

Building a predictive model to predict the likelihood of a customer churning

Dataset Link

.csv file for the same will be shared with the students

Dataset Description

The data provides historical customer transactions for a retailer, for purchases made on their e-commerce platform between Dec 2010 to Dec 2011. Customers are uniquely identified by their Customer ID (Col G).

Problem Statement

Assume that the company has been experiencing high customer churn and a reduction in repeat customers. **As part of this case study, you are required to build a predictive model to predict the likelihood of a customer churning**

Churn Definition (Target Variable definition)

For this problem, the churn should be defined as follows:

The training data should include all customers who have made a transaction between 1st Dec 2010 to 31st Aug 2011. Customers from this set who have NOT made any subsequent purchase in the period Sep 2011 to Dec 2011 should be labeled as **“Churn”**. All other customers with purchases made during this follow-up period should be labelled as **“Not Churn”**

- A. Target variable creation (based on churn definition provided) – 2M
- B. Customer level feature creation, from transaction dataset – 2M
- C. Construct a **Decision Tree using ID3** algorithm and classify whether the customer will churn/not Churn. – 4M
- D. **Investigate** whether ID3 can be used for constructing oblique DT or not? – 2M
- E. Evaluation of the model – 2M
- F. Use the above dataset to implement **SVM and ANN** for training and testing purposes – 5M
- G. Construct a **comparative table** for DT, SVM and ANN - 3M