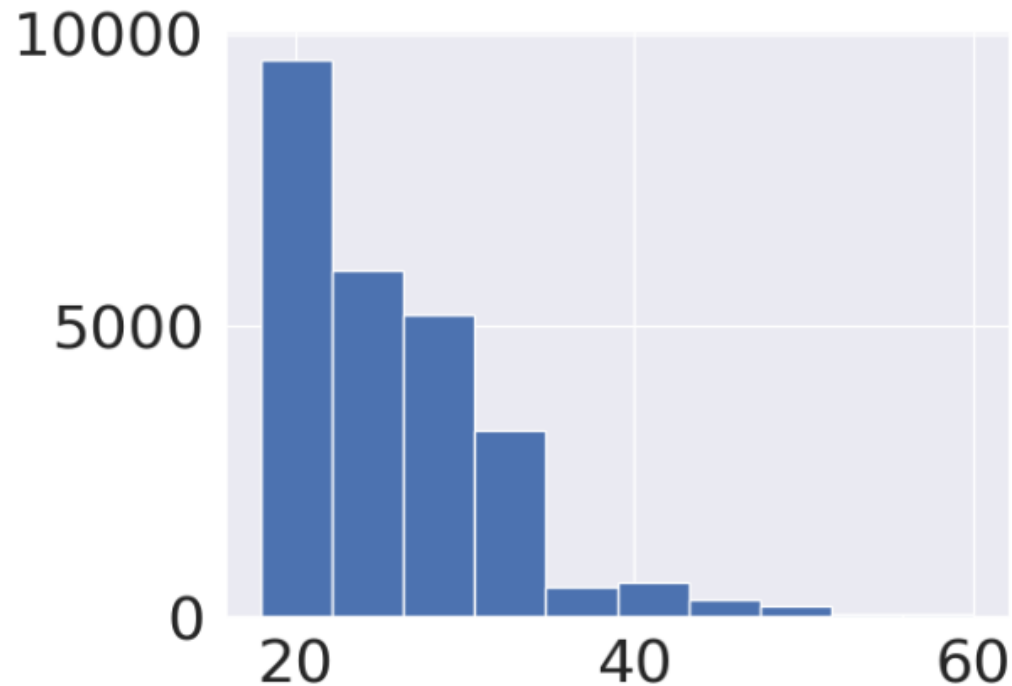


Mental Health

EDA

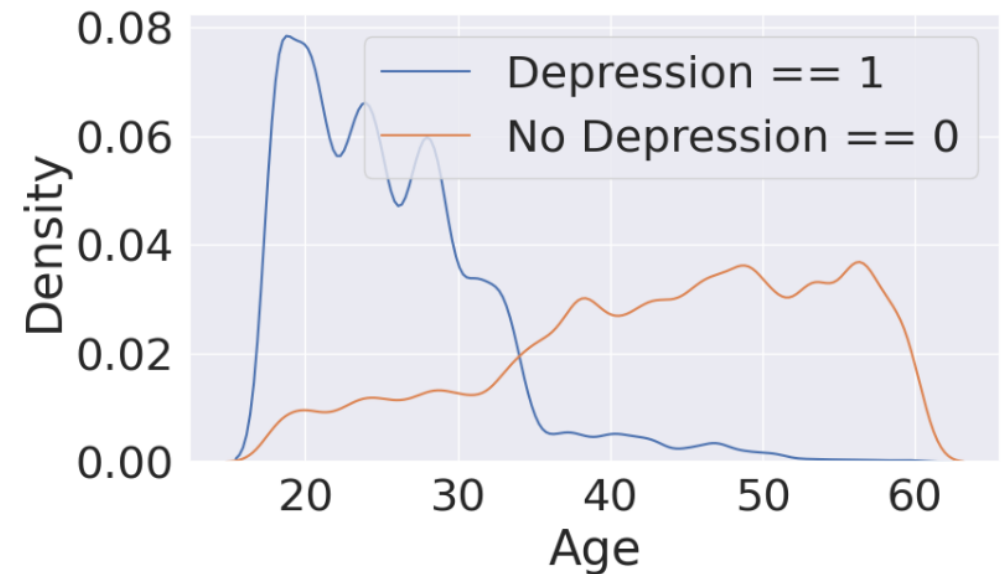
Age with Depression using Boxplot

나이에 대한 Depression 분포 시각화



Age with Depression using KDE

나이에 대한 Depression 분포 시각화 (커널 밀도 추정을 활용함)

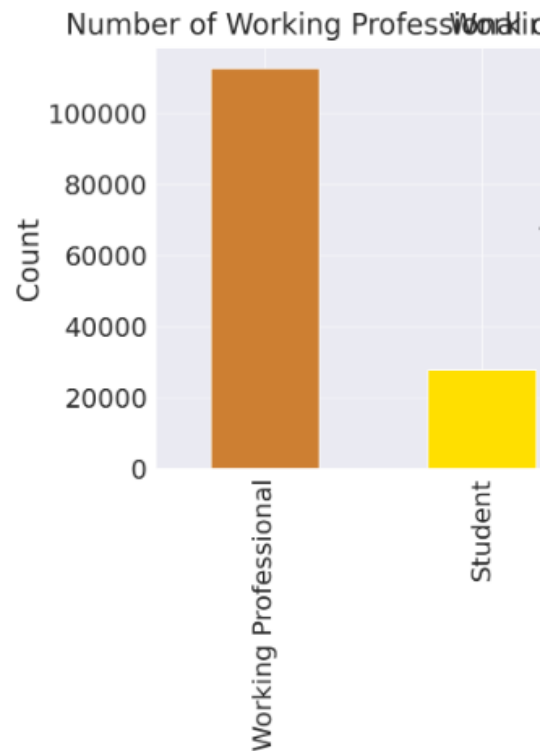


Mental Health

EDA

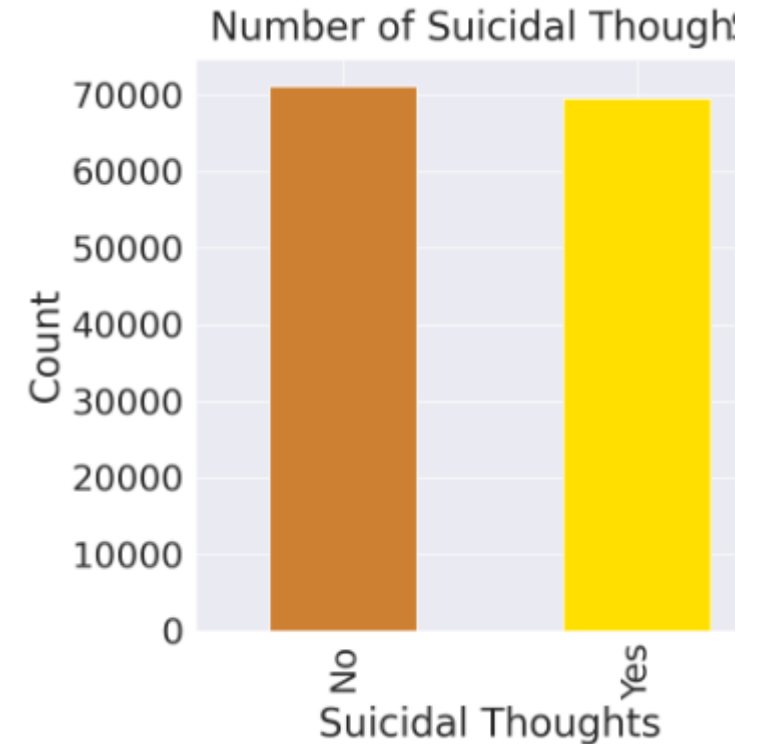
WPOS with Depression

직장인과 학생에 대한 Depression 비율 시각화



Suicidal Thoughts with Depression

자살 생각에 대한 Depression 비율 시각화

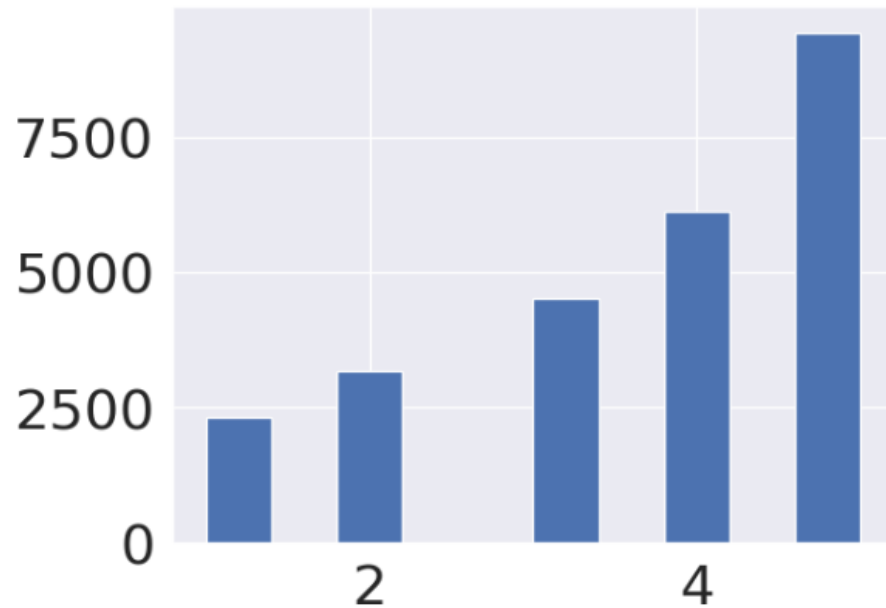


Mental Health

EDA

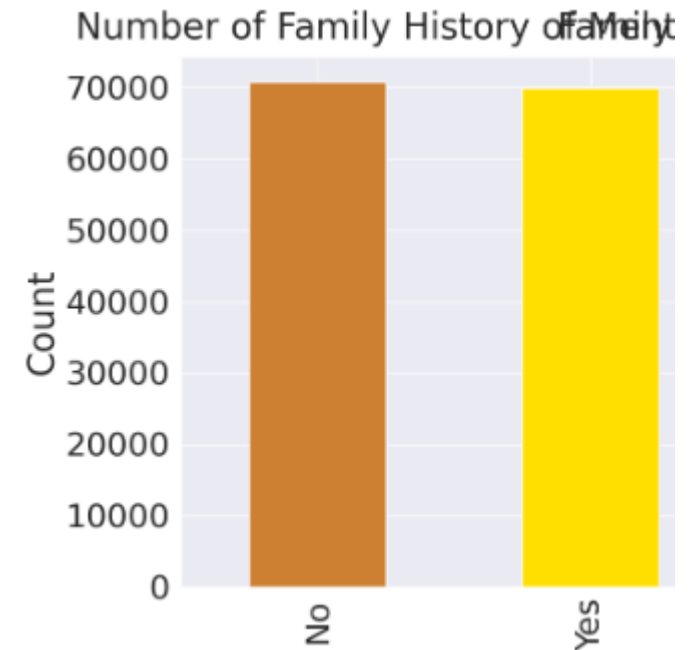
Financial Stress with Depression using Boxplot

재무적 스트레스에 대한 Depression 비율 시각화



Family History with Depression

가족력에 대한 Depression 비율 시각화



Mental Health

Preprocess

```
columns:          id          Percent of NaN Value : 0.000000%
columns:          Name        Percent of NaN Value : 0.000000%
columns:          Gender      Percent of NaN Value : 0.000000%
columns:          Age         Percent of NaN Value : 0.000000%
columns:          City        Percent of NaN Value : 0.000000%
columns:          Working Professional or Student Percent of NaN Value : 0.000000%
columns:          Sleep Duration Percent of NaN Value : 0.000000%
columns:          Have you ever had suicidal thoughts ? Percent of NaN Value : 0.000000%
columns:          Work/Study Hours Percent of NaN Value : 0.000000%
columns:          Family History of Mental Illness Percent of NaN Value : 0.000000%
columns:          Depression  Percent of NaN Value : 0.000000%
-----
columns:          Profession   Percent of NaN Value : 26.034115%
columns:          Academic Pressure Percent of NaN Value : 80.172708%
columns:          Work Pressure Percent of NaN Value : 19.842217%
columns:          CGPA        Percent of NaN Value : 80.171997%
columns:          Study Satisfaction Percent of NaN Value : 80.172708%
columns:          Job Satisfaction Percent of NaN Value : 19.836532%
columns:          Dietary Habits Percent of NaN Value : 0.002843%
columns:          Degree       Percent of NaN Value : 0.001421%
columns:          Financial Stress Percent of NaN Value : 0.002843%
```


Mental Health

Preprocess

Edit

Social Work Pressure = Academic + Work
Social Work Satisfaction = Study + Work

Delete

Name, CGPA, Degree, City, Profession,
Sleep Duration, Dietary Habits

Mental Health

Preprocess

Delete

Name, CGPA, Degree, City, Profession,
Sleep Duration, Dietary Habits

etc. (Null Data, Categorizing, ...)

Null Data

Social Work Satisfaction : Replace with Mean

Categorizing

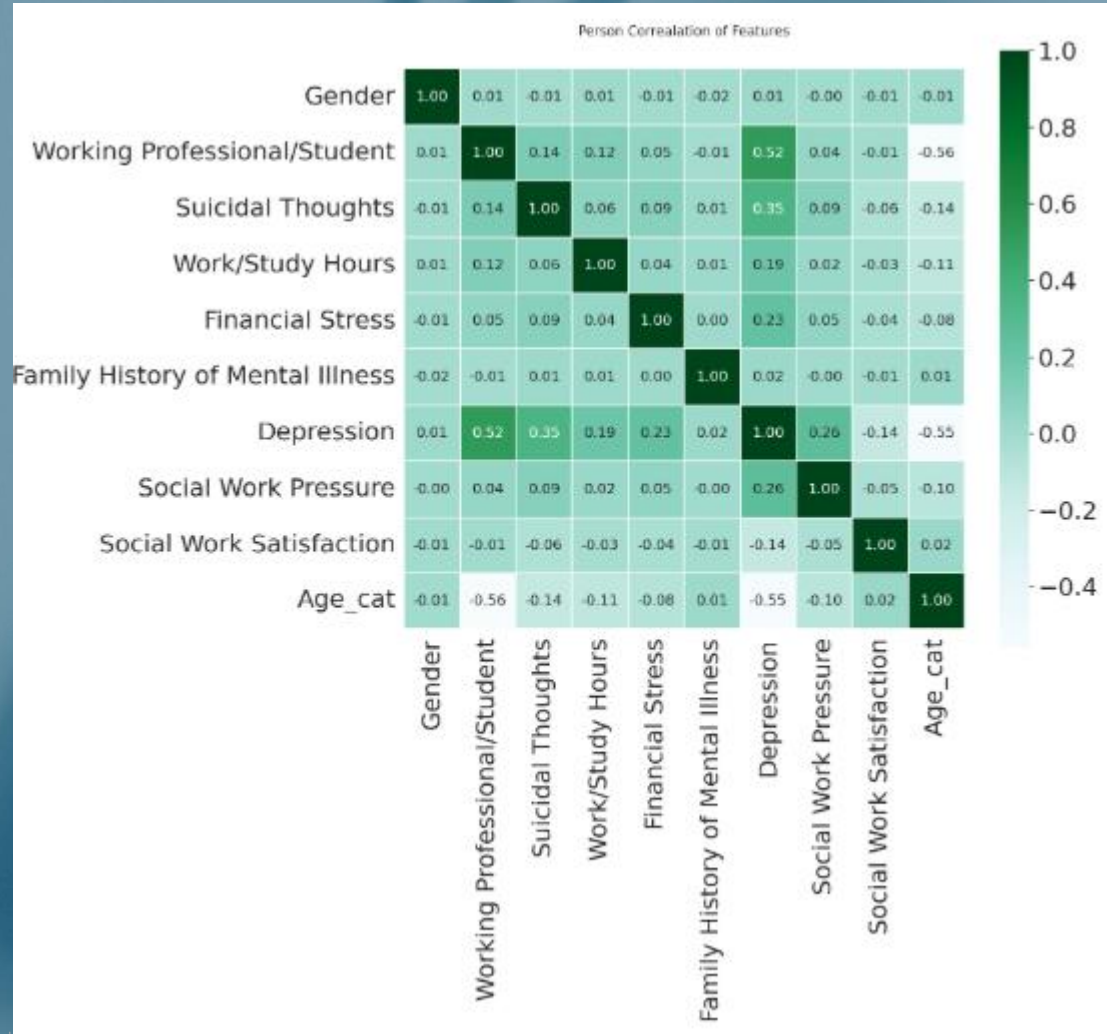
Age : 10단위 구분 (0 - 7)

Quantifying

Gender, WPOS, Suicidal Thoughts, etc. : 0/1

Mental Health

Correlation



Pearson Correlation Coefficient

연속형 데이터 X 연속형 데이터

Mental Health

Machine Learning

Step1. Import Packages

RandomForestClassifier

Ensemble

scikit-learn

Model

train_test_split

Metrics

모델의 성능 평가

Mental Health

Machine Learning

Step2. Setting the Target

-



Mental Health

Machine Learning

Step3. Adjust size of train & test set



Mental Health

Machine Learning

Step4. Random Forest



Mental Health

Machine Learning

Summary

```
#sklearn 모듈 설치
from sklearn.ensemble import RandomForestClassifier
from sklearn import metrics
from sklearn.model_selection import train_test_split

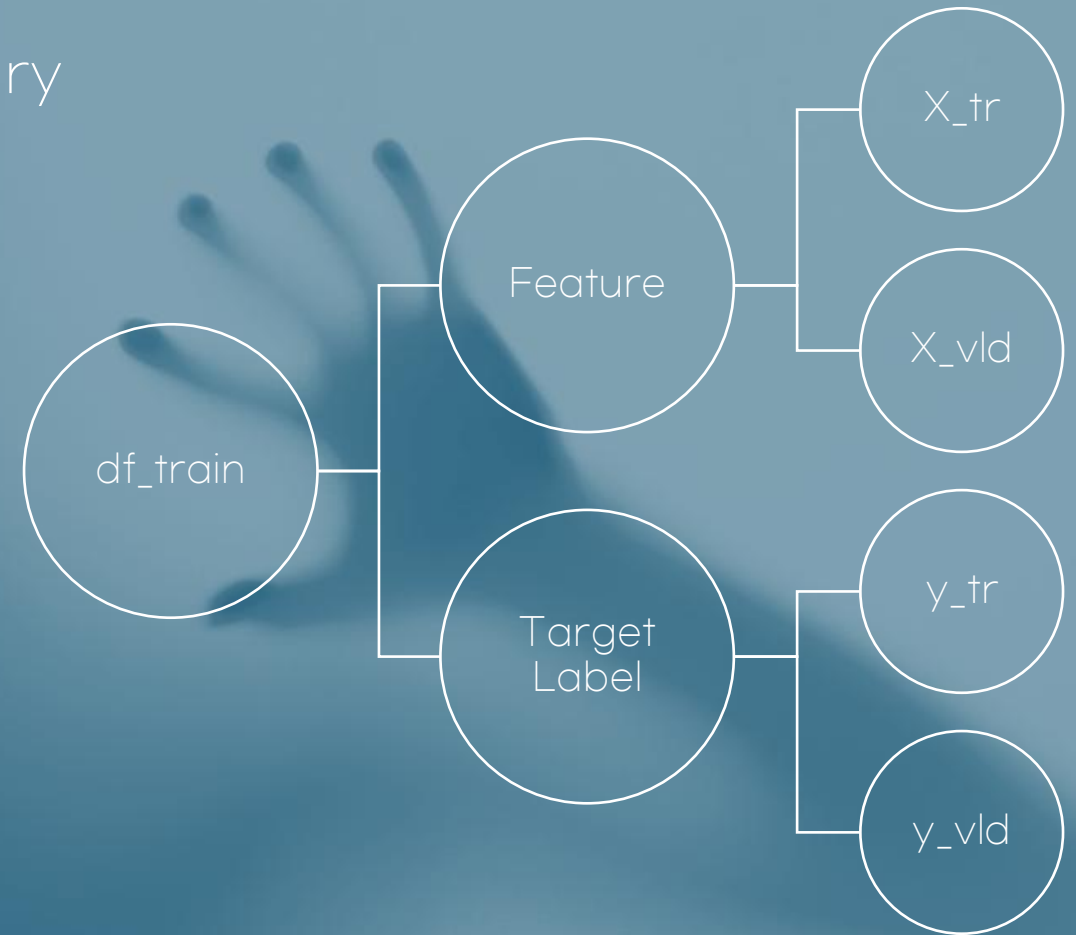
#X_train: Depression을 제외한 입력변수 설정
#target_label: Depression을 출력변수(타겟) 설정
#X_test: 테스트 파일을 Numpy 배열로 변환
X_train = df_train.drop('Depression', axis=1).values
target_label = df_train['Depression'].values
X_test = df_test.values

#훈련 및 검증 데이터를 테스트 세트(70%)와 검증 세트(30%)로 분리
#시드값을 고정하여 동일한 재현
X_tr, X_vld, y_tr, y_vld = train_test_split(X_train, target_label, test_size=0.3, random_state=2024)

#머신러닝 모델 설정 (RandomForestClassifier)
#fit함수로 데이터 입력
#prediction 변수에 예측값 입력 (변수는 Test의 입력데이터(X_vld))
model = RandomForestClassifier()
model.fit(X_tr, y_tr)
prediction = model.predict(X_vld)

#Pandas를 활용해 Feature Importance 추출
from pandas import Series

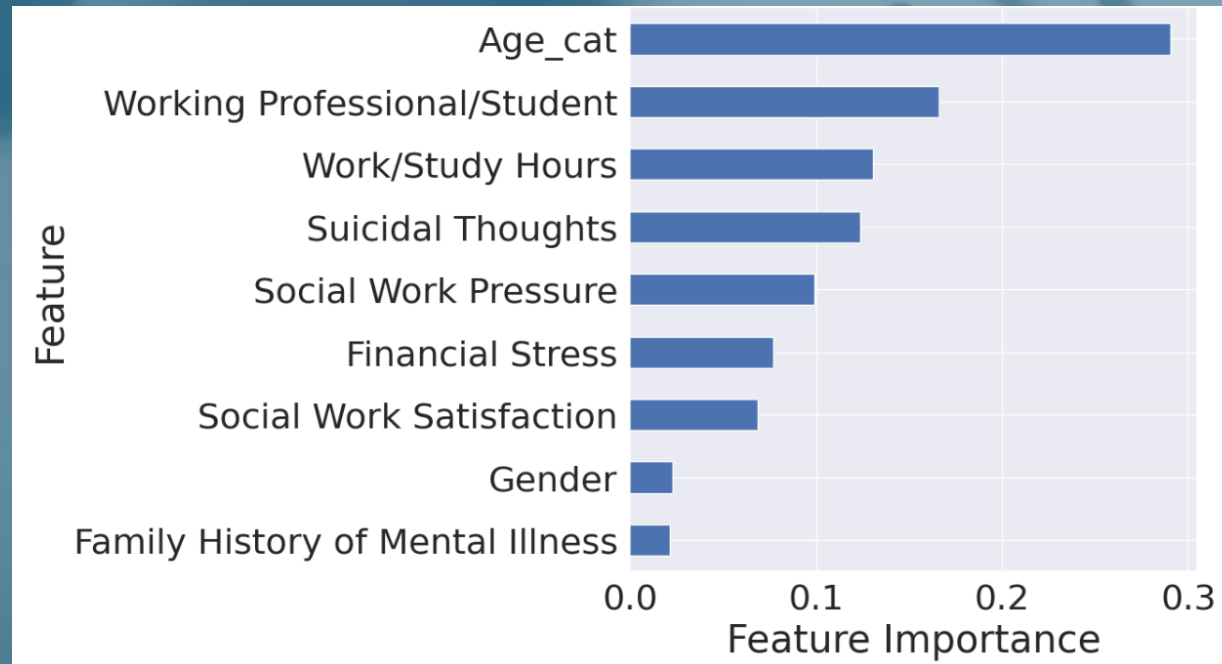
feature_importance = model.feature_importances_
Series_feat_imp = Series(feature_importance, index=df_test.columns)
```



Mental Health

Machine Learning

Conclusion



Mental Health

Machine Learning

Conclusion

-

```
print('총 {}명 중 {:.2f}% 정확도로 우울증 맞춤'.format(y_vid.shape[0], 100 * metrics.accuracy_score(prediction, y_vid)))
```

총 42210명 중 91.83% 정확도로 우울증 맞춤