DATA-448 IN-CLASS ASSIGNMENT FEBRUARY 8, 2023

Consider the `turnover.csv` data file (posted under the In-Class 9 assignment link). This file contains basic employment information of employees from some company. The goal is to build a binary classification to predict employee turnover. **In Python**, answer the following:

1. (3 points) Using the pandas library, read the csv data file and create a data-frame called `turnover`.
2. (6 points) Change `sales`, and `salary` from labels to dummy variables.
3. (6 points) Using the Box-Cox transformation, transform the `time_spend_company`. Also, transform the `number_project` and `average_montly_hours` to 0-1 scale.
4. (6 points) Create two interactions: one between `satisfaction_level` and `time_spend_company` call it `interaction_1`, and another one between `last_evaluation` and `promotion_last_5years` call it `interaction_2`.
5. (5 points) Using `interaction_1`, `interaction_2`, `satisfaction_level`, `last_evaluation`, `number_project` (standardize), `average_montly_hours` (standardize), `time_spend_company` (standardize), `Work_accident`, `promotion_last_5years`, `sales` (dummy variables), and `salary` (dummy variables) as the input variables and `left` as the target variable, split the data into two data-frames (taking into account the proportion of 0s and 1s) `train` (80%) and `test` (20%).
6. (8 points) Using `train` data-frame build a random forest model (with 500 trees and the maximum depth of each tree equal to 3). Then, use this model to make predictions on the `test` data-frame. Use the provided `precision_recall_cutoff.py` (posted under the In-Class 8 Assignment link) file to estimate the optimal cutoff value. Compute the classification report of this model.
7. (8 points) Using `train` (without the interaction features) data-frame build a random forest model (with 500 trees and the maximum depth of each tree equal to 3). Then, use this model to make predictions on the `test` data-frame. Use the provided `precision_recall_cutoff.py` (posted under the In-Class 8 Assignment link) file to estimate the optimal cutoff value. Compute the classification report of this model.
8. (3 points) Using the results from part 6 and 7, what model would use to predict `left`? Be specific.