Task 3 Sub-task 2

- Ensure you're able to explain the performance of your model, where did the model underperform?
- Why did you choose the evaluation metrics that you used? Please elaborate on your choices.
- Document the advantages and disadvantages of using the Random Forest for this use case.
- Do you think that the model performance is satisfactory? Give justification for your answer.
- (Bonus) - Relate the model performance to the client's financial performance with the introduction of the discount proposition. How much money could a client save with the use of the model? What assumptions did you make to come to this conclusion?

1. Based on the metrics performance and fitting time performance
2. Because this is a classification issue, four metrics would be considered including accuracy, precision, recall and f1 score.
3. Random forest
   Pros:
   a. Robust to outliers.
   b. Works well with non-linear data.
   c. Lower risk of overfitting.
   d. Runs efficiently on a large dataset.
   e. Better accuracy than other classification algorithms.
   Cons:
   a. Random forests are found to be biased while dealing with categorical variables.
   b. Slow Training.
   c. Not suitable for linear methods with a lot of sparse features
4. Based on the metrics performance, all the scores of metrics are over 0.8, which could be verified as satisfactory.
5. Based on the analysis of feature importance, electricity consumption of the past 12 months, forecasted bill of meter rental for the next 2 months, gross margin on power subscription, forecasted electricity consumption for next 12 months, total net margin and electricity consumption of the last month are the top important features.