

PE 프로그램 진행과정 (8.5)

01

사용 데이터별 결과비교

- 2021 data 사용

- 2020 ~ 2021 data 사용

- 2019 ~ 2021 data 사용

- 2018 ~ 2021 data 사용

2021년 데이터 사용

```
Epoch 96/100
146/146 [=====] - 1s 5ms/step - loss: 0.5988 - accuracy: 0.6720
Epoch 97/100
146/146 [=====] - 1s 5ms/step - loss: 0.6113 - accuracy: 0.6614
Epoch 98/100
146/146 [=====] - 1s 5ms/step - loss: 0.6015 - accuracy: 0.6706
Epoch 99/100
146/146 [=====] - 1s 5ms/step - loss: 0.6118 - accuracy: 0.6685
Epoch 100/100
146/146 [=====] - 1s 5ms/step - loss: 0.6010 - accuracy: 0.6758
```

```
: loss_and_metrics = model.evaluate(X_test, y_test, batch_size = 32)
37/37 [=====] - 2s 4ms/step - loss: 0.7810 - accuracy: 0.6177
```

데이터의 수 : 5820

사용한 피쳐 개수 : 9

acc = 0.6177

2020~2021년 데이터 사용

```
Epoch 96/100
421/421 [=====] - 2s 5ms/step - loss: 0.5935 - accuracy: 0.6819
Epoch 97/100
421/421 [=====] - 2s 5ms/step - loss: 0.5958 - accuracy: 0.6789
Epoch 98/100
421/421 [=====] - 2s 5ms/step - loss: 0.5993 - accuracy: 0.6742
Epoch 99/100
421/421 [=====] - 2s 5ms/step - loss: 0.5972 - accuracy: 0.6802
Epoch 100/100
421/421 [=====] - 2s 6ms/step - loss: 0.5970 - accuracy: 0.6737
```

```
] : loss_and_metrics = model.evaluate(X_test, y_test, batch_size = 64)
53/53 [=====] - 1s 2ms/step - loss: 0.6004 - accuracy: 0.6826
```

데이터의 수 : 16808

사용한 피쳐 개수 : 12

acc = 0.6826

PE 프로그램 진행과정 (8.5)

02

사용 데이터별 결과비교

- 2021 data 사용
- 2020 ~ 2021 data 사용
- 2019 ~ 2021 data 사용
- 2018 ~ 2021 data 사용

2019 ~ 2021년 데이터 사용

```
Epoch 97/100
696/696 [=====] - 3s 5ms/step - loss: 0.6008 - accuracy: 0.6686
Epoch 98/100
696/696 [=====] - 4s 5ms/step - loss: 0.5977 - accuracy: 0.6716
Epoch 99/100
696/696 [=====] - 4s 5ms/step - loss: 0.5974 - accuracy: 0.6733
Epoch 100/100
696/696 [=====] - 3s 5ms/step - loss: 0.6051 - accuracy: 0.6697
```

```
loss_and_metrics = model.evaluate(X_test, y_test, batch_size = 32)
```

```
174/174 [=====] - 0s 2ms/step - loss: 0.6429 - accuracy: 0.6635
```

데이터의 수 : 27802

사용한 피쳐 개수 : 12

acc = 0.6635

2018 ~ 2021년 데이터 사용

```
Epoch 97/100
978/978 [=====] - 5s 5ms/step - loss: 0.5925 - accuracy: 0.6773
Epoch 98/100
978/978 [=====] - 5s 5ms/step - loss: 0.5896 - accuracy: 0.6818
Epoch 99/100
978/978 [=====] - 5s 5ms/step - loss: 0.5914 - accuracy: 0.6831
Epoch 100/100
978/978 [=====] - 5s 5ms/step - loss: 0.5966 - accuracy: 0.6750
```

```
loss_and_metrics = model.evaluate(X_test, y_test, batch_size = 32)
```

```
245/245 [=====] - 1s 2ms/step - loss: 0.6018 - accuracy: 0.6734
```

데이터의 수 : 39118

사용한 피쳐 개수 : 12

acc = 0.6734

데이터의 수가 증가함에도
정확도측면에서는 큰 차이가 없음

PE 프로그램 진행과정 (8.5)

03

불필요한 데이터 줄이기

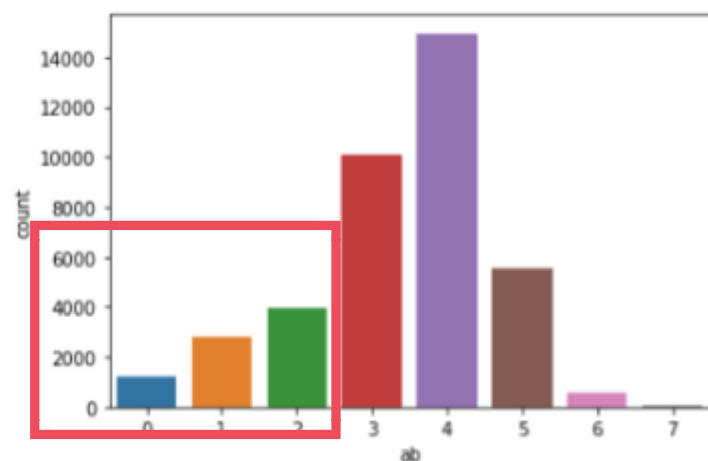
- 2018 ~ 2021 data 사용

- 타석 수 > 1

- 타석 수 > 2

```
sns.countplot(final_2018_2021['ab'])  
plt.show()
```

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators
FutureWarning



최소 2타석 이상의 기록들만 사용

```
Epoch 97/100  
879/879 [=====] - 4s 5ms/step - loss: 0.5692 - accuracy: 0.7131  
Epoch 98/100  
879/879 [=====] - 4s 5ms/step - loss: 0.5673 - accuracy: 0.7151  
Epoch 99/100  
879/879 [=====] - 4s 5ms/step - loss: 0.5727 - accuracy: 0.7084  
Epoch 100/100  
879/879 [=====] - 4s 5ms/step - loss: 0.5658 - accuracy: 0.7159
```

```
loss_and_metrics = model.evaluate(X_test, y_test, batch_size = 32)  
220/220 [=====] - 1s 2ms/step - loss: 0.5740 - accuracy: 0.7082
```

데이터의 수 : 35125

사용한 피쳐 개수 : 12

acc = 0.7082

최소 3타석 이상의 기록들만 사용

```
Epoch 97/100  
780/780 [=====] - 4s 5ms/step - loss: 0.5436 - accuracy: 0.7398  
Epoch 98/100  
780/780 [=====] - 4s 5ms/step - loss: 0.5460 - accuracy: 0.7352  
Epoch 99/100  
780/780 [=====] - 4s 5ms/step - loss: 0.5477 - accuracy: 0.7388  
Epoch 100/100  
780/780 [=====] - 4s 5ms/step - loss: 0.5363 - accuracy: 0.7458
```

```
loss_and_metrics = model.evaluate(X_test, y_test, batch_size = 32)  
195/195 [=====] - 1s 2ms/step - loss: 0.5463 - accuracy: 0.7422
```

데이터의 수 : 31170

사용한 피쳐 개수 : 12

acc = 0.7422

정확도가 기존보다 10%가량 향상된 것 처럼

보이지만,

1의 비율이 71%의 데이터임을 감안하면

성능 Bad



PE 프로그램 진행과정 (8.5)

04

불필요한 데이터 줄이기

- 2018 ~ 2021 data 사용
- 최소 타석 수 제한
- 선발 출장한 선수 기록만 사용
- 시즌 초 경기 제외(3월, 4월)

선발 출장한 선수 기록만 사용

```
Epoch 97/100
877/877 [=====] - 4s 5ms/step - loss: 0.5747 - accuracy: 0.7046
Epoch 98/100
877/877 [=====] - 4s 5ms/step - loss: 0.5702 - accuracy: 0.7061
Epoch 99/100
877/877 [=====] - 4s 5ms/step - loss: 0.5675 - accuracy: 0.7094
Epoch 100/100
877/877 [=====] - 4s 5ms/step - loss: 0.5718 - accuracy: 0.7059
```

```
loss_and_metrics = model.evaluate(X_test, y_test, batch_size = 32)
```

```
220/220 [=====] - 1s 2ms/step - loss: 0.5774 - accuracy: 0.7041
```

데이터의 수 : 35049
사용한 피쳐 개수 : 12

acc = 0.7041

선발 출장 + 시즌 초 경기 제외

```
Epoch 97/100
737/737 [=====] - 4s 5ms/step - loss: 0.5960 - accuracy: 0.6916
Epoch 98/100
737/737 [=====] - 4s 5ms/step - loss: 0.5988 - accuracy: 0.6890
Epoch 99/100
737/737 [=====] - 4s 5ms/step - loss: 0.6029 - accuracy: 0.6844
Epoch 100/100
737/737 [=====] - 4s 5ms/step - loss: 0.6021 - accuracy: 0.6850
```

```
loss_and_metrics = model.evaluate(X_test, y_test, batch_size = 32)
```

```
185/185 [=====] - 1s 2ms/step - loss: 0.6028 - accuracy: 0.6818
```

데이터의 수 : 29461
사용한 피쳐 개수 : 12

acc = 0.6818

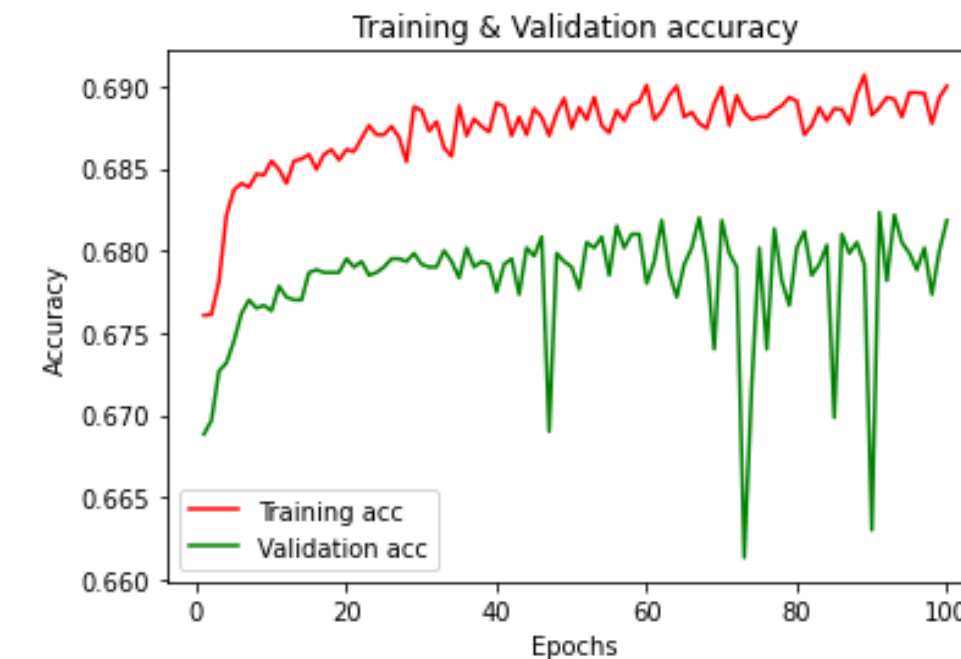
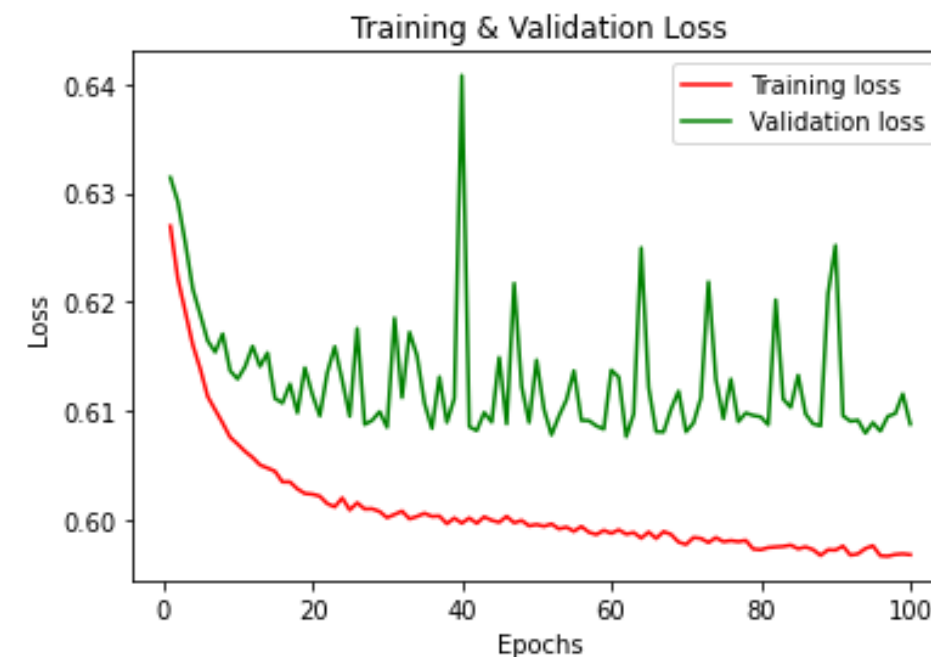
정확도가 조금씩 향상되는 모습

PE 프로그램 진행과정 (8.5)

05

Validation data set 구성

- 데이터 과적합 확인
- training / validation 정확도



Training data는 갈수록
loss가 감소하고 accuracy 증가하지만,
Validation data는 그래프가 상당히 불안정

PE 프로그램 진행과정 (8.5)

06

MLP model 개선

- 피쳐 개수 증가
- 활성화 함수 수정
- layer 간략화

MLP Model

```
: #Activation = Relu & Sigmoid
model = Sequential()

model.add(Dense(32, input_dim= width, kernel_initializer = 'glorot_uniform', activation = 'relu'))

# model.add(Dense(96, kernel_initializer = 'glorot_uniform', activation = 'relu'))
model.add(Dense(48, activation = 'relu'))
model.add(Dense(2, activation = 'sigmoid'))
```

Model training

```
model.compile(loss = 'binary_crossentropy', optimizer = 'rmsprop', metrics=['acc'])

history = model.fit(X_train, y_train, epochs = 200, batch_size = 64, validation_data=(X_val, y_val))
```

마지막 층 활성화 함수 : softmax > sigmoid

loss 함수 : categorical > binary crossentropy

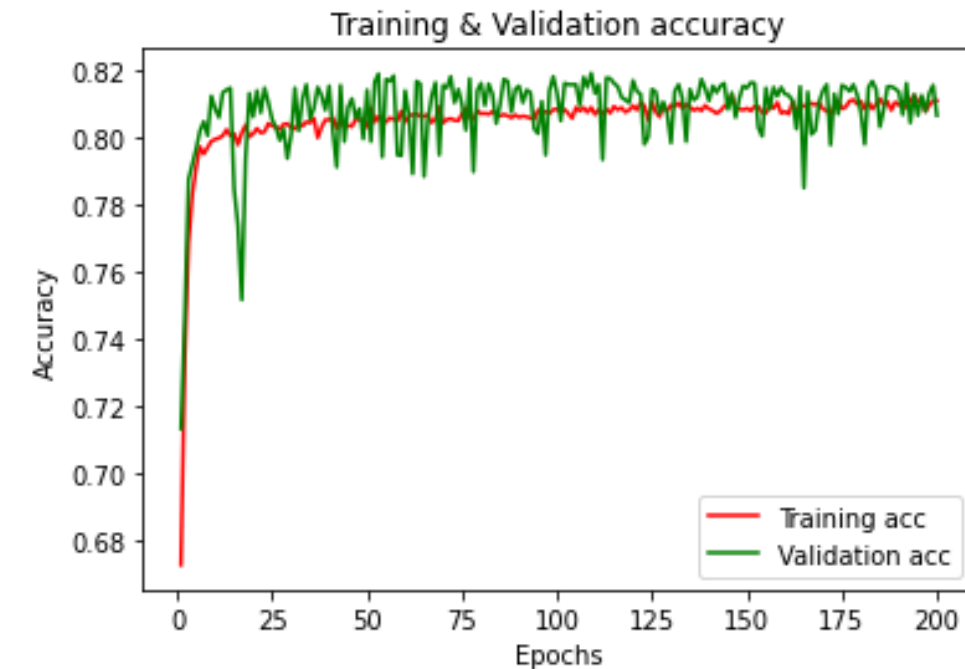
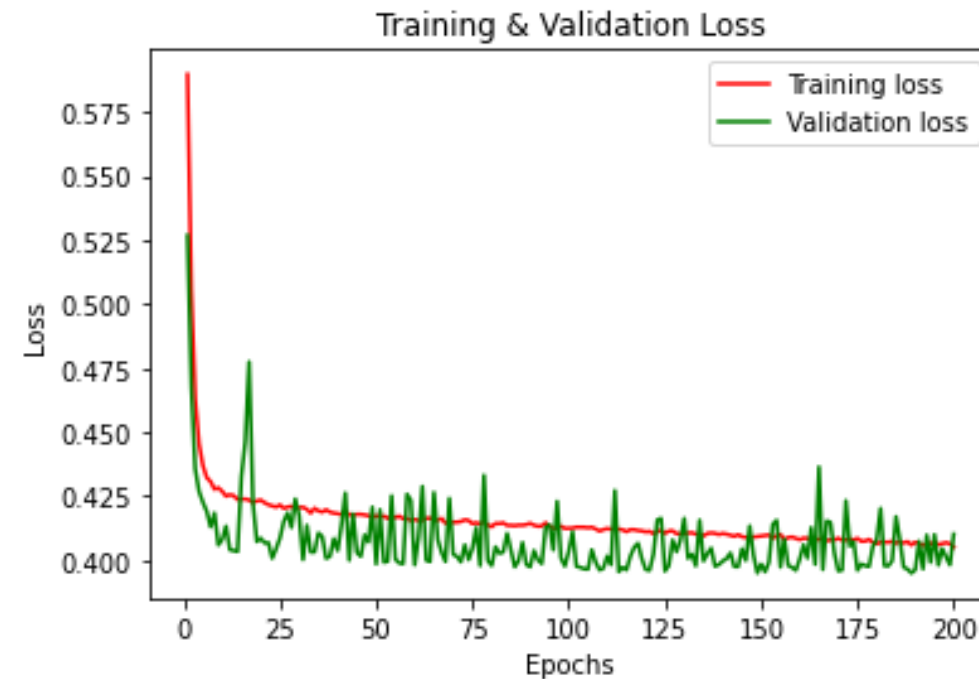
옵티마이저 : rmsprop

PE 프로그램 진행과정 (8.5)

07

과적합 및 결과 확인

- 데이터 과적합 확인
- training / validation 정확도



```
Epoch 197/200
244/244 [=====] - 0s 2ms/step - loss: 0.4065 - acc: 0.8107 - val_loss: 0.4046 - val_acc: 0.8080
Epoch 198/200
244/244 [=====] - 0s 2ms/step - loss: 0.4021 - acc: 0.8120 - val_loss: 0.4015 - val_acc: 0.8134
Epoch 199/200
244/244 [=====] - 1s 2ms/step - loss: 0.4062 - acc: 0.8098 - val_loss: 0.3982 - val_acc: 0.8158
Epoch 200/200
244/244 [=====] - 1s 2ms/step - loss: 0.4086 - acc: 0.8082 - val_loss: 0.4102 - val_acc: 0.8068
```

```
#Latest try: Activation Relu & Sigmoid
loss_and_metrics = model.evaluate(X_test, y_test, batch_size = 64)
138/138 [=====] - 0s 1ms/step - loss: 0.4036 - acc: 0.8127
```

데이터의 수 : 29395
사용한 피쳐 개수 : 13
acc = 0.8127