Hybrid Classification Algorithm for Customer Churn Prediction

CO421: Data Warehousing and Data Mining

About the Paper

Topic:

A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees

Authors:

Arno De Caigny, Kristof Coussement, Koen W. De Bock

Link:

https://www.sciencedirect.com/science/article/pii/S0377221718301243

Problem Statement

- Predict the number of customers leaving
- Estimation of a future churn probability for every customer based on the historical knowledge of the customer

Dataset Description

- The dataset chosen is from: Telecom Customer Churn
- Each row represents a customer, each column contains customer's attributes described on the column Metadata.

Raw Data

- 7043 rows (customers)
- 21 columns (features)

Fields	Fields	Fields
customerID	MultipleLines	StreamingMovies
Gender	InternetService	Contract
SeniorCitizen	OnlineSecurity	PaperlessBilling
Partner	OnlineBackup	PaymentMethod
Dependents	DeviceProtection	MonthlyCharges
Tenure	TechSupport	TotalCharges
PhoneService	StreamingTV	**Churn**

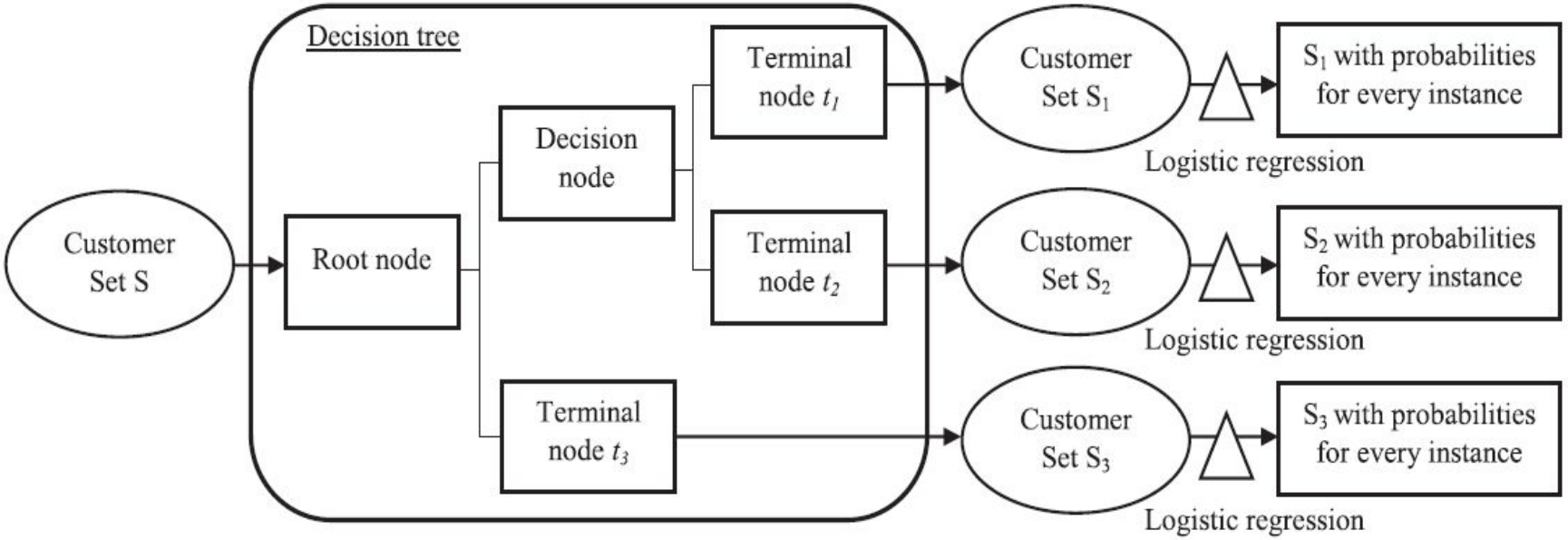
Existing Solutions

- Decision tree
- Logistic regression
- Random forest
- Logistic model trees

Proposed Approach (LLM)

- Logit leaf model
- A combination of decision trees and logistic regression
- The customer set S is split by the tree structure into customer subsets
- In the LLM algorithm, in a second step, a logistic regression with forward variable selection is fitted at every terminal node

Proposed Model Structure

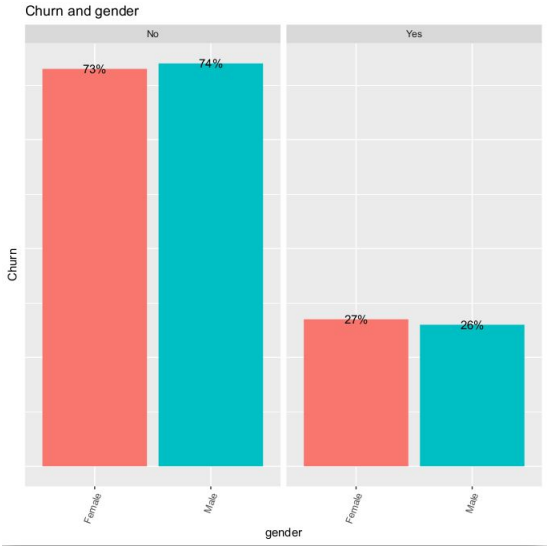


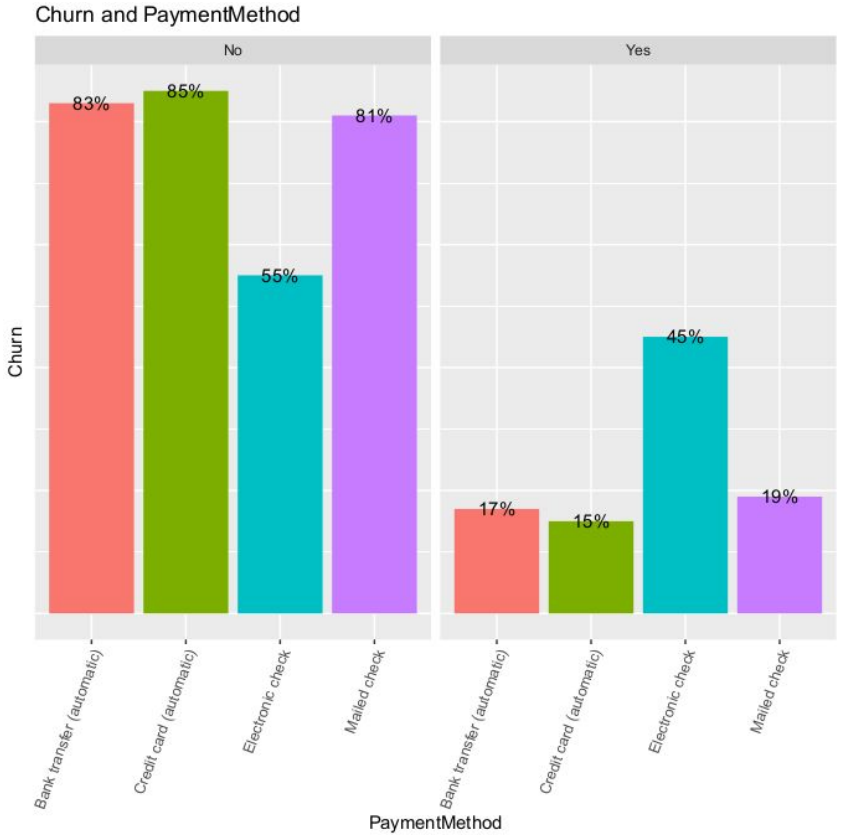
Progress

- Programming environment - R
- Dataset
- Data analysis by plots
- Data preprocessing

Data Analysis Plots

- Programming environment - R
- Dataset
- Data analysis by plots





Data Pre Processing

Data cleaning and preparation

- Checking for “null” values.
- Removing the columns we won’t analyze.
  - eg: customer ID, gender, phone service, multiple lines, monthly charges, total charges

Thank You!

Mehnaz Yunus - 16CO124

Mishal Shah - 16CO125

Sharanya Kamath - 16CO140