**DS ASSESMENT - BRIEF**

Dataset provided with this task was preprocessed and clean, hence no need to do data preprocessing. The features were already created, hence no further feature engineering I have used. Now, if this initial work leads further to a long-term project, then I would recommend to revise following additional steps:-

* Analysis of the original data instead of readymade features, I will create more robust features to increase the model performance.
* Next thing I will do data preprocessing i.e. outlier detection, normalization etc.
* As the only problem in dataset was class imbalance issue, hence we can try more ways to handle this problem or we can take an effective sample of the data (employees who did not churn).
* Churn analysis is more associated with time factor, hence we can treat this problem in time series nature.

In this way we can make sure to have an effective machine learning model. Now with current model in hands, how we can evaluate whether my model can replace the existing model or not? As shown in my code, we can evaluate the models using ML evaluation parameters such as precision, recall, F1-measure and accuracy. You can only rely on accuracy especially when there is class imbalance problem. We will test existing deployed model with new data (previously not seen by existing model as well as this new model), and then we test new model with unseen data. We will compare both models using these parameters and will evaluate which model is producing more false alarms. If my model will outperformed in performance, this means that we can replace the existing model with new model.

In machine learning there are some problems, for which you cannot rely on once-created-model, instead you have to update model after sometime. Let’s say we have problem of face recognition, we can have model for longer time period, because faces remain same for decades even. But let say we have model such as this problem, where we are working on features such as salary; now this feature change with time i.e. currency evaluation. Salaries in 2022 will be definitely greater than salaries in 2015, hence your old model might not be able to do predictions with that same accuracy as it was. Also we can new features, for example work from home feature was not that much famous before COVID-19. Hence we need to re-evaluate the model after few months and if required add new features (**but that should not be from scratch, we should have code written once and then just change the dataset and code will just ran on new data**).