

Understanding hypothesis and accuracy

customer churn in a telecom's supply data set

Hypothesis and null hypothesis

HYPOTHESIS: The monthly charges, the tenure of the service and the citizen status of the customers will predict their churn of the company.

NULL HYPOTHESIS: The above variables have no relation with their churn and the churn is caused by coincidence or other causes.

The strategic impact of finding an accurate result in this model

Not to lose money.

Testing how likely are these variables predicting the customers loss will help to detect how many of them denote the most likely to leave the company, in order to be able to do something to avoid losing money by losing customers.

Possible steps that should improve the model

Collecting more data about customers:

- Employment situation
- Salary
- Number of people living together and making use of the service
- Moving-house history (how many times)
- Satisfaction survey

Adding features in the model:

- Adding "contract" or more details/column about the contract
- Adding the data about HOW they use the service like "online security", etc

Handle the imbalance: collecting more data and/or oversampling.

Other statistical tests: correlation between variables.

The accuracy of the model achieved

The accuracy of the model was a 0.79 and this result suggests a good accuracy of the model. But after oversampling and remodeling, the result visualization is indicative that there is still a relevant percentage of customers that the model couldn't predict well their behaviour (dark zones).

