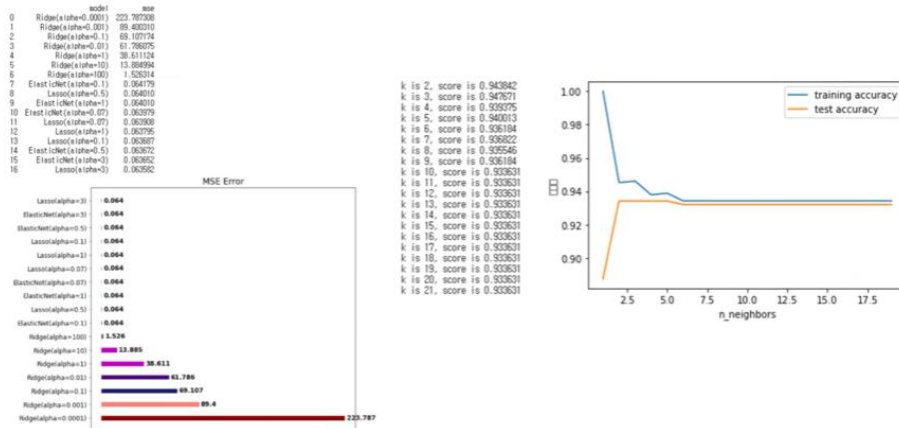


2조,3조,4조와의 비교 report

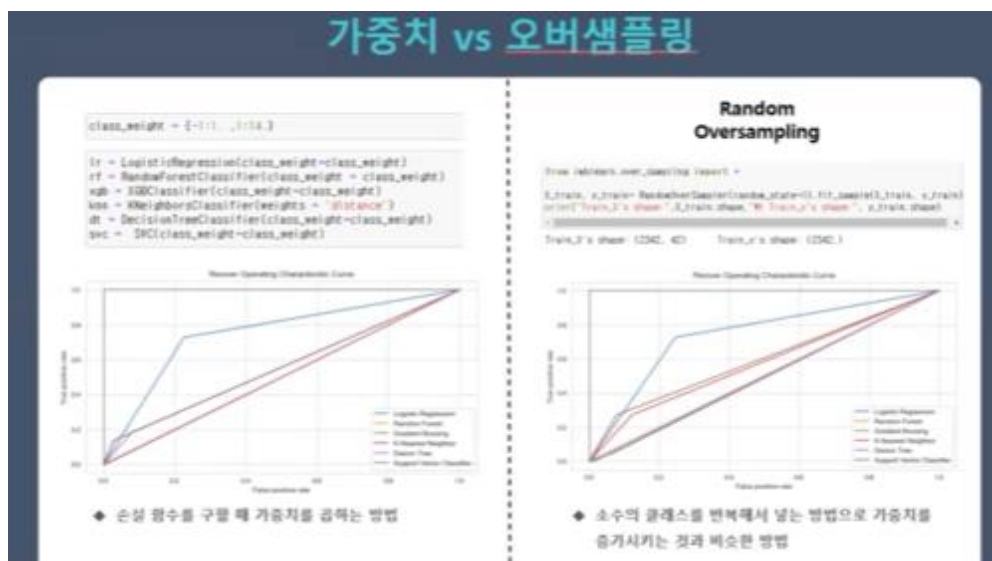
항태영

1. 시각화 자료

4. Model Fitting_Normal Data



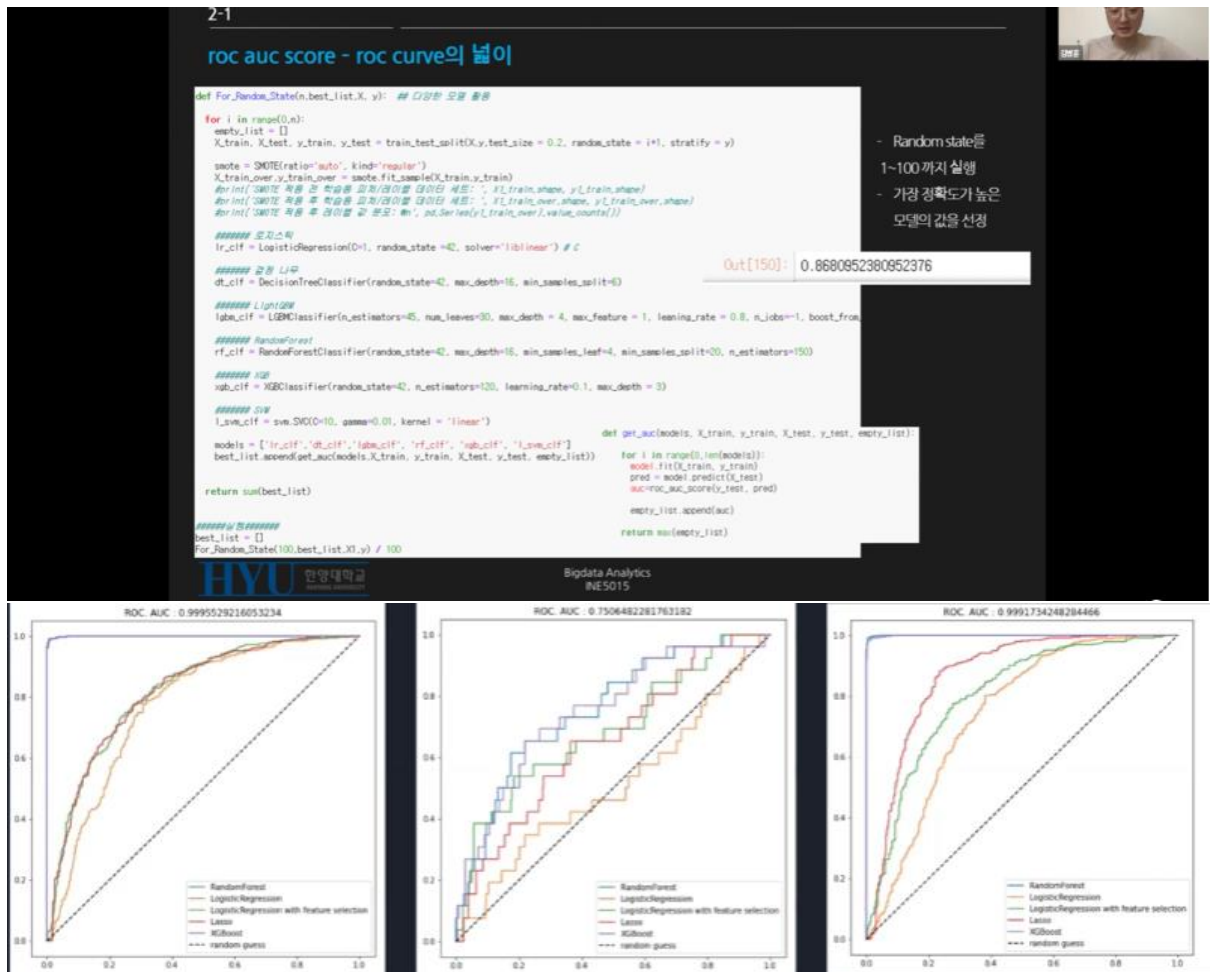
3조 - 모델 fitting 부분



4조 가중치 vs 오버샘플링 부분

단순히 코드를 통해 확인한 자료가 아닌 발표를 위한 시각화 자료의 필요성을 느꼈고, 이 부분은 교수님의 지적 사항이기도 하였다. 단, 2조와 같이 정리된 필요한 부분에서만 사용하는 방식 또한 좋다고 생각한다. 대신 이 방식에 있어서는 우리가 사용한 모델들의 이해가 충분히 이루어져 설명할 수 있어야 한다. 모델 이해에 대한 부분은 교수님이 여러 차례 다른 조의 발표에서 지적한 부분이기도 하다.

다음으로 아래 2조와 우리 조의 발표 자료를 비교해 보면



같은 roc curve에 대한 내용에서 우리의 발표 자료보다 2조의 발표자료가 한눈에 들어오고 청중 입장에서 발표한 내용의 이해해 도움이 되어 보인다.

2. 발표에 대한 결론(발표 평가 부분)

결론 및 계획

- 대체로 선형 모델보다 비선형 모델의 성능이 좋게 나왔다
 - pass와 fail의 결정 경계가 선형이 아닌 것 같다.
- 클래스 균형을 맞춰줬을때 대체로 성능이 좋게 나왔다
- Feature Scaling을 해줬을때 대체로 성능이 좋게 나왔다
- 모델 자체에 규제가 있거나 feature를 선택하는 효과가 있는게 더 좋은 성능을 낸다

다음단계로 각 모델에 대해서 교차검증을 통해 하이퍼파라미터를 맞추는 모델 튜닝을 할 것입니다.

9. 문제점 및 향후 계획

- VIF값이 높게 나옴 -> 다중공선성 해결을 위한 방법 생각
- 낮은 재현율 -> 정확도, F1-score값과 함께 개선할 수 있는 방법 생각
- PCA를 진행하여 모델링 한 결과 과적합 -> 다른 조합 생각

문제점 및 개선방안

문제점	개선방안 및 고려사항
<ul style="list-style-type: none"> • 소수 클래스인 1(불량물)에 대해 분류를 잘 못함 • 특징선택 과정에서 Threshold 기준을 정할 때 데이터에 과적합되는 경우와 클래스 불균형을 고려하지 않음 • 특징선택 과정에서 같은 auc값을 가질 때 적은 특징조합을 선택하였고 이로 인해 데이터의 손실이 발생하였을 수 있음 • Ensemble시 True 값을 중복으로 몇 개 이상 갖는 것을 선택할지 정할 때 적은 특징 조합에 많은 True를 선택함으로써 데이터의 손실이 발생하였을 수 있음 • 모델 분석 시 하이퍼파라미터 및 kfold를 이용한 <u>과적합</u> 방지를 하지 않음 	<ul style="list-style-type: none"> • 특징선택 과정에서 Cross Validation (ex kfold)을 이용하고 클래스 불균형을 고려하여 threshold를 구함 • 특징선택 과정에서 같은 성능을 가질 때 많은 특징의 조합을 선택하여 ensemble로 인해 데이터가 손실하는 부분을 개선 • 모델 분석 시 하이퍼파라미터 조정 및 kfold를 통한 <u>과적합</u> 방지 • Pca 사용 후 원래 데이터로 변환하여 결과 해석 시 영향을 미친 정확한 변수 파악 및 해석 • <u>결측치</u> 대체 비율 및 knn 이웃 수 비교분석을 통해 추가적으로 고려

각 2조 3조 4조와 같이 문제점 시사와 개선방안에 대한 내용을 마지막에 삽입하였다. 이 부분은 아래 표를 통해 알 수 있다.

평가 구분	평가 기준 (총점)	C	B	A
토론	문제 정의 (6)	문제정의를 모호하고 분석결과만 리포트함	문제 정의가 명확하고 데이터 해석을 성실하게 수행함	문제 정의가 명확하고 데이터 해석 후 시사점을 제시함
	자료 수집 (6)	기존 자료 수집만 수행함	기존 자료의 비교분석을 수행함	데이터 분석 결과와 연계하여 모델링 방향을 제시함
	자료 분석 (7)	코드를 실행 결과만 리포트	결과에 대한 해석을 성실하게 수행함	결과해석을 수행하고 문제점을 정확히 도출함
	개선방안 수립 (6)	개선 방안의 추진 방법이 명확하지 않음	추진방법은 명확하나 구체성이 없음	개선방안이 명확하고 구체적으로 제시되어 있음
프로젝트	최종 보고서 (15)	보고서의 형식과 내 용이 미흡함	문제 정의부터 개선방 안까지 전개가 명확하 지 않음	형식과 내용이 명확하 지만 창의성이 부족함
	발 표 (10)	발표자료가 장황하고, 슬라이드를 보면서 발표함	발표자료는 잘 준비하 였지만, 발표 연습이 부족함	발표가 아주 명쾌하며 시간을 잘 지킴

해당 평가표를 보았을 때 우리가 속한 점수인 B와 2~4조가 속한 A의 차이는 문제점과 개선방안에 대한 부분인 것으로 보인다. 우리와 같은 점수를 받은 7조와 비교를 해 보면 7조는 결과부분을 포함하여 발표하였다. 그러나, 발표 내용에 대한 지적으로써 B를 받은 것으로 보이므로 이번 발표에서 문제점과 개선방안에 대한 부분이 추가되었다면 더 높은 점수를 받을 수 있었을 거라고 생각한다.