MVP of HR dataset- Promotion Prediction

In this folder, the Minimum Viable Product (MVP) is provided.

The initial steps in detail are provided in the Jupyter notebook. Then, the data explored in addition to the EDA phase is done.

The MVP notebook summarizes the following:

1. Employee's dataset

- Explore the features
- Drop any unnecessary features
- Clean the data
- Data Analysis and Visualization

2. Prediction Model

- Data cleansing to train the model
- Convert the categories variables into Numerical variables
- Train and compare the scores of multiple ranking algorithms.
- Apply Extreme Gradient Boosting.
- Assuming "Accuracy" and "Recovery" rate have the exact cost to the company, thus it will used the F1 score, which is a balance of both scores

3. Predictions

The final result is a database that will show which employee will be promoted (1) and which will not (0) and the probability of these employees being promoted according to this model.

	department	education	gender	recruitment_channel	no_of_trainings	age	previous_year_rating	length_of_service	KPIs_met >80%	awards_won?	avg_training_score
0	2	1.0	1	0	1	24	NaN	1	1	0	77
1	7	1.0	0	1	1	31	3.0	5	0	0	51
2	0	1.0	1	1	1	31	1.0	4	0	0	47
3	5	1.0	0	1	3	31	2.0	9	0	0	65
4	6	1.0	1	0	1	30	4.0	7	0	0	61