DETERMINING EFFICIENT CLASSIFICATION ALGORITHMS WITH FEATURE SELECTION AND DATA BALANCING TO PREDICT CUSTOMER CHURN

Nafisseh Heiat, College of Business, Montana State University-Billings, 1500 University Drive, Billings, MT 59101. 406-657-2224, nheiat@msubillings.edu

ABSTRACT

In many industries customer turnover or churn is an important concern. In this study a number of classification algorithms are analyzed to determine the most efficient model for predicting customer churn in telecommunications.

INTRODUCTION

In many industries specially telecommunication field customer turnover or churn is an important concern. Churn is a term for subscribers switching from one telecommunication provider to another. In this study I am interested in voluntary churn. Voluntary churn is when the customer decides to leave the current provider for a variety of reasons like subscription costs, quality of service, more attractive incentives from competition etc. Customers switch from one provider to another results in loss of considerable profit. According to Lu Junxiang telecommunications industry experiences an average of 30-35 percent annual churn rate and it costs 5-10 times more to recruit a new customer than to retain an existing one.

In order to identify high risk customers that may switch to another provider and manage churning rate, we need to develop reliable models. Once a reliable and accurate churn model is developed and tested, companies may use the model for churn management. In this study we analyzed a number of classification algorithms to determine the most efficient model for predicting customer churn.

DATA PROCESSING

The dataset used in this study has 3333 entries and 21 attributes. The attributes are shown in table 1.

Table 1. Attributes in telecommunication dataset.

S.No.	Attribute name
1	State
2	Account. Length
3	Area. Code
4	Phone
5	Int .l .Plan
6	VMail.Plan
7	VMail.Message
8	Day.Mins
9	Day.Calls
10	Day.Charge
11	Eve.Mins
12	Eve.Calls
13	Eve.Charge
14	Night.Mins
15	Night.Calls
16	Night.Charge
17	Intl.Mins
18	Intl.Calls
19	Intl.Charge
20	CustServ.Calls
21	Churn.

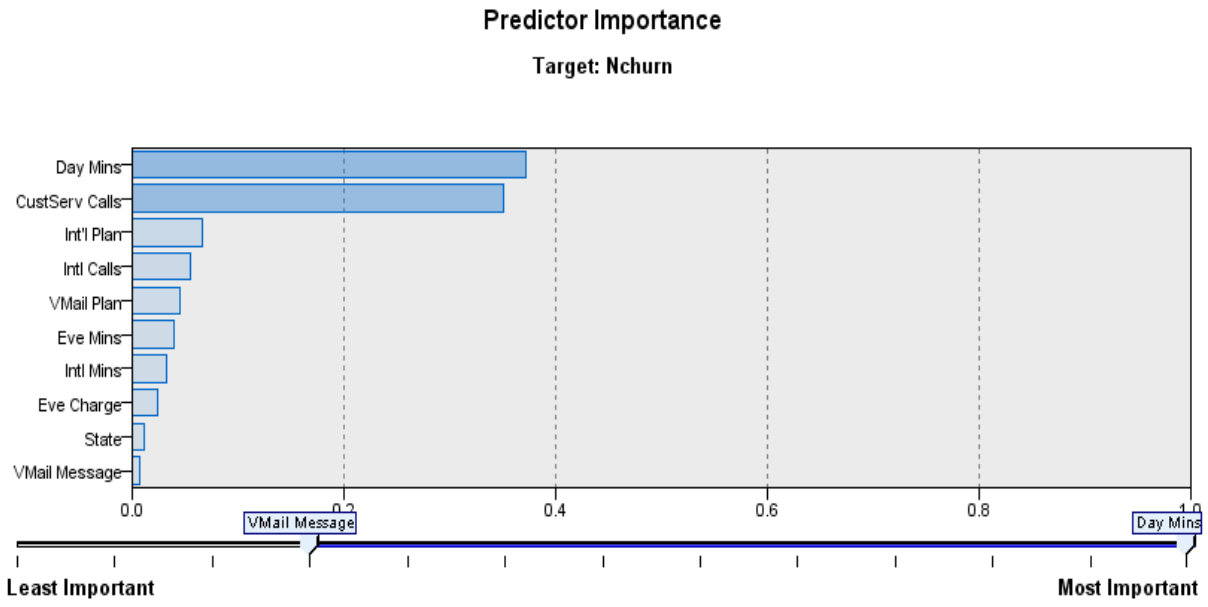
METHODOLOGY

Auto-Classification of SPSS Modeler is used to determine the best algorithm for the dataset used in this study. The most accurate algorithm, C5 classification, is applied to telecommunication dataset.

FINDINGS

The most important attributes according to C5 algorithm for customer churn is displayed in Figure 1.

Figure 1. The most important attributes influencing customer churn



REFRENCES

References are available upon request.