Company

Employee Absenteeism Project

ISSUE / PROBLEM

The seeks to improve employee absenteeism and answer the following question:

What's likely to make the employee leave the company?

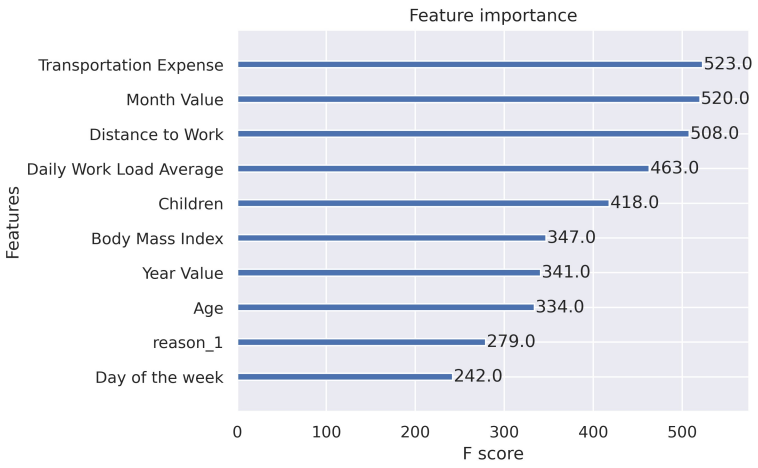
RESPONSE

Since the variable we are seeking to predict is categorical, the team could build either a logistic regression or a tree-based machine learning model or a xgboost.

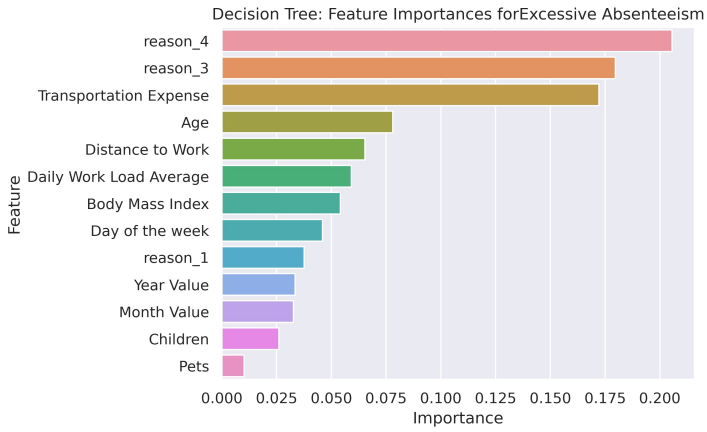
The model slightly outperforms the xgboost and decision tree model.

IMPACT

This model helps predict whether an employee will have excessive absenteeism or disengaged employee to identify which factors are most influential. These insights can help HR make decisions to improve employee



Barplot above shows the most relevant variables: '*Transportation Expense*', '*Month Value*', '*Distance to work*', '*Daily Work Load Average*' and '*Children*', etc.



In the decision tree model above, '*reason_4*', '*reason_1*', '*Transportation Expense*', '*Age*', '*Distance to Work*', and `*Daily Work Load Average*` have the highest importance. These variables are most helpful in predicting the outcome variable, `*Wxcessive Absenteeism*`.

INSIGHTS/NEXT STEPS

- Transportation expense between ~190 to 300 USD have the most employee with excessive absenteeism. We have to look to this further whether they can work from home, as the distance with this range account of ~10 to 40 distance to work but not included as feature importance.
- We can see that there are employees who are overworked increase in ~240-300 hours.
- We can see the increase of children also decrease employee with excessive absenteeism. There are employees without children who are disengaged employees. We can also elaborate by separating to two groups without children (0) and with children (1).
- Body mass index and month value.

Next Steps

It may be justified to still have some concern about data leakage and we only have small data of 700. Feture engineering is required to confirm the results further.