HR Analytics: Job Change of Data Scientists

The dataset here is related to the HR analytics of a company that is active in Big Data and Data Science, to estimate the candidates looking for a job change, and the factors influencing that

We have main keys points to note in this dataset:

• Most of the data is categorical

• There is an imbalance in the classes

• There are lots of missing values

The goal of this Model is to predict the probability of a candidate work for the company.

I tried two models to get the best correlation, but none is getting good result. I will be exploring the data deeply and changing the features to enhance the models performance.

• Algorithms tested are Logistic Regression and Random Forest

I want to find out which factors are most important in by find out the correlated value in according to "target":

	city_development_index	experience_Imputed	company_size_Imputed	training_hours	target
target	0.34	0.17	0.08	0.02	1