Clustering

Eoin Flynn

26 March 2018

Bond University
Data Science
Final Assignment

Contents

Introduction

In this report we will using an unsupervised machine learning technique called clustering to identify groups of customers within our dataset. The model will produce a set of clusters and return a "typical customer" for those groupings. A typical customer is the type of person we expect to see in that group, for example a typical customer who uses tech support might be a senior who pays month-to-month. Being able to identify groups of customers will allow you to make more informed decisions when marketing new products or entering new markets.

Cluster Model

After testing a number of different models, we were able to find one which has the greatest amount on insight into your customers without being too specific, or too general. The model we created contains 14 clusters of customers, each with their own unique characteristics. If we focus on group five for example we can find some information that can be used to make more informed management decisions.

gender	0.4896907
SeniorCitizen	0.2061856
Partner	0.8092784
Dependents	0.3608247
tenure	71.1340206
PhoneService	1.0000000
InternetService	1.0000000
Contract	0.0206186
PaperlessBilling	0.7783505
MonthlyCharges	111.6033505
TotalCharges	7955.4891753
Churn	0.0824742

This output shows us what the typical customer looks like in this group. They are very unlikely to churn, they have high-monthly charges, they are most likely on a yearly or bi-yearly contract, they have a phone service and internet service, they have a partner, and are most likely not a senior citizen. From this information we can make many recommendations to the business such as:

Recommendations

- Try to push all customers onto one or two year contracts.
- Offer customers on longer term contracts phone services as well as internet services.
- Try to have all customers on phone and internet plans.
- If someone is a not a senior citizen and does have a partner then offer them a longer term contract.
- Gender should not be used to determine the likelihood of a person churning since the typical customer is 48% likely to be female, and 52% likely to be male, meaning that there is no real predictive value in this number.

Conclusion

The model that we have created here should be used for identifying groups of customers within the company. Identifying these new groups will allow you to better utilise your marketing and sales budgets, as you will be able to create more specific offers which can be used to entice customers to stay with your network. Using a model with fourteen clusters gives you the ability to identify clusters of customers which are not so specific

that they only apply to a minuscule amount of your customers, but also no so general that you gain no real insight.