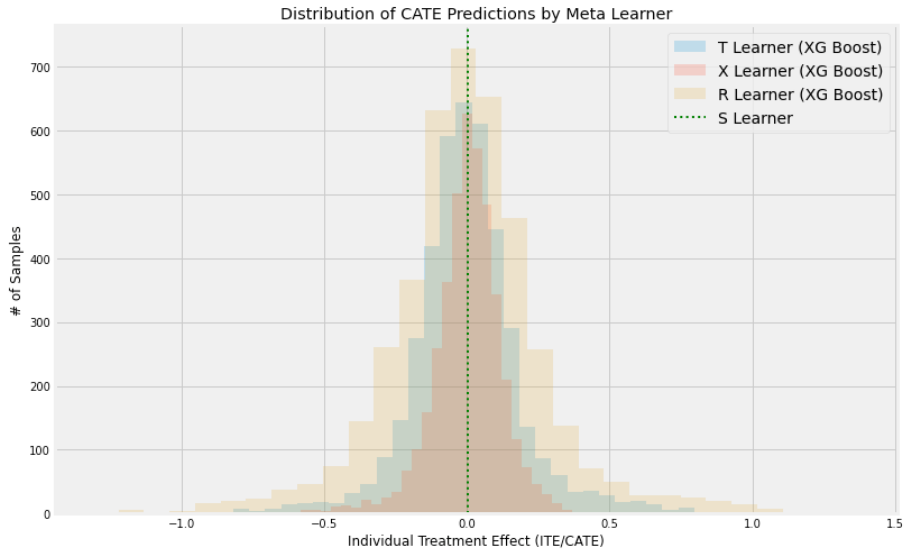
Causal Inference Project – Employee Attrition Rate

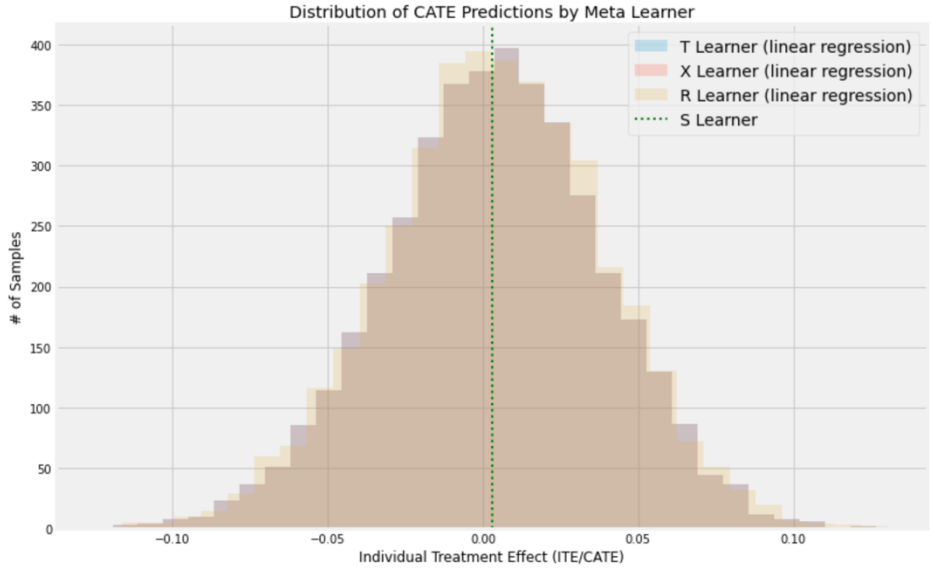
Target variable: the employee attrition rate
Treatment: compensation and benefits type 2/type 3; a total of 2664 type 2, 1456 type3
X variables: 37 variables after making categorical data into dummy variables

1 Average Treatment Effect (ATE) Estimation with S,T,X, and R learners: ATE estimate is small.

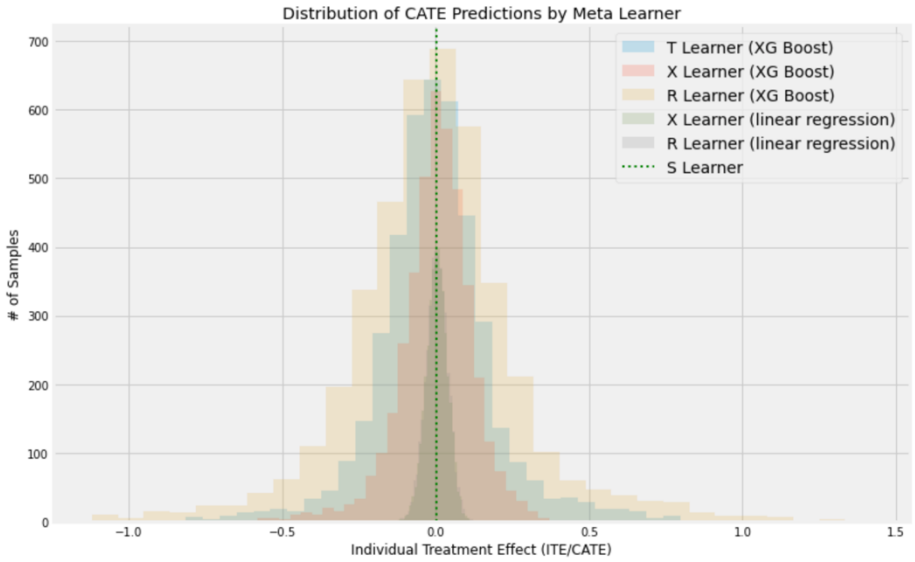
2 Individual Treatment Effect (ITE/CATE)



From the plot above, XG Boost with T learner is in the middle kind of a normal distribution.

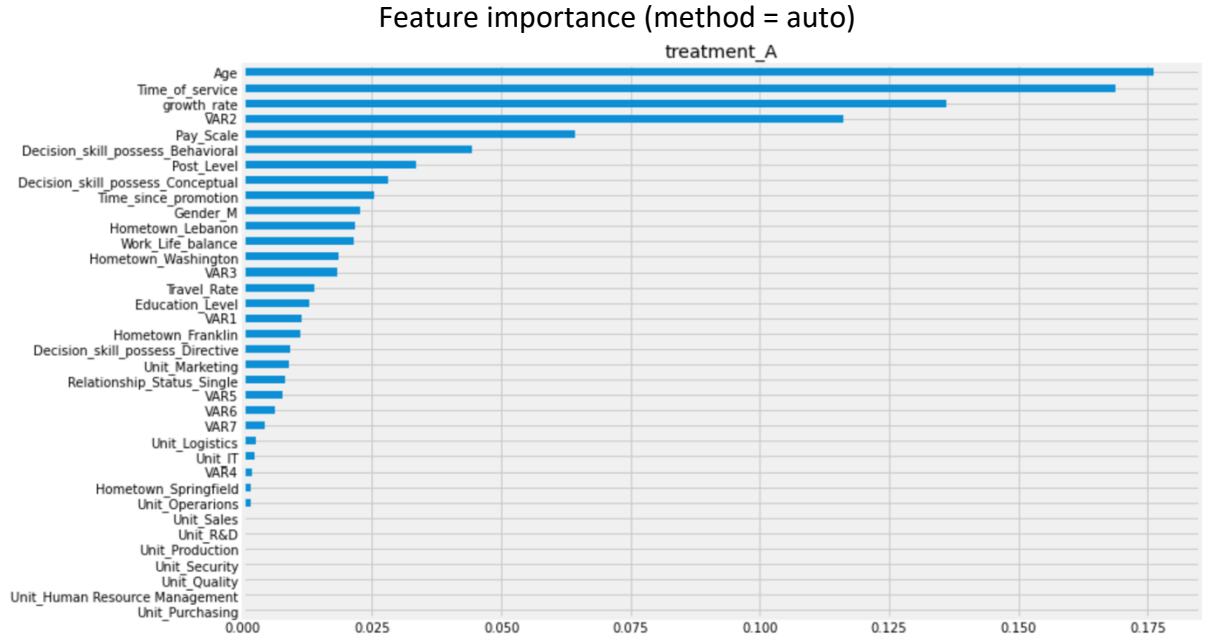


From the plot above, we can see that the distribution of linear regression with T, X, and R learners has approximately a similar distribution area.



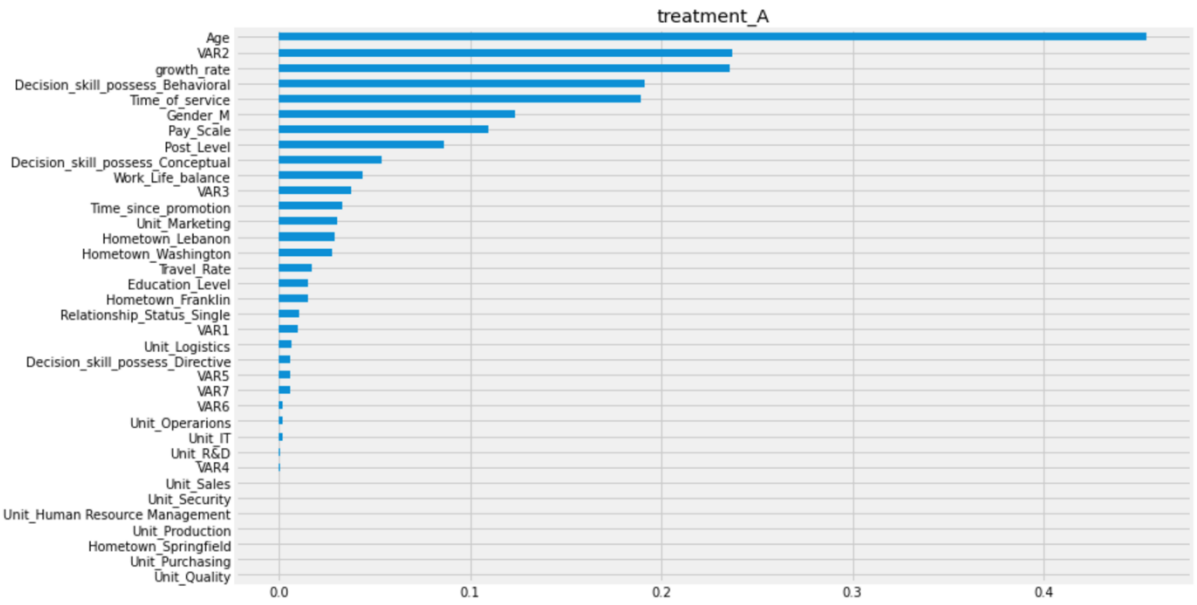
Set type 2 benefits as treatment A, type 3 benefits control variable.

3 LRSRegressor With S Learner:



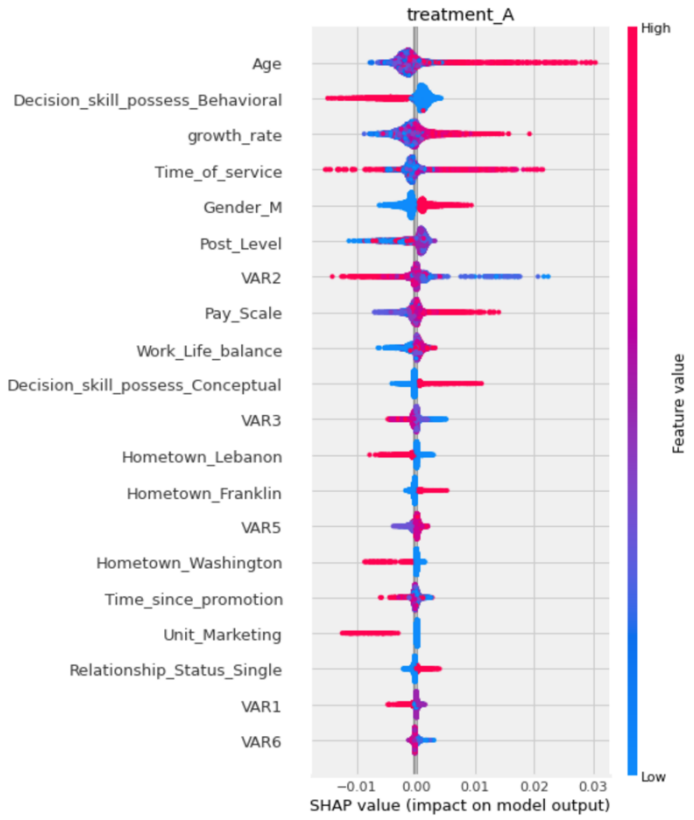
From the plot of feature importance with the auto method, we can tell the age, time of service, growth rate, VAR 2, and pay scale are the top 5 most important features. And most of the unit department is not important at all, except marketing and logistics.

Feature importance (method = permutation)

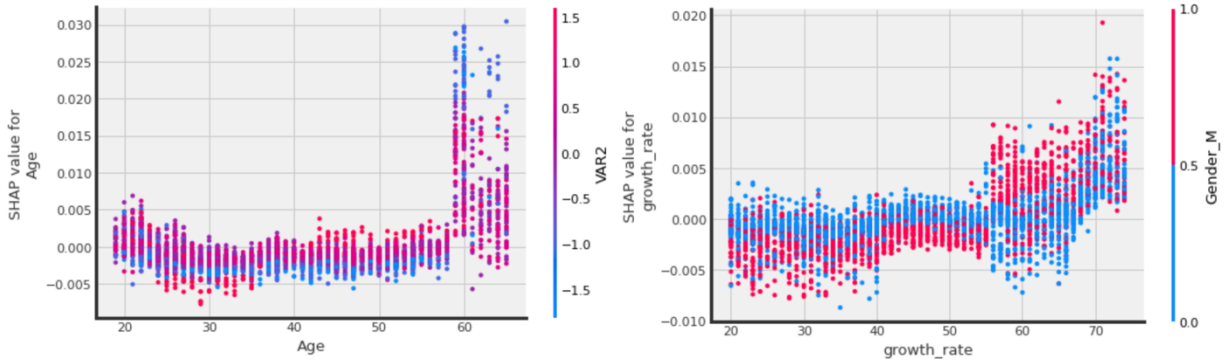


From the plot of feature importance with the permutation method, we can tell the age, VAR2, growth rate, decision skill possess behavioral, and time of service are the top 5 important features. And most of the unit department is not important at all, except marketing and logistics.

Shape Values



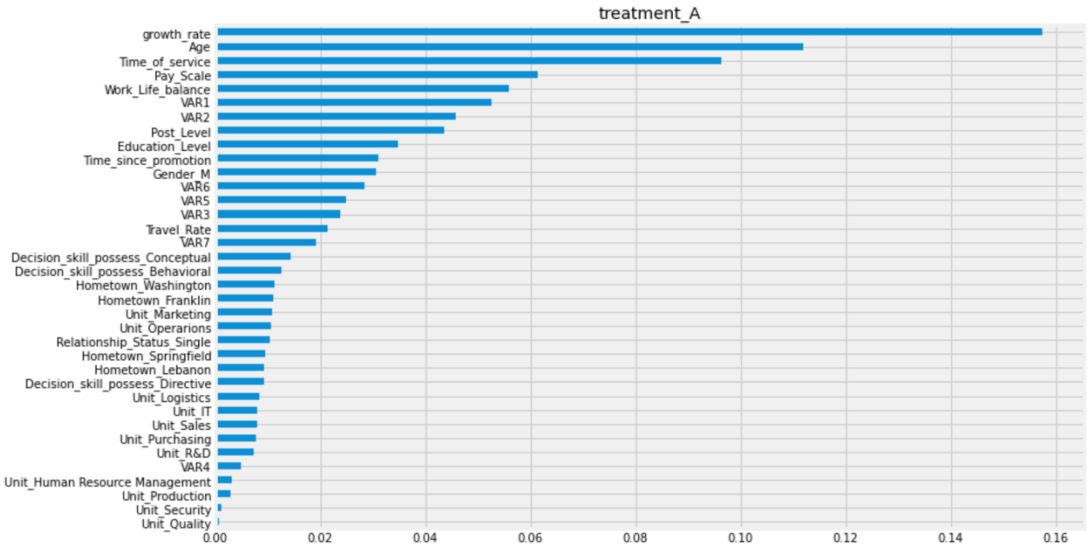
Age is the most important feature. Higher age affects the predictions in a positive direction, and lower age affects the predictions in a negative direction. We can state that the younger employee has a lower attrition rate. The second important feature is that the employee possess behavioral decision skill. When the employee has the more behavioral skill, it affects the predictions in a negative direction.



The strong interaction with age is VAR2. The strong interaction with growth rate is gender. When the growth rate is less than 55, the female employee has a higher shape value than the male. But when the growth rate is more than 55, the male has a higher SHAP value than the female employee.

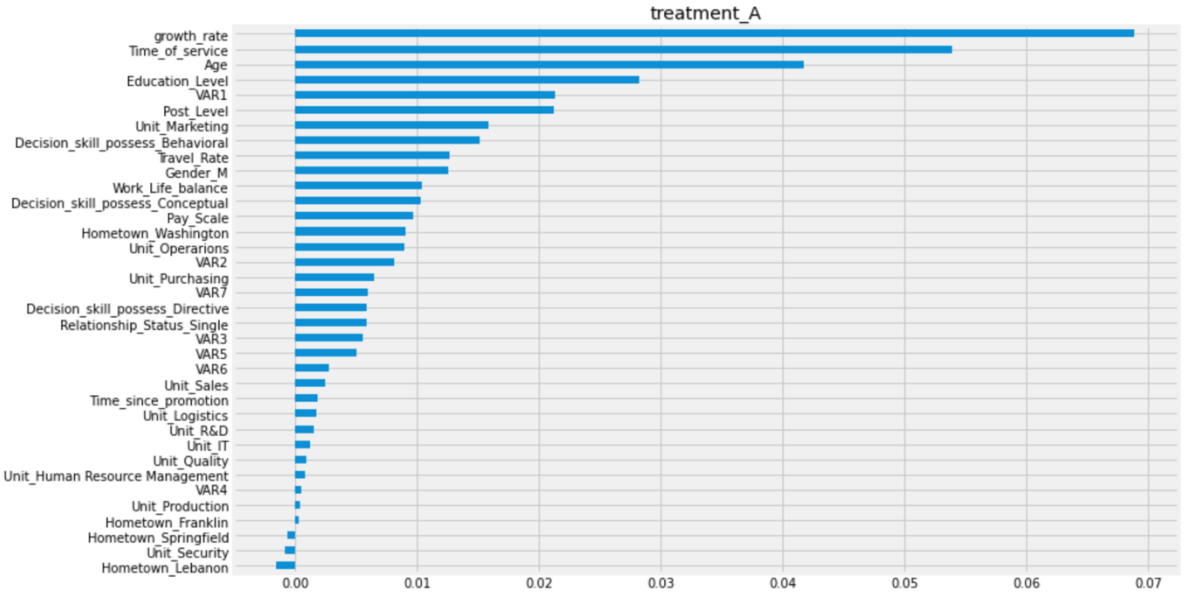
5 XGBTRegressor With T Learner:

Feature Importance (method = auto)



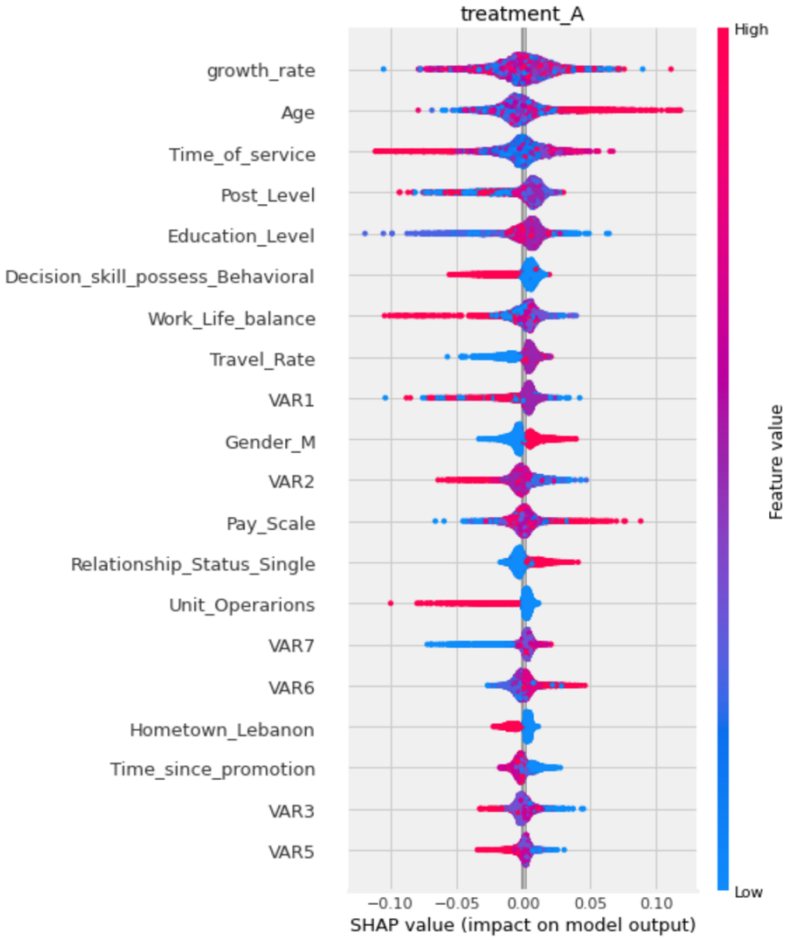
From the plot of feature importance with the auto method, we can tell the growth rate, age, time of service, pay scale, and work-life balance are the top 5 most important features. And the units are the least important features in XG Boost with T learner.

Feature Importance (method = premutation)

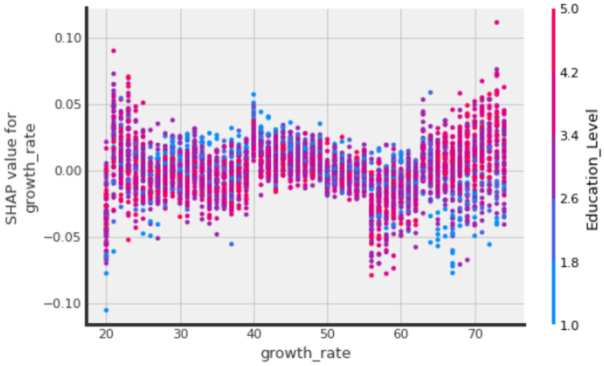


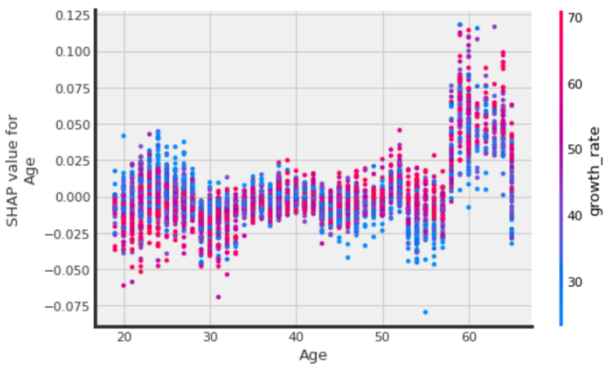
From the plot of feature importance with the auto method, we can tell the growth rate, time of service, age, education level, and VAR 1 are the top 5 most important features. And the units and hometown are the least important features in XG Boost with T learner.

Shape Values

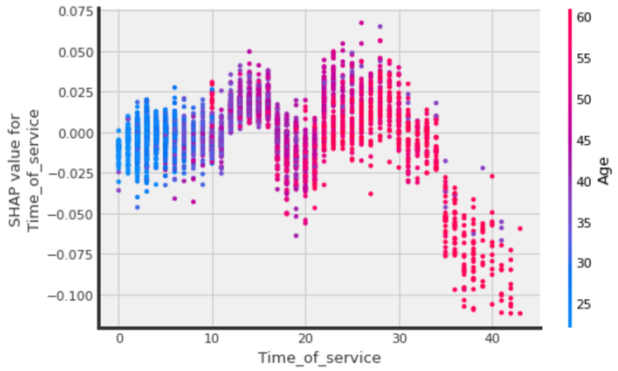


Growth rate is the most important feature. But it is hard to tell the relationship between the growth rate and the prediction direction. The second important feature is age. Higher age affects the predictions in a positive direction, and lower age affects the predictions in a negative direction. We can state that the younger employee has a lower attrition rate. The third important feature is the time of service in the company. The less time of service will affect the predictions in a negative direction.





The strong interaction with the growth rate is education level. The strong interaction with age is the growth rate. Less growth rate will stand out when the age is less than 30. When the age is between 40-60, the less growth rate employee has a lower SHAP value.



The strong interaction with a time of service is the age. They have a clear relationship that when the time of service is lower than 10, the age of the employee is below 35. Longer the service time, the older the age. The SHAP value for the time of service becomes negative when the age is between 30-40+ with a higher time of service.