착과량: 과수에서 열매가 열린 수 또는 무게를 나타내는 용어.

텍스트, 지도, 도표이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

수고: 나무의 높이

수관폭: 나무의 너비

새순: 나무나 풀의 줄기에서 새로 돋아나는 연한 잎이나 싹

엽록소: 식물의 광합성 작용을 하는 식물성 색소.

수고와 착과량의 관계: 낮은 수고는 양질의 착과량을 증대 시킴. 사과나무의 상부 쪽에 햇빛이 잘 들게 하여 발육이 좋아지고, 품질 좋은 착과를 유도할 수 있음.

수관폭과 착과량의 관계: 수관폭이 넓으면 광합성 효율이 높아지고 뿌리또한 넓어 영양분 흡수에 도움을 줌.

새순과 착과량의 관계: 새순과 착과량은 경쟁관계임. 새순이 많으면 양분 경합, 빛 가림 등으로 열매가 잘 비대하지 못함.

엽록소와 착과량의 관계: 엽록소는 광합성과 영양 공급을 하여 착과량을 늘릴 수 있음.

즉 위 특성들은 착과량과 연관이 있음.

데이터 확인 과정 후

EDA

데이터간 상관계수 확인. 새순과 엽록소는 빼고 수관폭간 상관관계 확인

각각의 상관관계는 모두 0.8 이상으로 다중공선성 의심. 수관폭간 컬럼 데이터가 비슷하여 이는 학습에 방해가 될것임.

하나씩 빼보면서 제일 좋은 특성 찾기.

새순은 9월부터 11월 중순까지는 상관관계가 음의 상관관계임.

엽록소는 전부 음의 상관관계인데 그 정도가 크지 않음.

아니면 각각 일주일 간격으로 데이터를 바꿔서 성능평가 해보기.

음의 상관관계 정도가 크더라도 모델에 중요한 정보가 될 수 있음.

지금까지 수관폭평균+ 염록소제거 조합이 제일 잘 나옴.

내일 일주일 간격으로 새순, 염록소 데이터 나눠서 수관폭평균 데이터 조합으로 해보기.

순서: 새순과 엽록소 데이터를 먼저 일주일 단위로 전처리 하고, 수관폭 관련 컬럼 3개를 각각 성능평가 해보기.

엽록소 데이터는 전부 음의 상관관계이지만, 그 크기가 크지 않았고, 새순 데이터는 절반이 음의 상관관계였지만, 전부 0.8이상의 강한 음의 상관관계였음.

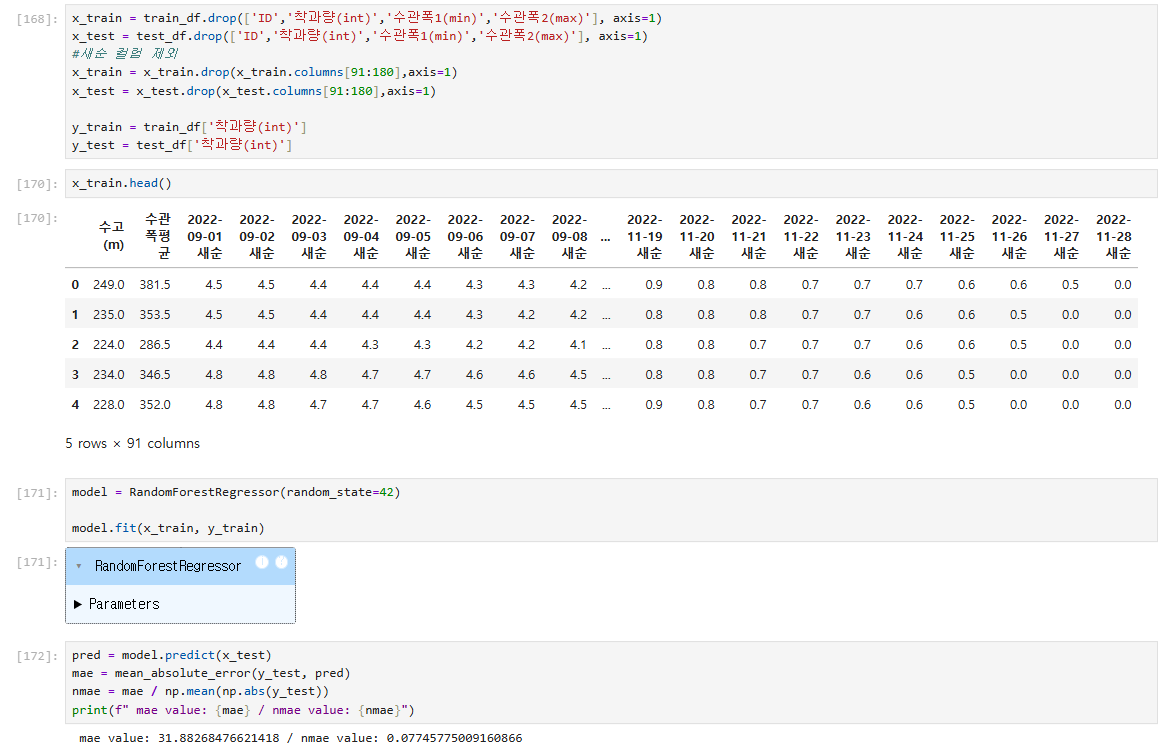
음의 상관관계여도 예측 모델에 중요한 정보가 될수있음. 하지만 비슷한 데이터가 너무 많고 이는 과적합이 생길수도있고 너무 많은 차원임.

새순 데이터와 엽록소 데이터를 일주일 단위로 전처리하면 차원 축소 효과와 과적합을 방지할 수 있다고 생각함.

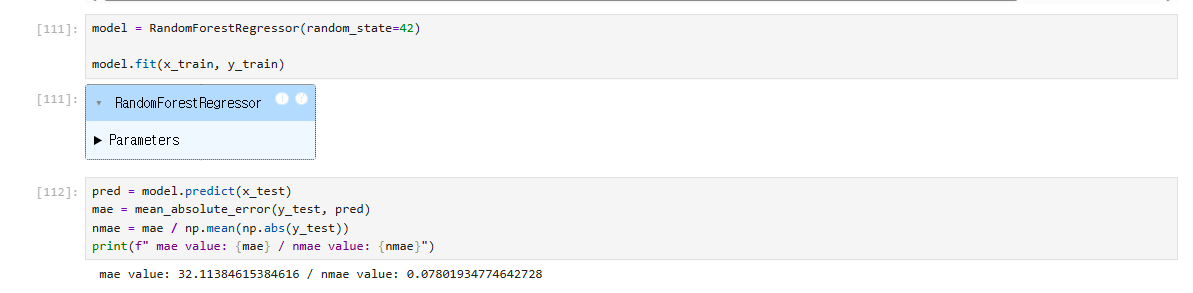
이런 경우 새순과 엽록소간 다중공선성이 있을 수 있음. 하지만 모든 독립변수가 착과량에 영향을 미치니까 릿지를 사용해서 다중공선성을 낮출 수 있음.

릿지는 변수간 다중공선성을 낮추고 과적합을 방지할 수 있음. 계수를 0에 가깝게 만들지만 0으로 만들지는 않음. 사용할 수 있는 변수로 만들어줌. 대신 모든 독립변수들을 먼저 스케일링 해야함. -> 릿지는 하이퍼파라미터를 수정해야되서 사용 불가함

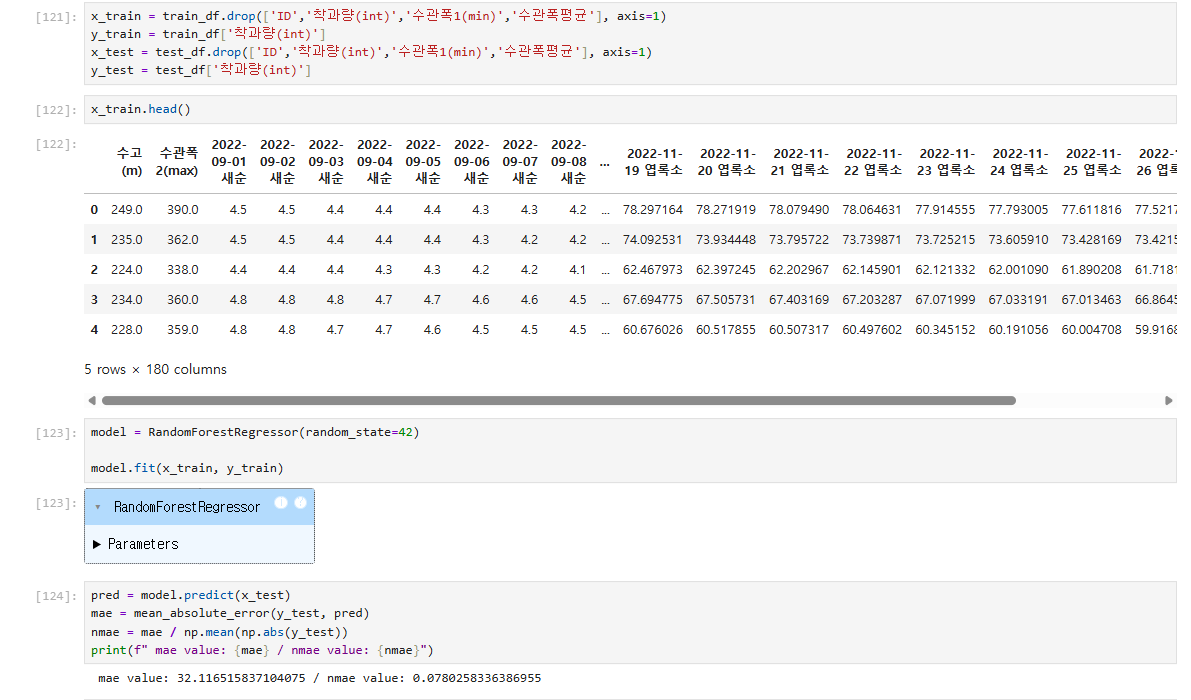
엽록소 컬럼 제거한 결과(수고, 수관폭평균으로만)



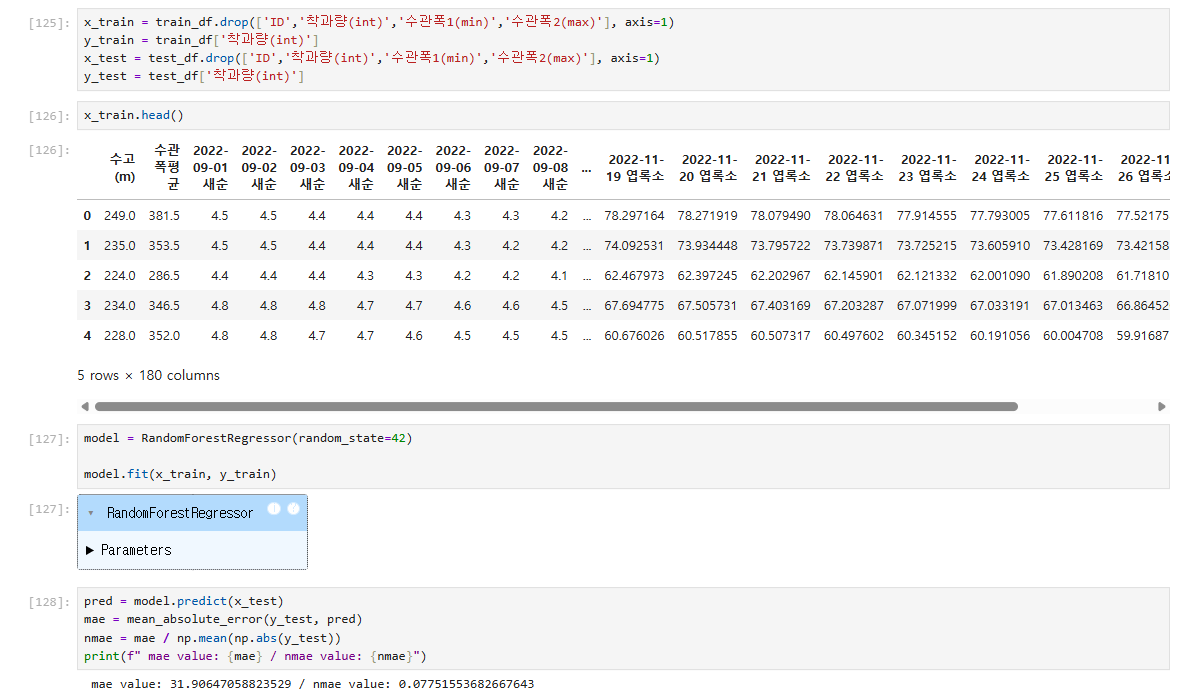
수관폭1로 한거 결과



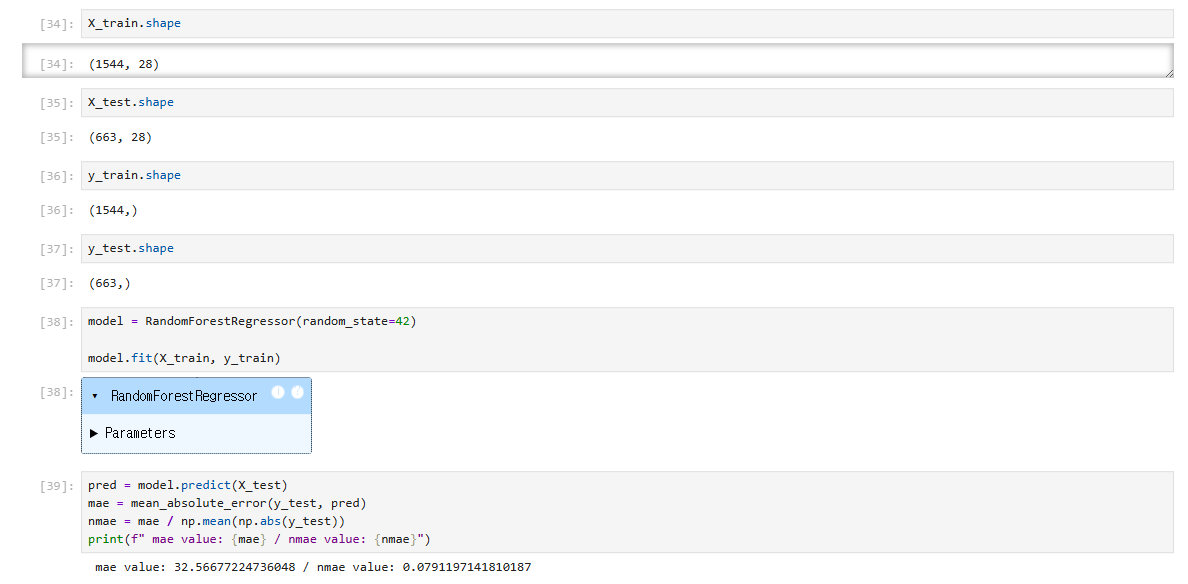
수관폭2로 한거 결과



수관폭평균으로 한거 결과



새순 엽록소 일주일치로 한거 결과



새순 엽록소 서로 곱한거 결과

