# **결론**

SHAP importance로 확인한 결과

**Force**-Grade I을 예측하는 데는 Grade\_I이 예측 값을 높이는데 긍정적 영향, CA125는 부정적 영향  
(CA125가 좀 더 적은 곳에 Grade I이 분포되어 있지만 크게 분포 차이가 없기 때문에 부정적 영향을 미치는 것으로 보임.)  
-Grade II를 예측하는 데는 Grade\_I, Age, CA125가 부정적 영향, 폐경\_No, Less than이 조금 긍정적 영향  
-others를 예측하는 데는 CA125, Age가 긍정적 영향, Grade\_I, Less than 50%가 부정적 영향

**로컬 막대 플롯(Waterfall Plot) / 전역 막대 / Summary  
\*예측하는 대상**-Grade I: Grade I이 가장 큰 긍정적 영향을 미침. (순서대로 Age, Less than50%, CA125, Method\_Diliation / curettage)  
-Grade II: Grade II가 가장 큰 부정적 영향을 미침. (순서대로 Age, CA125, Less than 50%, 폐경\_No, Method\_Diliation/curettage)  
-others: Age가 긍정적 영향, Grade\_I은 부정적 영향, CA125는 긍정적 영향 등

## **데이터 처리 및 분석**

Grade column은 I, II, others로 구분

Stage column은 Ia, Ib로 통일.

Postoperative에서 Group내에 존재하는 인원 파악  
-Group 1~4에 해당하는 인원은 228명, 나머지는 24명

Preoperative의 침습 깊이와 Postoperative의 침습 깊이 일치도 비교  
-문제되는 부분은 전체의 약 5.15% 오차율 존재

Preoperative의 Grade와 Postoperative의 Grade 일치도 비교  
-문제되는 부분은 전체의 약 4.36% 오차율 존재.

Grade의 분포도 확인

Stage의 분포도 확인.  
- 데이터 불균형 문제 존재.

Grade에 대한 수치형 범주형에 따른 특성 및 분포도 확인(countplot)

Violin을 이용한 label\_Grade, Age, Menopause에 따른 관계 확인  
Violin을 이용한 label\_Grade, CA125, Menopause에 따른 관계 확인  
Violin을 이용한 label\_Grade, Tumor size, Menopause에 따른 관계 확인

## **모델 평가 및 시도**

평가 지표: F1 socre의 macro average

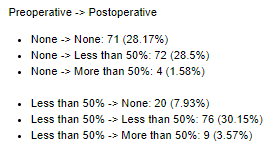
XGBoost의 하이퍼 파라미터 설정  
 - estimator, max\_depth, 방법 – Random Search

Random oversampling 적용 (data imbalance 문제 해결 위해)  
- SMOTE, Random oversampling

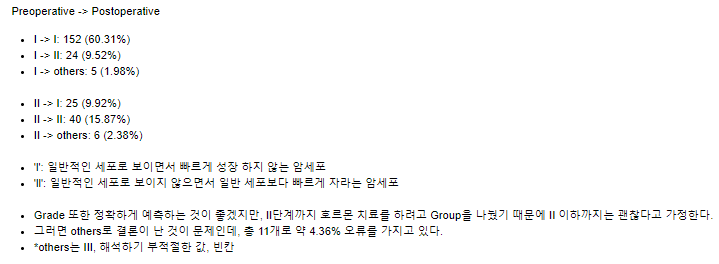
Class weight를 조절  
- pos\_weight

SHAP importance 적용하여 분석

## **요청사항**

1. 우선 label data에서의 Stage column은 Ia에 너무 편향되어 있는 문제가 존재.  
   -> Ib에 대한 데이터가 추가적으로 요구됨.
2. 침습 깊이 및 Grade를 Group 4개로 나누는 것보다 이 4개의 그룹을 1개로 묶는 것은 어떻게 생각하시는지?? 이 4개의 Group이 호르몬 치료를 받아도 괜찮다면 그룹을 1개로 묶는 것이 더 좋아보임.  
     
   침습 깊이의 confusion\_matrix를 확인해보면,  
     
   -> 제대로 맞추는 것은 28.17 + 30.15 = 약 58.32%  
   그럼, 제대로 못맞춘다는 경우는 약 41.68%라는 것. -> 오차율이 너무 크다.  
   하지만, Less than 50%까지 괜찮다는 가정하에 그룹을 묶게 되면 일치도가 높아짐.  
   이를 제외한 수술 후 More than 50%인 심각한 bad case인 경우는 약 5.15%이다.  
   But, 이렇게 나뉘었을 때도 문제가 발생할 수 있다. -> data imbalance  
   -> 이를 해결하기 위한 More than 50% case 데이터가 좀 더 필요하다.  
    Q. 수술 전 More than 50%라고 판명된 것은 무조건 수술 후 More than 50% 인지?

Grade 또한 같은 맥락. (confusion\_matrix)

  
-> 제대로 맞추는 것은 60.31 + 15.87 = 약 76.18%  
이 케이스도 마찬가지로 I, II을 한 개의 case로 볼 지, 아니면 I, II를 따로 볼지 고민이 필요함.   
-> 이는 Grade I, II가 호르몬 치료를 감행하는 데 얼마나 영향을 미치고 위험한지에 따라서 하나로 묶을지 아니면 따로 꼭 분리해서 볼 지 결정해야할 것으로 보임.  
  
침습 깊이는 정확히 맞추는 것이 좀 낮긴 때문에 되도록 None과 Less than 50%을 하나의 case로 보는 것이 좋아 보이고 아니면 범위를 50%가 아닌 예를 들어 호르몬 치료가 가능한 안정적인 범위인 30%로 줄여서 확실히 호르몬 치료가 가능한 case들을 크게 묶는 것이 중요해 보입니다.  
  
Grade도 마찬가지로 I, II을 괜찮다면 하나의 case로 보면 좋을 듯합니다.  
-> 대신 others한 case가 좀 더 필요해 보입니다.

4개의 Group들을 하나의 Group으로 묶던가 새로운 Group으로 재정의 해서 묶는 것이 필요해 보입니다.

**Q.**

1. 수술 전 Grade와 침습 깊이(stage) 결과를 얻을 때, 정량적인 평가 기준이 있는지? 아니면 의사마다 개인 차이가 있는지?  
-> 의사가 개인적으로 cell type, region을 보고 판단하는지?  
자궁내막조직검사 및 MRI로 Grade와 Stage를 예상.

2. Method에서 방식은 어떤 경우에 따라 방식을 달리 하는지??

3. 수술 전 Grade와 수술 후 Grade가 달라지는 이유는?  
수술 전 Grade는 자궁내막조직 검사로 나온 결과이고, 수술 후 Grade는 Route로 들어가서 정확한 암 조직을 떼서 확인한 결과인지?

4. 수술 전 종양 크기가 작고, 수술 후 종양 크기가 큰 경우는 MRI상으로 잘 보이지 않아서 그럴 수 있다고 생각을 하는데,  
 수술 전 종양 크기가 크고, 수술 후 종양 크기가 작은 이유는 어떤 이유 때문에 그런 결과가 나오는지?

5. 침습 깊이와 종양 크기는 어떤 관계를 가지는지? (별개의 문제인지?)  
수술 후 침습 깊이와 종양 크기를 보면, 종양 크기가 0인데도 Less than 50%을 가지는 것이 있고, 심지어 More than 50%를 가지는 환자(184)가 있다. 왜 이런 경우가 생기는지??

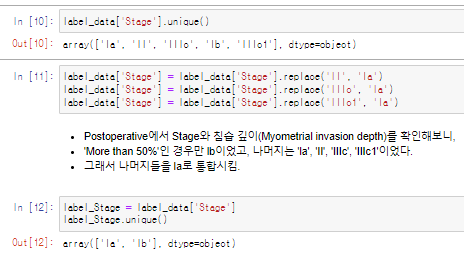
# **데이터 처리 및 분석**

## Grade (Postoperative) columns



* Grade는 I, II, others로 구분.

## Stage columns

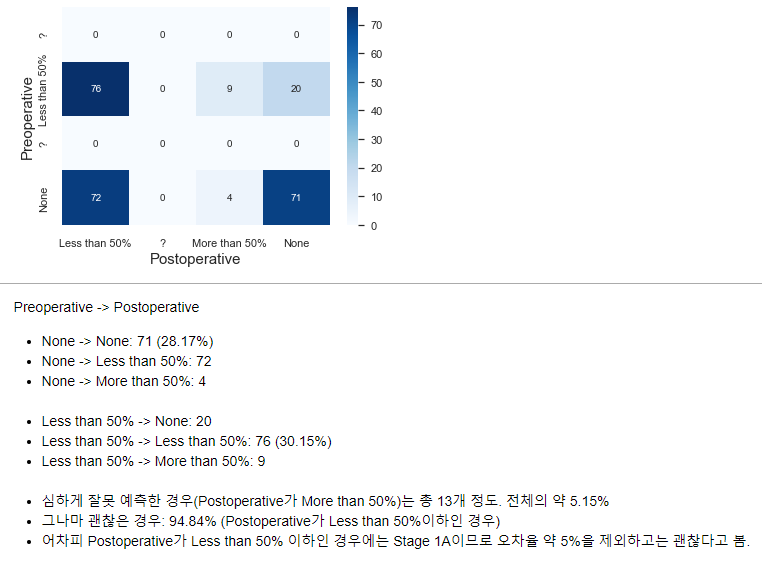


* 1a, 1b로 통합시킴.

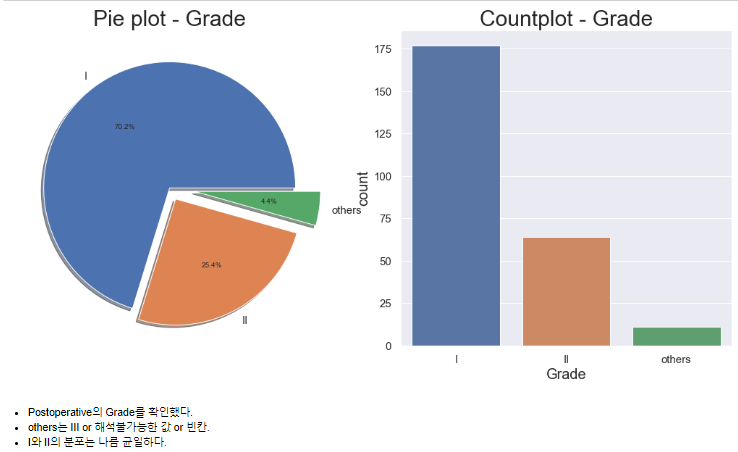
## 3. Postoperative에서의 Group 나누기.

Group 1: 68 / Group 2: 102 / Group 3: 16 / Group 4: 42  
전체 환자 수 : 252명, Group 1~4: 228, 나머지: 24(Grade='others' 이거나 침습 깊이가 More than 50%일 때)

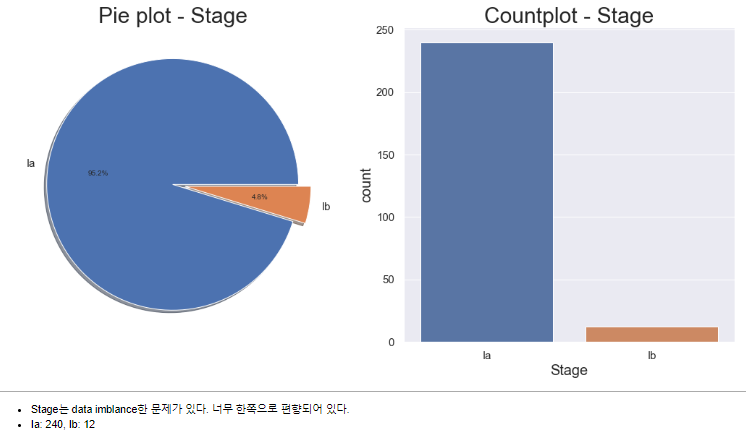
## 4. Preoperative의 침습 깊이와 Postoperative의 침습 깊이 일치도 비교 (confusion\_matrix)



## 5. Grade의 분포도 확인



## 6. Stage의 분포도 확인



* Stage는 데이터 불균형 문제 존재. (Ib의 데이터가 부족하다.)

## 5. Preoperative의 Grade와 Postoperative의 Grade 일치도 비교 (confusion\_matrix)

