**Case studies on Decision Tree**

**Case: 1 - Auto insurance policy records -**

Adam, an Analytics consultant works with First Auto Insurance Company. His manager gave him data having policy level and loss amount related details of a group of customers. He asked him to “identify” the customer you paid the premium on time and who are not able to pay based on following characteristics:

**Data Dictionary:**

1. Policy Number
2. Age-Band ; A = 15-25 yrs., B = 26-59 yrs, C = 60+
3. Years of Driving experience
4. Number of Vehicles
5. Gender ; M = Male, F = Female
6. Vehicle Age
7. Married ; Married, Single
8. Premium payment ; 1= premium paid, 0 = premium not paid
9. Losses

**Question**:

Develop predictive model (Decision Tree) to identify the probability of the customers paid the premium or not?

**Case: 2 - Romanov, an Analytics consultant –**

Romanov, an Analytics consultant works with Credit One Bank. His manager gave him data having “Credit” and personal information of a group of customers. Some of the customers had defaulted in making the payment on balance due. He asked him to “identify” and “quantify” the factors responsible for defaults in a multivariate fashion and find out the probability of default corresponding to each of the customers. Romanov has no knowledge of running a multivariate regression se) which predicts probability of occurrence of an event (default in this case). The variable in the data as follows:

**Data Dictionary:**

1. Title: Credit data

2. Source: Credit One Bank

3. Number of Instances: 5000

4. Name of Dataset: Analysis\_of\_Default

6. Number of Attributes: 20 (7 numerical, 13 categorical)

7. Attribute description

Attribute 1: (Qualitative / Categorical)

Status of existing checking account

A11: ... < 0 USD

A12: 0 <= ... < 10000 USD

A13: ... >= 10000 USD

A14: no checking account

Attribute 2: (numerical)

Duration in month

Attribute 3: (Qualitative / Categorical)

Credit history

A30: no credits taken/all credits paid back duly

A31: all credits at this bank paid back duly

A32: existing credits paid back duly till now

A33: delay in paying off in the past

A34:critical account/other credits existing(not at this bank)

Attribute 4: (Qualitative / Categorical)

Purpose

A40: car (new)

A41: car (used)

A42: furniture/equipment

A43: radio/television

A44: domestic appliances

A45: repairs

A46: education

A47: (vacation - does not exist?)

A48: retraining

A49: business

A410: others

Attribute 5: (numerical)

Credit amount

Attribute 6: (Qualitative / Categorical)

Savings account/bonds

A61: ... < 1000 USD

A62: 1000 <= ... < 5000 USD

A63: 5000 <= ... < 10000 USD

A64: .. >= 10000 USD

A65: unknown/ no savings account

Attribute 7: (Qualitative / Categorical)

Present employment since

A71: unemployed

A72: ... < 1 year

A73: 1 <= ... < 4 years

A74: 4 <= ... < 7 years

A75: .. >= 7 years

Attribute 8: (numerical)

Installment rate in percentage of disposable income

Attribute 9: (Qualitative / Categorical)

Personal status and sex

A91: male : divorced/separated

A92: female: divorced/separated/married

A93: male : single

A94: male : married/widowed

A95: female: single

Attribute 10: (Qualitative / Categorical)

Other debtors / guarantors

A101: none

A102: co-applicant

A103: guarantor

Attribute 11: (numerical)

Present residence since

Attribute 12: (Qualitative / Categorical)

Property

A121: real estate

A122: if not A121: building society savings agreement/

life insurance

A123: if not A121/A122: car or other, not in attribute 6

A124: unknown / no property

Attribute 13: (numerical)

Age in years

Attribute 14: (Qualitative / Categorical)

Other installment plans

A141: bank

A142: stores

A143: none

Attribute 15: (Qualitative / Categorical)

Housing

A151: rent

A152: own

A153: for free

Attribute 16: (numerical)

Number of existing credits at this bank

Attribute 17: (Qualitative / Categorical)

Job

A171: unemployed/ unskilled - non-resident

A172: unskilled - resident

A173: skilled employee / official

A174: management/ self-employed/

highly qualified employee/ officer

Attribute 18: (numerical)

Number of people being liable to provide maintenance for

Attribute 19: (Qualitative / Categorical)

Telephone

A191: none

A192: yes, registered under the customer’s name

Attribute 20: (Qualitative / Categorical)

foreign worker

A201: yes

A202: no

8. Default on Payment due

1 (Defaulted) 0 (No Default)

**Question** –

Prepare the predictive model to identify the default rate.

**Case Study: Business School Data**

Nowadays, business schools are conducting entrance tests for their admission of candidates. Almost all these tests have three components. Therefore, success and failure of MBA aspirants may depend on the written score, group discussion, and interview scores. The higher is the score, the higher are the chances of the aspirants to succeed, that is, the probability of an aspirants for an admission to a business school may depend on the scores in the three components.

New age coaching institute in a metropolitan city is engaged in training the graduate aspirants for business school entrance test for last 10 years. It has faculty members who are taking different mock tests, conducting group discussions, and mock interviews to enhance the skills of the students in these areas. They have records for all these tests and collect information on their successful and unsuccessful students every year after the results are declared by the top business schools. The institute wants to find the probability of a student becoming successful on the basis of their test scores.

A sample of 3000 students from the institute has been collected relating of the following variables with the information on the success and failure.

Questions:

1. Fit a predictive model for the data and calculate the probability for the successful students with different scores in different components?

**Case – Credit Card Payment** –

IDBI, bank working on a credit card data of the customers. The data is having the name of bank's customers. Further the bank had asked to pull the information from bank's database pertaining to the customer list. The information will be around the credit cards issued by the bank.

**Data Dictionary** –

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable Name** | **Name of Customer** | **Customer ID** | **Number of Credit Cards** | **Age of Customer Last Birthday** | **Gender of Customer** | **Marital Status of Customer** | **Annual Salary** | **Monthly Credit Card Usage** | **Payment Made on time** |
| **Value Stored** |  |  |  |  | Male  Female | Divorced  Married  Never Married | ? | High (<75%)  Low  Low (<25%)  Medium (<50%)  Very High (>75%) | 1 = yes  2 = no |

Question –

1. Develop the predictive model to estimate the probability of the customers you will make the make of credit card on time and also develop the score card and interpret the data?

**Case - Predictive Modelling In Telecom Industry** –

**Business Problem**: Company X in telecom domain has strong market share. The market share has reached a point where capturing new customers is quite expensive and has low ROI. Company X planned to focus on its existing customers for growth in its revenue. To capture growth, company decided to target its existing customers for higher value plans through an up-sell campaign. Since the company has a very huge database of existing customers, it is not possible to target each and every customer and hence marketing team proposed to develop a propensity model for identifying the customers whom they should target in this campaign.

Data Dictionary –

|  |  |  |
| --- | --- | --- |
| **Variable** | **Level** | **Code** |
| Responder |  |  |
| Region\_cd | Var1 | 1. North 2. South 3. Central |
| Income | Var2 | 1. High 2. Low 3. Medium |
| Occupation | Var3 | 1. Government 2. Private 3. Unemployed |
| Age | Var4 |  |
| Gender | Var5 | 1. Male 2. Female |
| No\_of\_Serv\_Purchase | Var6 |  |
| Val\_of\_Purchase | Var7 |  |
| Mth\_Since\_Lst\_Purchase | Var8 |  |
| Pymt\_type | Var9 | 1. Card 2. Cash |
| No\_of\_calls | Var13 |  |
| No\_of\_Outgoing\_calls | Var14 |  |
| No\_of\_InComing\_calls | Var15 |  |
| No\_of\_Roaming\_calls | Var16 |  |
| No\_of\_International\_calls | Var17 |  |
| No\_of\_SMS | Var18 |  |
| Total\_min\_of\_calls | Var19 |  |
| Total\_Bill\_Amt | Var22 |  |
| Total\_Bill\_Amt\_Voice | Var23 |  |
| Total\_Bill\_Amt\_VAS | Var24 |  |
| Total\_Bill\_Amt\_SMS | Var25 |  |
| Avg\_No\_Days\_Late\_Payment | Var26 |  |
| Payment made by customers | Var 27 | 1. yes 2. No |

Questions –

1. Develop the predictive model to predict the probability of the customers they will make the payment? Develop the score card based on the model? Interpret the case?