```
In [1]: # 데이터 분석 준비
import pandas as pd
import numpy as np
```

In [2]: # 데이터 로딩

데이터 로딩 및 확인

```
EPL_df = pd.read_csv('./datasets/results_00_19.csv', encoding='cp949')
        EPL_test = pd.read_csv('./datasets/test20_22.csv', encoding='cp949')
In [3]: print(EPL_df.columns.values)
        ['Season' 'DateTime' 'HomeTeam' 'AwayTeam' 'FTHG' 'FTAG' 'FTR' 'HTHG'
          'HTAG' 'HTR' 'Referee' 'HS' 'AS' 'HST' 'AST' 'HC' 'AC' 'HF' 'AF' 'HY'
          'AY' 'HR' 'AR']
In [4]: # preview the data
        EPL df.head()
                                            AwayTeam FTHG FTAG FTR HTHG HTAG HTR ...
                                                                                            HST AST HC
                                                                                                         AC HF AF
                                                                                                                              HR
           Season
                     DateTime HomeTeam
                                                                                                                     HY AY
Out[4]:
                      2000-08-
                                                                0
                                                                    Н
                                                                           2
                                                                                 0
                                                                                                       6
                                                                                                           6
                                                                                                                           2
                                                                                                                               0
            Jan.00
                                 Charlton
                                              Man City
                                                                                      Н
                                                                                              14
                                                                                                              13
                                                                                                                  12
                   19T00:00:00Z
                      2000-08-
            Jan.00
                                  Chelsea
                                             West Ham
                                                                2
                                                                                 0
                                                                                              10
                                                                                                   5
                                                                                                              19
                                                                                                                  14
                                                                                                                           2
                                                                                                                               0
                   19T00:00:00Z
                      2000-08-
            Jan.00
                                 Coventry Middlesbrough
                                                                3
                                                                    Α
                                                                                      D
                                                                                              3
                                                                                                   9
                                                                                                              15
                                                                                                                  21
                                                                                                                       5
                                                                                                                           3
                                                                                                                               1
                   19T00:00:00Z
                      2000-08-
                                                         2
                                                                2
                                                                    D
                                                                                 2
                                                                                                   6
                                                                                                       5
                                                                                                           8
                                                                                                                  13
                                                                                                                               0
            Jan.00
                                   Derby
                                          Southampton
                                                                                                              11
                   19T00:00:00Z
                      2000-08-
                                                                           2
                                                         2
                                                                0
                                                                                 0
                                                                                              8
                                                                                                   6
                                                                                                       6
                                                                                                                           3
            Jan 00
                                              Everton
                                                                    Н
                                                                                      Н
                                                                                                           4
                                                                                                              21
                                                                                                                  20
                                                                                                                       1
                                                                                                                               0
                                   Leeds
                  19T00:00:00Z
        5 rows × 23 columns
In [5]: EPL df.info() # null 값 존재 x
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 7600 entries, 0 to 7599
        Data columns (total 23 columns):
         #
              Column
                        Non-Null Count Dtype
         0
                         7600 non-null
              Season
                                          object
          1
              DateTime
                         7600 non-null
                                          object
              HomeTeam
                         7600 non-null
                                          object
          3
              AwayTeam
                         7600 non-null
                                          object
          4
              FTHG
                         7600 non-null
                                          int64
          5
              FTAG
                         7600 non-null
                                          int64
          6
              FTR
                         7600 non-null
                                          object
          7
              HTHG
                         7600 non-null
                                          int64
          8
              HTAG
                         7600 non-null
                                          int64
          9
              HTR
                         7600 non-null
                                          object
          10
              Referee
                         7600 non-null
                                          object
                         7600 non-null
              HS
          11
                                          int64
          12
              AS
                         7600 non-null
                                          int64
                         7600 non-null
                                          int64
          13
              HST
                         7600 non-null
              AST
                                          int64
          14
          15
              HC
                         7600 non-null
                                          int64
          16
              AC
                         7600 non-null
                                          int64
              HF
          17
                         7600 non-null
                                          int64
          18
              ΑF
                         7600 non-null
                                          int64
          19
              HY
                         7600 non-null
                                          int64
          20
              ΑY
                         7600 non-null
                                          int64
          21
             HR
                         7600 non-null
                                          int64
                         7600 non-null
          22 AR
                                          int64
        dtypes: int64(16), object(7)
        memory usage: 1.3+ MB
In [6]: EPL_df.describe()
```

```
HTHG
                                                                 HTAG
                                                                                                                                                      AC
Out[6]:
           count 7600.000000
                               7600.000000
                                             7600.000000
                                                          7600.000000 7600.000000
                                                                                      7600.000000
                                                                                                    7600.000000
                                                                                                                 7600.000000
                                                                                                                               7600.000000 7600.000000 7
           mean
                     1.528947
                                   1.140132
                                                 0.682500
                                                              0.498947
                                                                           13.526316
                                                                                         10.566842
                                                                                                       6.255395
                                                                                                                     4.828158
                                                                                                                                   6.129211
                                                                                                                                                4.787632
                     1.296113
                                   1.131548
                                                 0.831894
                                                              0.714742
                                                                            5.230781
                                                                                          4.529502
                                                                                                       3.351286
                                                                                                                     2.815406
                                                                                                                                   3.100482
                                                                                                                                                2.730690
             std
            min
                     0.000000
                                   0.000000
                                                 0.000000
                                                              0.000000
                                                                            0.000000
                                                                                          0.000000
                                                                                                       0.000000
                                                                                                                     0.000000
                                                                                                                                   0.000000
                                                                                                                                                0.000000
            25%
                     1.000000
                                   0.000000
                                                 0.000000
                                                              0.000000
                                                                           10.000000
                                                                                          7.000000
                                                                                                       4.000000
                                                                                                                     3.000000
                                                                                                                                   4.000000
                                                                                                                                                3.000000
            50%
                     1.000000
                                   1.000000
                                                 0.000000
                                                              0.000000
                                                                           13.000000
                                                                                         10.000000
                                                                                                       6.000000
                                                                                                                     4.000000
                                                                                                                                  6.000000
                                                                                                                                                4.000000
            75%
                     2.000000
                                   2.000000
                                                 1.000000
                                                              1.000000
                                                                           17.000000
                                                                                         13.000000
                                                                                                       8.000000
                                                                                                                     6.000000
                                                                                                                                   8.000000
                                                                                                                                                6.000000
            max
                     9.000000
                                   9.000000
                                                 5.000000
                                                              5.000000
                                                                           43.000000
                                                                                        30.000000
                                                                                                      24.000000
                                                                                                                    20.000000
                                                                                                                                  20.000000
                                                                                                                                               19.000000
```

```
Missing Value 처리
        # check missing values in train dataset
        EPL df.isnull().sum()
        Season
Out[7]:
        DateTime
                    0
        HomeTeam
                    0
        AwayTeam
                    0
        FTHG
                    0
        FTAG
                    0
        FTR
                    0
        HTHG
                    0
        HTAG
                    0
                    0
        HTR
        Referee
                    0
        HS
                    0
        AS
                    0
        HST
                    0
        AST
                    0
        HC
                    0
                    0
        AC
        HF
                    0
                    0
        ΑF
        HY
                    0
                    0
        ΔY
        HR
                    0
        AR
                    0
        dtype: int64
In [8]: EPL_df[EPL_df.DateTime.isnull()] # 결측이 있는 행을 찾음
          Season DateTime HomeTeam AwayTeam FTHG FTAG FTR HTHG HTAG HTR \dots
                                                                             HST AST HC AC HF AF HY AY HR AR
Out[8]:
        0 rows × 23 columns
In [9]:
        EPL df = EPL df.dropna() # 레코드의 대부분이 결측이므로 레코드를 삭제
In [10]:
        # 시각화 패키지
        import seaborn as sns
         import matplotlib.pyplot as plt
         %matplotlib inline
         경기 결과와 경기 기록 비교
```

```
# 홈팀 어웨이팀으로 나뉘어있는 feature를 분석하기 좋게 하나의 새로운 feature로 표현
          EPL df['DIFF FG'] = EPL df['FTHG'] - EPL df['FTAG'] # 홈팀 풀타임 골 - 어웨이팀 풀타임 골
         EPL_df['DIFF_HG'] = EPL_df['HTHG'] - EPL_df['HTAG'] # 홈팀 하프타임 골 - 어웨이팀 하프타임 골
EPL_df['DIFF_SHOOT'] = EPL_df['HS'] - EPL_df['AS'] # 홈팀 슈팅 수 - 어웨이팀 슈팅 수
          EPL_df['DIFF_ST'] = EPL_df['HST'] - EPL_df['AST'] # 홈팀 유효슈팅 수 - 어웨이팀 유효슈팅 수
          EPL_df['DIFF_FOUL'] = EPL_df['HF'] - EPL_df['AF'] # 홈팅 파울 - 어웨이팀 파울
          EPL df['DIFF CONER'] = EPL df['HC'] - EPL df['AC']# 홈팀 코너킥 - 어웨팀 코너킥
          EPL_df['DIFF_YC'] = EPL_df['HY'] - EPL_df['AY'] # 홈팀 옐로카드 - 어웨이팀 옐로카드
          EPL df['DIFF RC'] = EPL df['HR'] - EPL df['AR'] # 홈팀 레드카드 - 어웨이팀 레드카드
         list dif = ["FTR", "DIFF FG", "DIFF HG", "DIFF SHOOT", "DIFF ST", "DIFF FOUL", "DIFF CONER", "DIFF YC", "DIFF RC"]
         EPL_df[list_dif].groupby(['FTR'], as_index=False).mean().sort_values(by='FTR', ascending=False)
                 DIFF_FG DIFF_HG DIFF_SHOOT
                                               DIFF_ST DIFF_FOUL DIFF_CONER DIFF_YC DIFF_RC
Out[12]:
                                                         -0.746387
                 1.884387
                          0.836498
                                      5.034287
                                               3.215642
                                                                     1.491924 -0.516860
                                                                                     -0.085577
              Н
              D
                 0.000000
                          0.006792
                                      2.684431
                                               1.080982
                                                         -0.635841
                                                                     1.549634 -0.377743 -0.019331
              A -1.713027 -0.727863
                                     -0.191006 -1.191470
                                                         -0.214650
                                                                     0.910987 -0.075568
```

데이터 전처리: 속성 조정

```
In [13]: # train / test split
          from sklearn.model_selection import train_test_split
          # 필요한 feature만 뽑아 새로운 df에 저장
In [14]:
          # 풀타임 결과(H, D, R), 슈팅수, 유효슈팅수, 파울, 코너킥, 옐로카드, 레드카드 EPL_df_ext = EPL_df[['FTR', 'HS', 'AS', 'HST', 'AST', 'HF', 'AF', 'HC', 'AC', 'HY', 'AY', 'HR', 'AR']]
          print(EPL_df_ext)
                FTR
                     HS
                          AS
                               HST
                                    AST
                                          HF
                                               ΑF
                                                   HC
                                                        AC
                                                            HY
                                                                 ΑY
                                                                      HR
                                                                          AR
          0
                                                                           0
                  Н
                      17
                           8
                                14
                                       4
                                          13
                                               12
                                                     6
                                                         6
                                                              1
                                                                  2
                                                                       0
          1
                  Н
                      17
                          12
                                10
                                       5
                                          19
                                               14
                                                     7
                                                         7
                                                              1
                                                                  2
                                                                       0
                                                                           0
          2
                                       9
                                               21
                                                                  3
                          16
                                 3
                                          15
          3
                  D
                       6
                          13
                                 4
                                       6
                                          11
                                               13
                                                     5
                                                         8
                                                              1
                                                                  1
                                                                       0
                                                                           0
          4
                  H 17
                          12
                                 8
                                       6
                                          21
                                               20
                                                     6
                                                         4
                                                              1
                                                                  3
                                                                       0
                                                                           0
                                                     3
                                                                           0
          7595
                  Α
                      14
                                 3
                                       3
                                          12
                                               11
                                                              1
                                                                       1
          7596
                     31
                            5
                                                     9
                  Н
                                10
                                       4
                                           7
                                                         0
                                                                  1
                                                                       0
                                                4
                                                              1
                                                                           0
                                                5
          7597
                  Α
                      3
                          14
                                 2
                                       6
                                         11
                                                     2
                                                         4
                                                              1
                                                                  0
                                                                       0
                                                                           0
          7598
                  Н
                     13
                           5
                                 4
                                       3
                                           9
                                               16
                                                     9
                                                         1
                                                              0
                                                                  1
                                                                       0
                                                                           0
          7599
                  D 10
                          13
                                         16
                                              13
          [7600 rows x 13 columns]
In [15]: # FTR 무승부 지우고 binary 형태로
# 무승부 경기를 제외하고 홈팀이 이긴 것을 RESULT의 1로
           EPL_df_ext = EPL_df_ext[EPL_df_ext.FTR != 'D']
          EPL_df_ext['RESULT'] = np.where(EPL_df_ext['FTR']=='H', 1, 0)
EPL_df_ext.drop('FTR', axis=1, inplace=True)
          print(EPL_df_ext)
          EPL_test_ext = EPL_test_ext[EPL_test_ext.FTR != 'D']
          EPL_test_ext['RESULT'] = np.where(EPL_test_ext['FTR']=='H', 1, 0)
EPL_test_ext.drop('FTR', axis=1, inplace=True)
          print(EPL test ext)
                 HS
                     AS
                          HST
                                AST
                                     HF
                                          ΑF
                                               HC
                                                   AC HY
                                                            AY
                                                                 HR
                                                                      AR RESULT
          0
                 17
                                      13
                                          12
                      8
                            14
                                                6
                                                     6
                                                         1
                                                                  0
                                                                                1
          1
                 17
                     12
                            10
                                  5
                                      19
                                          14
                                                7
                                                     7
                                                         1
                                                              2
                                                                  0
                                                                       0
                                                                                1
          2
                  6
                     16
                            3
                                  9
                                      15
                                          21
                                                8
                                                     4
                                                         5
                                                              3
                                                                  1
                                                                       0
                                                                                0
          4
                 17
                      12
                             8
                                  6
                                      21
                                          20
                                                6
                                                     4
                                                         1
                                                              3
                                                                  0
                                                                       0
                                                                                1
          6
                 16
                      3
                            10
                                       8
                                           8
                                                6
                                                     1
                                                              1
                                                                  0
                                                                       0
                                                                                1
                                  2
                                                         1
          7594
                                  7
                 13
                      13
                            5
                                      11
                                           9
                                                2
                                                     5
                                                         1
                                                              0
                                                                  0
                                                                       0
                                                                                0
          7595
                 14
                            3
                                  3
                                      12
                                          11
                                                                  1
                                                                                0
                            10
                                                9
          7596
                       5
                                  4
                                       7
                                           4
                                                     0
                                                                  0
                                                                       0
                 31
                                                         1
                                                              1
                                                                                1
                                      11
          7597
                  3
                     14
                            2
                                  6
                                           5
                                                2
                                                     4
                                                         1
                                                              0
                                                                  0
                                                                       0
                                                                                0
          7598
                 13
                       5
                                          16
                                                9
          [5686 rows x 13 columns]
                                         AF
                HS AS HST
                               AST HF
                                              HC
                                                  \mathsf{AC}
                                                       HY
                                                           AY
                                                                HR
                                                                    AR
                                                                         RESULT
          0
                 5
                     13
                                 6
                                    12
                                         12
                                                   3
                                                            2
                                                                 0
                                                                      0
                                                                               0
                                               7
                            3
                                    14
                                                        2
          1
                 5
                     9
                                 5
                                         11
                                                   3
                                                            1
                                                                 0
                                                                      0
                                                                               1
          2
                22
                     6
                            6
                                 3
                                      9
                                          6
                                               q
                                                   0
                                                        1
                                                            0
                                                                 0
                                                                      0
                                                                               1
          3
                15
                     15
                            3
                                 2
                                    13
                                          7
                                               8
                                                   7
                                                        2
                                                            2
                                                                 0
                                                                      0
                                                                               0
                 7
                     13
                            1
                                    12
                                          9
                                               2
                                                   5
                                                        1
                                                            1
                                                                      0
                                                                               0
          683
                 8
                     10
                            1
                                 5
                                      8
                                         12
                                               3
                                                   8
                                                            0
                                                                 0
                                                                      0
                                                                               0
                 9
          684
                     11
                            8
                                    12
                                         14
                                                                               0
          685
                                          6
                                               4
                                                        0
                15
                      5
                            7
                                 1
                                      2
                                                                 0
                                                                      0
                                                                               1
                                                   6
                                                            1
          686
                12
                     11
                            3
                                 3
                                    11
                                         12
                                               3
                                                   4
                                                        1
                                                            1
                                                                 0
                                                                      0
                                                                               1
          687
                17
                                         10
                                               6
                                                   7
                                                            1
                                                                      0
                                                                               1
                    18
                                    12
          [534 rows x 13 columns]
In [16]: # 데이터 내에서 학습 집합과 테스트 집합을 나눔(8:2)
           # train, test = train_test_split(EPL_df_ext, test_size=0.2, random_state=12)
          train = EPL df ext
           test = EPL test ext
          X_train = train.drop("RESULT", axis=1) # 열 지움
          Y_train = train["RESULT"
          X test = test.drop("RESULT", axis=1)
           Y test = test["RESULT"]
          X train.shape, Y train.shape, X test.shape, Y test.shape
          print(X_train)
```

```
HS AS HST AST HF AF HC AC HY AY HR AR
      17 8
17 12
                     4 13 12
5 19 14
1
                10
                    9 15 21 8 4
6 21 20 6 4
2 8 8 6 1
      6 16
      17 12
                                              3 0 0
1 0 0
4
                8
                                           1
              10
6
      16 3
                                           1
                                               0 0 0
4 1 0
              5
                    7 11 9
3 12 11
7594 13 13 5
7595 14 7 3
7596 31 5 10
                              9
                                           1
                    4 7 4 9 0
6 11 5 2 4
      3 14
7597
                                                0
7598 13
          5
                         9
                             16
                      3
```

기계학습

[5686 rows x 12 columns]

```
In [17]: # 기계 학습
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC, LinearSVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB # 이산형이 아닌 연속형에 대한
from sklearn.linear_model import Perceptron
from sklearn.linear_model import SGDClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.neural_network import MLPClassifier
from sklearn.cluster import KMeans
```

Random Forest

```
In [18]: from sklearn.ensemble import RandomForestClassifier

# n_estimators: 결정 트리 개수(클수록 수행 시간 증가)
rf = RandomForestClassifier(n_estimators = 100, oob_score = True, random_state=123456)

rf.fit(X_train, Y_train)
from sklearn.metrics import accuracy_score

pred_rf = rf.predict(X_test)

# oob_score = out of bag score 예측이 얼마나 정확한지에 대한 추정치
# print(f'Out of bag score estimate: {rf.oob_score_:.9}')
print(f'accuracy: {accuracy_score(Y_test,pred_rf): .4f}')

accuracy: 0.7697
```

Xgboost

Support Vector Machines

```
In [20]: # Support Vector Machines
# SVM 모델 학습
svc = SVC()
svc.fit(X_train, Y_train)
Y_pred_svc = svc.predict(X_test)
acc_svc = svc.score(X_test, Y_test)

print(f'accuracy: {acc_svc: .4f}')
accuracy: 0.7903
```

Logistic Regression

```
# Logistic Regression training
logreg = LogisticRegression(solver='liblinear')
logreg.fit(X_train, Y_train)
# Logistic Regression prediction
Y_pred_logreg = logreg.predict(X_test)
acc_log = logreg.score(X_test, Y_test)
print(f'accuracy: {acc_log: .4f}')
```

보완필요

accuracy: 0.7903

```
In [22]: #####보완필요####
import statsmodels.api as sm

model_fb = sm.Logit.from_formula("RESULT ~ HR+AS+HST+AST+HF+AF+HC+AC+HY+AY+HR+AR",EPL_df_ext )
result_fb = model_fb.fit()
print(result_fb.summary())
```

Optimization terminated successfully.

Current function value: 0.509022

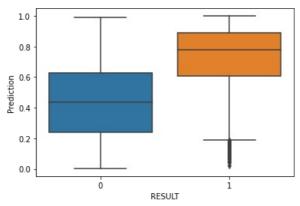
Iterations 6

Logit Regression Results

_____ _____ Dep. Variable: RESULT No. Observations: 5686 Df Residuals: Model: Logit 5674 Method: MLE Df Model: 11 0.2331 Mon, 25 Apr 2022 Pseudo R-squ.: Date: Time: 22:36:57 Log-Likelihood: -2894.3 converged: True LL-Null: -3774.1 Covariance Type: nonrobust LLR p-value: 0.000 ______ coef std err P>|z| [0.025 0.975] Intercept 0.3346 0.199 1.678 0.093 -0.056 0.725 -1.0371 0.133 -7.804 0.000 -1.298 -0.777 AS 0.0171 0.011 1.543 0.123 -0.005 0.039 23.876 **HST** 0.3035 0.013 0.000 0.279 0.328 **AST** -0.3711 0.018 -21.087 0.000 -0.406 -0.337 ΗF 0.0233 0.009 2.513 0.012 0.005 0.041 -0.0088 0.009 -0.990 0.322 -0.026 0.009 ΑF -8.408 0.000 HC -0.0996 0.012 -0.123 -0.076 AC0.1260 0.014 9.291 0.000 0.099 0.153 HY -0.2059 0.029 -7.015 0.000 -0.263 -0.148 0.0858 0.002 0.141 0.028 3.038 0.030 AY AR 0.6668 0.125 5.332 0.000 0.422 0.912

```
import seaborn as sns
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline

EPL_df_ext["Prediction"] = result_fb.predict(EPL_df_ext)
sns.boxplot(x="RESULT", y="Prediction", data=EPL_df_ext)
plt.show()
```



```
In [24]: model_fb = sm.Logit.from_formula("RESULT ~ HST + HF+ AC + AY + AR", EPL_df_ext )
    result_fb = model_fb.fit()
    print(result_fb.summary())
```

Optimization terminated successfully.

Current function value: 0.597698

Iterations 6

Logit Regression Results

N N D T	Dep. Variab Model: Method: Date: Time: converged:	Мо	Lo n, 25 Apr 2 22:30	ogit Df R MLE Df M 2022 Pseu 5:57 Log- True LL-N		:	5686 5680 5 0.09951 -3398.5 -374.1
(Covariance ⁻	lype: 	nonrol	oust LLR	p-value: 		4.361e-160
		coef	std err	Z	P> z	[0.025	0.975]
]	Intercept	-1.0399	0.131	-7.924	0.000	-1.297	-0.783
H	HST	0.2449	0.011	23.192	0.000	0.224	0.266
H	HF	-0.0161	0.008	-2.094	0.036	-0.031	-0.001
F	AC	0.0174	0.011	1.614	0.106	-0.004	0.038
F	ΑY	0.0600	0.024	2.534	0.011	0.014	0.106
F	AR	0.6467	0.112	5.749	0.000	0.426	0.867
=			========	========	========	=======	========

k-Nearest Neighbor

```
In [25]: knn = KNeighborsClassifier(n_neighbors = 3)
knn.fit(X_train, Y_train)
Y_pred_knn = knn.predict(X_test)
acc_knn = knn.score(X_test, Y_test)
print(f'accuracy: {acc_knn: .4f}')
accuracy: 0.6966
```

Naive Bayes classifiers

```
In [26]: # Gaussian Naive Bayes
gaussian = GaussianNB()
gaussian.fit(X_train, Y_train)
Y_pred_NB = gaussian.predict(X_test)
acc_gaussian = gaussian.score(X_test, Y_test)
print(f'accuracy: {acc_gaussian: .4f}')
accuracy: 0.7154
```

Decision tree

```
In [27]: # Decision Tree
    decision_tree = DecisionTreeClassifier()
    decision_tree.fit(X_train, Y_train)
    Y_pred_DT = decision_tree.predict(X_test)
    acc_decision_tree = decision_tree.score(X_test, Y_test)
    print(f'accuracy: {acc_decision_tree: .4f}')
    accuracy: 0.6966
```

Artificial Neural Network

```
In [28]: # 수정
ANN = MLPClassifier(solver='lbfgs', alpha=1, hidden_layer_sizes=(30, 10), random_state=1, max_iter=5000)
ANN.fit(X_train, Y_train)
Y_pred_ANN = ANN.predict(X_test)
acc_ANN = ANN.score(X_test, Y_test)
print(f'accuracy: {acc_ANN: .4f}')
accuracy: 0.7715
```

Keras: ANN 모델

```
In [29]: from keras import layers, models, datasets from keras.utils import np_utils

Using TensorFlow backend.
```

```
In [30]: def plot(hist):
```

```
loss_ax.plot(hist.history['loss'], 'y', label='train loss')
loss_ax.plot(hist.history['val loss'], 'r', label='val loss')
        acc_ax.plot(hist.history['accuracy'], 'b', label='train acc')
acc_ax.plot(hist.history['val_accuracy'], 'g', label='val acc')
        loss_ax.set_xlabel('epoch')
        loss_ax.set_ylabel('loss')
        acc ax.set ylabel('accuray')
         loss_ax.legend(loc='upper left')
        acc ax.legend(loc='lower left')
         plt.show()
In [31]: import keras
      from keras.models import Sequential
      from keras.layers import Dense, Dropout, Activation
      model = Sequential()
      model.add(Dense(64, activation='relu', input_dim=12))
      model.add(Dropout(0.2))
      model.add(Dense(64, activation='relu'))
      model.add(Dropout(0.2))
      model.add(Dense(1, activation='sigmoid'))
      model.compile(loss='binary_crossentropy',optimizer='adam', metrics=['accuracy'])
      history = model.fit(X train, Y train, epochs=200,
           batch size=32, validation split=0.2)
      performance_test = model.evaluate(X_test, Y_test, batch_size=32)
      classes = model.predict(X_test, batch_size=128)
      performance test
      2022-04-25 22:37:11.265254: I tensorflow/core/platform/cpu feature guard.cc:145] This TensorFlow binary is opti
      mized with Intel(R) MKL-DNN to use the following CPU instructions in performance critical operations: SSE4.1 S
      SE4.2 AVX AVX2 FMA
      To enable them in non-MKL-DNN operations, rebuild TensorFlow with the appropriate compiler flags.
      2022-04-25 22:37:11.265758: I tensorflow/core/common_runtime/process_util.cc:115] Creating new thread pool with
      default inter op setting: 12. Tune using inter_op_parallelism_threads for best performance.
      Train on 4548 samples, validate on 1138 samples
      Epoch 1/200
      4548/4548 [==
                         - val accuracy: 0.7671
      Epoch 2/200
      - val_accuracy: 0.7671
      Epoch 3/200
      4548/4548 [=
                        =========] - 0s 103us/step - loss: 0.5717 - accuracy: 0.7040 - val loss: 0.5030
      - val_accuracy: 0.7680
      Epoch 4/200
      4548/4548 [=================== - os 104us/step - loss: 0.5659 - accuracy: 0.7128 - val loss: 0.4911
      - val accuracy: 0.7654
      Fnoch 5/200
      - val accuracy: 0.7698
      Epoch 6/200
      val accuracy: 0.7768
      Epoch 7/200
      - val accuracy: 0.7786
      Epoch 8/200
      - val_accuracy: 0.7830
      Epoch 9/200
      4548/4548 [====
                  val_accuracy: 0.7786
      Epoch 10/200
      - val accuracy: 0.7865
      Epoch 11/200
      4548/4548 [==
                      ==========] - 0s 105us/step - loss: 0.5312 - accuracy: 0.7381 - val_loss: 0.4986
      - val accuracy: 0.7786
      Fnoch 12/200
      - val_accuracy: 0.7786
      Epoch 13/200
      - val_accuracy: 0.7935
      Epoch 14/200
      - val_accuracy: 0.7847
      Epoch 15/200
```

fig, loss_ax = plt.subplots()
acc_ax = loss_ax.twinx()

- val_accuracy: 0.7882

```
Epoch 16/200
4548/4548 [==
             ========] - 0s 104us/step - loss: 0.5200 - accuracy: 0.7482 - val_loss: 0.4790
- val accuracy: 0.7777
Epoch 17/200
- val_accuracy: 0.7856
Epoch 18/200
- val_accuracy: 0.7856
Epoch 19/200
- val_accuracy: 0.7900
Epoch 20/200
       4548/4548 [===
val_accuracy: 0.7900
Epoch 21/200
4548/4548 [=============== ] - 0s 106us/step - loss: 0.5153 - accuracy: 0.7454 - val loss: 0.4692
- val accuracy: 0.7865
Epoch 22/200
4548/4548 [===
       - val_accuracy: 0.7768
Epoch 23/200
val_accuracy: 0.7873
Epoch 24/200
- val_accuracy: 0.7909
Epoch 25/200
4548/4548 [==
             :========] - 0s 109us/step - loss: 0.5123 - accuracy: 0.7489 - val loss: 0.4590
- val_accuracy: 0.7882
Epoch 26/200
- val_accuracy: 0.7891
Epoch 27/200
- val_accuracy: 0.7856
Epoch 28/200
- val_accuracy: 0.7891
Epoch 29/200
val_accuracy: 0.7856
Epoch 30/200
val_accuracy: 0.7847
Epoch 31/200
4548/4548 [==
                =====] - 1s 115us/step - loss: 0.5092 - accuracy: 0.7456 - val loss: 0.4641
val accuracy: 0.7873
Epoch 32/200
4548/4548 [==
             :========] - 1s 118us/step - loss: 0.5086 - accuracy: 0.7524 - val loss: 0.4626
val_accuracy: 0.7891
Epoch 33/200
4548/4548 [===
             :========] - 1s 121us/step - loss: 0.5060 - accuracy: 0.7478 - val loss: 0.4627
- val_accuracy: 0.7926
Epoch 34/200
4548/4548 [==
        - val_accuracy: 0.7891
Epoch 35/200
4548/4548 [==
           - val accuracy: 0.7847
Fnoch 36/200
- val_accuracy: 0.7891
Epoch 37/200
- val_accuracy: 0.7900
Epoch 38/200
val_accuracy: 0.7847
Epoch 39/200
- val_accuracy: 0.7803
Epoch 40/200
4548/4548 [==
            ==============] - 1s 112us/step - loss: 0.4992 - accuracy: 0.7577 - val loss: 0.4645
- val_accuracy: 0.7891
Epoch 41/200
val_accuracy: 0.7847
Epoch 42/200
4548/4548 [==
             :========] - 0s 109us/step - loss: 0.5032 - accuracy: 0.7504 - val loss: 0.4715
val accuracy: 0.7821
Epoch 43/200
- val_accuracy: 0.7882
Epoch 44/200
4548/4548 [==
            =========] - 0s 108us/step - loss: 0.5017 - accuracy: 0.7489 - val_loss: 0.4552
val accuracy: 0.7900
Epoch 45/200
```

```
- val accuracy: 0.7821
Epoch 46/200
4548/4548 [==
               =======] - 0s 107us/step - loss: 0.4995 - accuracy: 0.7487 - val loss: 0.4748
- val_accuracy: 0.7803
Fnoch 47/200
4548/4548 [===
         - val accuracy: 0.7821
Fnoch 48/200
4548/4548 [===
       val accuracy: 0.7882
Epoch 49/200
- val_accuracy: 0.7830
Epoch 50/200
val_accuracy: 0.7891
Epoch 51/200
- val_accuracy: 0.7891
Epoch 52/200
4548/4548 [========
             =========] - 0s 108us/step - loss: 0.4967 - accuracy: 0.7568 - val loss: 0.4634
- val_accuracy: 0.7794
Epoch 53/200
4548/4548 [==
            :==========] - 0s 108us/step - loss: 0.5006 - accuracy: 0.7526 - val loss: 0.4728
val_accuracy: 0.7803
Epoch 54/200
4548/4548 [===
       val_accuracy: 0.7917
Epoch 55/200
4548/4548 [==
              ========] - 0s 107us/step - loss: 0.4895 - accuracy: 0.7551 - val loss: 0.4734
val_accuracy: 0.7812
Epoch 56/200
4548/4548 [===
       - val_accuracy: 0.7891
Epoch 57/200
- val accuracy: 0.7856
Epoch 58/200
- val_accuracy: 0.7803
Epoch 59/200
4548/4548 [===
       - val_accuracy: 0.7856
Epoch 60/200
- val accuracy: 0.7777
Epoch 61/200
4548/4548 [==
        - val accuracy: 0.7891
Epoch 62/200
val_accuracy: 0.7926
Epoch 63/200
4548/4548 [==
                 =====] - 0s 106us/step - loss: 0.4889 - accuracy: 0.7551 - val_loss: 0.4627
val_accuracy: 0.7821
Epoch 64/200
- val_accuracy: 0.7794
Epoch 65/200
4548/4548 [==
                ======] - 0s 106us/step - loss: 0.4883 - accuracy: 0.7579 - val loss: 0.4622
- val_accuracy: 0.7900
Epoch 66/200
- val_accuracy: 0.7865
Epoch 67/200
- val_accuracy: 0.7856
Epoch 68/200
val accuracy: 0.7847
Epoch 69/200
4548/4548 [==================== ] - 0s 108us/step - loss: 0.4869 - accuracy: 0.7586 - val loss: 0.4670
val accuracy: 0.7856
Epoch 70/200
- val accuracy: 0.7873
Epoch 71/200
4548/4548 [===
             - val accuracy: 0.7909
Epoch 72/200
4548/4548 [==
                   ≔=] - 0s 106us/step - loss: 0.4821 - accuracy: 0.7632 - val loss: 0.4734
- val accuracy: 0.7821
Epoch 73/200
4548/4548 [=======
             :========] - 0s 106us/step - loss: 0.4813 - accuracy: 0.7544 - val loss: 0.4615
- val accuracy: 0.7873
Epoch 74/200
4548/4548 [==
             :=========] - 0s 106us/step - loss: 0.4855 - accuracy: 0.7658 - val loss: 0.4722
```

val_accuracy: 0.7926

Epoch 75/200

```
- val_accuracy: 0.7900
Epoch 76/200
- val_accuracy: 0.7882
Epoch 77/200
val_accuracy: 0.7838
Epoch 78/200
4548/4548 [==
           :=========] - 0s 106us/step - loss: 0.4804 - accuracy: 0.7643 - val loss: 0.4808
val accuracy: 0.7742
Epoch 79/200
val accuracy: 0.7856
Epoch 80/200
4548/4548 [===
           - val_accuracy: 0.7847
Epoch 81/200
- val_accuracy: 0.7873
Epoch 82/200
- val_accuracy: 0.7768
Epoch 83/200
- val_accuracy: 0.7821
Epoch 84/200
- val_accuracy: 0.7935
Epoch 85/200
val_accuracy: 0.7856
Epoch 86/200
4548/4548 [==================== ] - 0s 109us/step - loss: 0.4771 - accuracy: 0.7698 - val loss: 0.4768
val_accuracy: 0.7786
Epoch 87/200
val_accuracy: 0.7821
Epoch 88/200
4548/4548 [==:
           :=========] - 0s 109us/step - loss: 0.4739 - accuracy: 0.7628 - val loss: 0.4621
- val accuracy: 0.7856
Epoch 89/200
4548/4548 [==
             val_accuracy: 0.7891
Epoch 90/200
4548/4548 [=====
         - val_accuracy: 0.7873
Epoch 91/200
4548/4548 [==
           ========] - 0s 109us/step - loss: 0.4746 - accuracy: 0.7700 - val loss: 0.4704
- val_accuracy: 0.7847
Epoch 92/200
val accuracy: 0.7891
Fnoch 93/200
4548/4548 [==================== ] - 0s 106us/step - loss: 0.4727 - accuracy: 0.7685 - val loss: 0.4949
- val_accuracy: 0.7847
Epoch 94/200
val_accuracy: 0.7821
Epoch 95/200
- val accuracy: 0.7803
Epoch 96/200
val accuracy: 0.7882
Epoch 97/200
4548/4548 [==
           :=========] - 0s 106us/step - loss: 0.4644 - accuracy: 0.7700 - val loss: 0.4849
val_accuracy: 0.7794
Epoch 98/200
4548/4548 [===
       - val accuracy: 0.7821
Epoch 99/200
4548/4548 [==
           :=========] - 0s 106us/step - loss: 0.4637 - accuracy: 0.7726 - val loss: 0.4767
- val accuracy: 0.7882
Epoch 100/200
4548/4548 [=============== ] - 0s 108us/step - loss: 0.4666 - accuracy: 0.7806 - val loss: 0.4845
- val_accuracy: 0.7847
Epoch 101/200
- val accuracy: 0.7750
Epoch 102/200
val_accuracy: 0.7654
Epoch 103/200
- val_accuracy: 0.7803
Epoch 104/200
```

- val accuracy: 0.7794

```
Epoch 105/200
val accuracy: 0.7900
Epoch 106/200
- val_accuracy: 0.7856
Epoch 107/200
val_accuracy: 0.7830
Epoch 108/200
4548/4548 [==
               ========] - 1s 112us/step - loss: 0.4639 - accuracy: 0.7751 - val loss: 0.5002
- val_accuracy: 0.7821
Epoch 109/200
4548/4548 [===
             :==========] - 1s 116us/step - loss: 0.4668 - accuracy: 0.7726 - val loss: 0.4967
val_accuracy: 0.7838
Epoch 110/200
4548/4548 [==
              ========] - 1s 114us/step - loss: 0.4661 - accuracy: 0.7715 - val loss: 0.4931
val_accuracy: 0.7733
Epoch 111/200
4548/4548 [===
            - val accuracy: 0.7900
Epoch 112/200
4548/4548 [===
              ========] - 1s 116us/step - loss: 0.4608 - accuracy: 0.7720 - val loss: 0.4914
- val accuracy: 0.7847
Epoch 113/200
val_accuracy: 0.7786
Epoch 114/200
val_accuracy: 0.7777
Epoch 115/200
- val_accuracy: 0.7882
Epoch 116/200
- val_accuracy: 0.7856
Epoch 117/200
- val_accuracy: 0.7935
Epoch 118/200
4548/4548 [===
              :=========] - 1s 110us/step - loss: 0.4641 - accuracy: 0.7661 - val loss: 0.4912
val_accuracy: 0.7891
Epoch 119/200
val accuracy: 0.7794
Epoch 120/200
4548/4548 [===
             =========] - 1s 111us/step - loss: 0.4663 - accuracy: 0.7773 - val_loss: 0.4928
val_accuracy: 0.7856
Epoch 121/200
- val_accuracy: 0.7900
Epoch 122/200
4548/4548 [===
             ==========] - 1s 111us/step - loss: 0.4595 - accuracy: 0.7773 - val loss: 0.4951
val_accuracy: 0.7821
Epoch 123/200
- val accuracy: 0.7830
Epoch 124/200
4548/4548 [====
          val_accuracy: 0.7873
Epoch 125/200
4548/4548 [============== - ] - 1s 112us/step - loss: 0.4567 - accuracy: 0.7792 - val loss: 0.4933
- val accuracy: 0.7882
Epoch 126/200
val_accuracy: 0.7865
Epoch 127/200
4548/4548 [====
       - val accuracy: 0.7821
Epoch 128/200
- val_accuracy: 0.7750
Epoch 129/200
4548/4548 [===
               =======] - 0s 107us/step - loss: 0.4589 - accuracy: 0.7735 - val loss: 0.5032
val accuracy: 0.7777
Epoch 130/200
val_accuracy: 0.7812
Epoch 131/200
4548/4548 [===
              :========] - 1s 112us/step - loss: 0.4544 - accuracy: 0.7715 - val loss: 0.5014
val_accuracy: 0.7838
Epoch 132/200
- val_accuracy: 0.7830
Epoch 133/200
4548/4548 [=================== ] - 0s 109us/step - loss: 0.4579 - accuracy: 0.7759 - val loss: 0.5037
val accuracy: 0.7812
Epoch 134/200
```

```
- val_accuracy: 0.7900
Epoch 135/200
4548/4548 [==
               =======] - 1s 111us/step - loss: 0.4509 - accuracy: 0.7817 - val loss: 0.5058
val accuracy: 0.7821
Epoch 136/200
val_accuracy: 0.7847
Epoch 137/200
4548/4548 [===
                =======] - 1s 110us/step - loss: 0.4516 - accuracy: 0.7777 - val_loss: 0.5084
val_accuracy: 0.7873
Epoch 138/200
- val_accuracy: 0.7794
Epoch 139/200
- val_accuracy: 0.7821
Epoch 140/200
- val_accuracy: 0.7873
Epoch 141/200
- val_accuracy: 0.7900
Epoch 142/200
4548/4548 [============== - ] - 1s 111us/step - loss: 0.4559 - accuracy: 0.7744 - val loss: 0.5079
val accuracy: 0.7909
Epoch 143/200
4548/4548 [===
             =========] - 1s 111us/step - loss: 0.4546 - accuracy: 0.7801 - val_loss: 0.5080
- val_accuracy: 0.7838
Epoch 144/200
4548/4548 [===
         - val accuracy: 0.7794
Epoch 145/200
4548/4548 [====
       val accuracy: 0.7794
Epoch 146/200
4548/4548 [==
                   :====l - 0s 109us/step - loss: 0.4522 - accuracy: 0.7762 - val loss: 0.5099
- val accuracy: 0.7838
Epoch 147/200
4548/4548 [====
       val_accuracy: 0.7821
Epoch 148/200
4548/4548 [===
         - val_accuracy: 0.7698
Epoch 149/200
- val accuracy: 0.7900
Epoch 150/200
4548/4548 [=================== ] - 0s 108us/step - loss: 0.4544 - accuracy: 0.7764 - val loss: 0.5137
- val accuracy: 0.7821
Epoch 151/200
val_accuracy: 0.7873
Epoch 152/200
val_accuracy: 0.7768
Epoch 153/200
4548/4548 [======
          - val_accuracy: 0.7794
Epoch 154/200
4548/4548 [===
             =========] - 1s 112us/step - loss: 0.4487 - accuracy: 0.7792 - val loss: 0.5147
- val_accuracy: 0.7768
Epoch 155/200
4548/4548 [====
         - val_accuracy: 0.7882
Epoch 156/200
4548/4548 [===
              :========] - 1s 111us/step - loss: 0.4488 - accuracy: 0.7810 - val loss: 0.5185
val accuracy: 0.7777
Fnoch 157/200
val_accuracy: 0.7698
Epoch 158/200
- val_accuracy: 0.7777
Epoch 159/200
4548/4548 [============== - ] - 1s 112us/step - loss: 0.4459 - accuracy: 0.7819 - val loss: 0.5300
- val accuracy: 0.7759
Epoch 160/200
val_accuracy: 0.7786
Epoch 161/200
4548/4548 [============== - ] - 1s 111us/step - loss: 0.4509 - accuracy: 0.7799 - val loss: 0.5211
- val_accuracy: 0.7742
Epoch 162/200
4548/4548 [===
             val_accuracy: 0.7812
Epoch 163/200
4548/4548 [===
           val accuracy: 0.7794
```

Epoch 164/200

```
- val accuracy: 0.7786
Epoch 165/200
4548/4548 [==
              ========] - 0s 108us/step - loss: 0.4369 - accuracy: 0.7889 - val loss: 0.5422
- val_accuracy: 0.7838
Epoch 166/200
4548/4548 [====
         - val_accuracy: 0.7759
Epoch 167/200
4548/4548 [===
            =========] - 0s 106us/step - loss: 0.4413 - accuracy: 0.7834 - val loss: 0.5268
- val_accuracy: 0.7777
Epoch 168/200
val accuracy: 0.7786
Fnoch 169/200
val accuracy: 0.7750
Epoch 170/200
val_accuracy: 0.7873
Epoch 171/200
4548/4548 [=================== ] - 0s 107us/step - loss: 0.4402 - accuracy: 0.7834 - val loss: 0.5319
val_accuracy: 0.7715
Epoch 172/200
- val_accuracy: 0.7750
Epoch 173/200
val_accuracy: 0.7803
Epoch 174/200
4548/4548 [===
        - val accuracy: 0.7821
Epoch 175/200
4548/4548 [===
           ==========] - 1s 113us/step - loss: 0.4474 - accuracy: 0.7902 - val_loss: 0.5339
val_accuracy: 0.7838
Epoch 176/200
4548/4548 [====
       val_accuracy: 0.7803
Epoch 177/200
4548/4548 [===
            =========] - 0s 109us/step - loss: 0.4459 - accuracy: 0.7858 - val_loss: 0.5259
val_accuracy: 0.7750
Epoch 178/200
- val_accuracy: 0.7838
Epoch 179/200
- val_accuracy: 0.7794
Epoch 180/200
- val_accuracy: 0.7759
Epoch 181/200
- val_accuracy: 0.7865
Epoch 182/200
- val accuracy: 0.7742
Epoch 183/200
val_accuracy: 0.7838
Epoch 184/200
4548/4548 [===
            val_accuracy: 0.7794
Epoch 185/200
- val_accuracy: 0.7733
Epoch 186/200
4548/4548 [==
                =====] - 0s 110us/step - loss: 0.4374 - accuracy: 0.7872 - val loss: 0.5437
- val_accuracy: 0.7742
Epoch 187/200
4548/4548 [===
           :=========] - 0s 106us/step - loss: 0.4363 - accuracy: 0.7856 - val loss: 0.5491
val_accuracy: 0.7953
Epoch 188/200
- val accuracy: 0.7689
Epoch 189/200
val accuracy: 0.7742
Epoch 190/200
4548/4548 [=================== ] - 0s 106us/step - loss: 0.4469 - accuracy: 0.7828 - val loss: 0.5134
- val accuracy: 0.7759
Epoch 191/200
- val_accuracy: 0.7777
Epoch 192/200
val accuracy: 0.7786
Epoch 193/200
4548/4548 [==:
            ========] - 0s 107us/step - loss: 0.4367 - accuracy: 0.7874 - val loss: 0.5397
```

- val_accuracy: 0.7750

```
Epoch 194/200
                :=========] - 0s 107us/step - loss: 0.4450 - accuracy: 0.7790 - val_loss: 0.5438
    4548/4548 [===
    val_accuracy: 0.7671
    Epoch 195/200
    - val_accuracy: 0.7768
    Epoch 196/200
    - val_accuracy: 0.7812
    Epoch 197/200
    - val accuracy: 0.7847
    Epoch 198/200
    - val accuracy: 0.7856
    Epoch 199/200
    - val accuracy: 0.7724
    Epoch 200/200
                4548/4548 [=======
    val_accuracy: 0.7750
    534/534 [====
                Out[31]: [0.5510589545139213, 0.7640449404716492]
In [32]: classes2 list = classes.tolist()
    model keras ANN = []
    for i in classes2 list:
      model keras ANN.append(round(i[0]))
```

모델 검증

```
In [33]: #report
         from sklearn.metrics import classification_report
         print(classification report(Y test, Y pred ANN, digits=4))
                       precision
                                     recall f1-score support
                                     0.7000
                                               0.7490
                    0
                           0.8053
                          0.7468
                                     0.8394
                                               0.7904
                                                            274
                    1
                                               0.7715
                                                            534
             accuracy
                                     0.7697
                          0.7760
                                               0.7697
                                                            534
            macro avq
                                    0.7715
                                               0.7702
         weighted avg
                          0.7753
                                                            534
```

앙상블모델 만들기

```
In [34]:
         model knn = knn.predict(X test)
         model_decision_tree = decision_tree.predict(X_test)
         model_gaussian = gaussian.predict(X_test)
         model_logreg = logreg.predict(X_test)
         model_svc = svc.predict(X_test)
         model ANN = ANN.predict(X test)
         # model keras ANN
In [35]: score = []
         # for i in range(0,962):
         for i in range(0,534):
             list = []
             #list.append(model knn[i])
             #list.append(model decision tree[i])
             #list.append(model_gaussian[i])
             #list.append(model_svc[i])
             list.append(model_logreg[i])
             list.append(model_ANN[i])
             list.append(model keras ANN[i])
             if list.count(1) > list.count(0):
                 score.append(1)
             else:
                 score.append(0)
         print(Y_test)
```

```
1
2
                1
                 1
          3
          4
                0
          683
                0
          684
                0
          685
                1
          686
                1
          687
         Name: RESULT, Length: 534, dtype: int64
In [36]: ensemble = pd.DataFrame({
                  "result": Y_test,
"prediction_ensemble": score
          ensemble['RESULT'] = np.where(ensemble["result"]==ensemble["prediction_ensemble"], 1, 0)
In [37]: #앙상블모델 정확도
          acc_Ens = round(((sum(ensemble['RESULT'])/534))*100, 2)
          acc_Ens
Out[37]: 77.72
In [38]: print(classification_report(Y_test,score))
                        precision
                                     recall f1-score
                                                         support
                     0
                             0.81
                                       0.70
                                                 0.75
                                                             260
                     1
                             0.75
                                       0.85
                                                 0.80
                                                             274
                                                             534
                                                 0.78
              accuracy
                             0.78
                                       0.78
             macro avg
                                                 0.78
                                                             534
         weighted avg
                             0.78
                                       0.78
                                                 0.78
                                                             534
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js