Eser İnan Arslan

1. What statistical technique did you choose and why is it suitable for this task?

Answer: I prefer XGBRegressor because it is a scalable and accurate implementation of gradient boosting machines and it has proven to push the limits of computing power for boosted trees algorithms as it was built and developed for the sole purpose of model performance and computational speed.

1. How did your first model perform?

Answer: By the help of small dataset, it worked fast and quite accurate.

1. What statistical metrics and techniques did you use to measure model’s performance?

Answer: I choose both “Mean Absolute Error” and “Mean Absolute Percentage Error”.

A picture containing table

Description automatically generated

At first I got “inf” as a result in MAPE because some actual values are zero. So I need to convert them as 0.1 then it solved.

1. Did you take any steps to improve the model?

Answer: I try to create new columns which I believe to improve my results. Moreover I use gridsearch to find best hyper parameters.

1. Does age variable have a statistically significant effect on number of calls?

Answer: It has small effect, ist correlation score is lower than the others.(negative correlation)

1. Does isgroup2 variable have a statistically significant effect on number of calls?

Answer: According to correlation calculation, I saw that it effect my results directly.(possitive correlation)

Text

Description automatically generated

**Bonus Question:** What is the probability of 33 year old group2 customer who is active for 75 days will make more than 4 calls?

Answer: According to my model, the estimated number of call will be 1.2