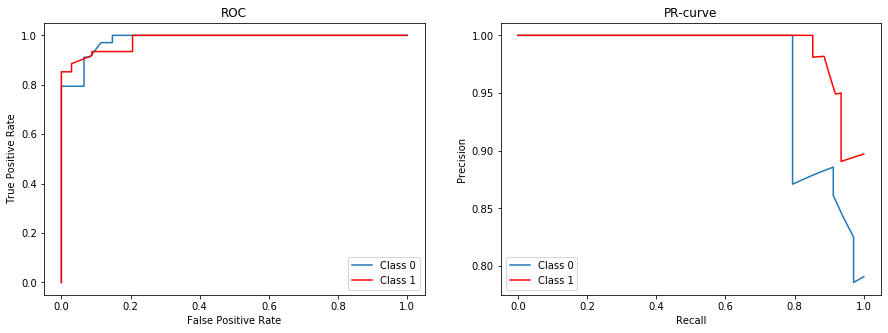
Report of the results

I investigated the dataset. It had enough columns for research, but it did not have many rows (there were 472 rows). Therefore, I was supposed to not do any row removing, because the dataset would result in having a bad model and would produce worse predictions. First I cleaned some data. Some included: not necessary columns (I explained in detail in Jupiter notebook why the columns were unnecessary), I expanded date-type columns to several columns to be used for usable training dataset. Second, I did imputations of numerical, categorical and text data. The imputation strategies that I have chosen are all in the notebook. And finally, I trained model with trained and test scores. All the scores were very high with this model and I decided not to move on to other models such as Logistic R., KNN, and SVM etc. You can see below how perfect the Roc and PR-curves are for this model. The Accuracy of test set was approx. 93%.



Moreover, to mention the status of Success/Failed of a company it is good for a company to look on the below 10 results of most important features and see what they lack THE most for Success. So the first most important thing is that for founders & cofounders of the company it is better that they have high education in business area, which is very logical, as you as a founders & cofounders of a company should be knowledgeable to before you have a startup. Next thing for a startup team is if the team has no focus on structured or unstructured data then this will influence very much their startup’s future Success/Fail. The third important things are analytics. This means a startup should be very engaged in analytics industry. There are many more of important features, but overall this is the idea behind which a company’s status is influenced mostly.

