

Customer Churn Prediction with Pyspark

Problem Statement:

Using customer churn data from a telecommunications company, can we predict which customers will churn?

Context:

In the service industry there are many factors that lead a customer to choose your service or to cancel their services. Studies indicate that it is less expensive to keep current customers than to acquire new ones. Therefore, it is in a company's interest to be able to identify customers that are at a risk of churning to intervene or correct the reasons why a customer may leave to a new service provider.

Criteria for Success:

Identify the customers that are at the greatest risk of churning using statistical models. The models will need to present a high degree of accuracy to predict the common features of customers that are at risk of churning.

Scope of Solution Space:

This problem will aid decision makers in identifying common patterns and features of high churn clients.

Constraints:

The amount of features available on the users interactions within the service they are utilizing.

Stakeholders:

Those companies and organizations who provide a service to their customers and also have a high degree of churn and competition within their industry. These industries could be subscription based telecommunications. The main stakeholders are the senior leadership in these organizations and their competitors.

Data Sources:

<https://www.kaggle.com/blastchar/telco-customer-churn>