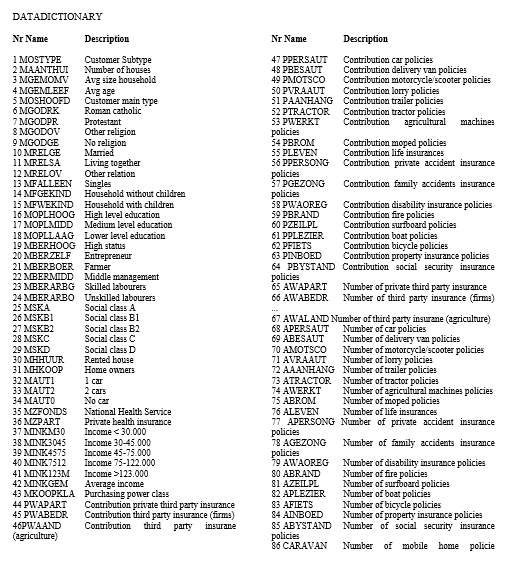
Terdapat 86 variabel/atribut yang akan digunakan dalam penelitian, dan terdapat 5822 pengamatan. Data teersebut berisis informasi pelanggan dari sebuah perusahaan asuransi. Data tersebut terdiri dari 86 variabel dan termasuk data penggunaan produk dan data socio-demograpic dari kode zip area (data tersebut berasal dari suatu area yang sama). Semua data tersebut merupakan data dari customer yang tinggal pada area yang sama.

Permasalahan yang ingin diselesaikan dari data tersebut adalah untuk mengetahui prediksi (menemukan model) costumer mana yang memiliki potensial untuk membeli “cavaran insurance policy”. 85 variabel lainnya yang terdapat dalam dataset merupakan variable predictor (input) yang akan digunakan untuk menganalisis apakah pelanggan tersebut memiliki potensi atau tidak. Target dalam dataset adalah variable “CAVARAN-number of mobile home policy” dimana kode 1 menunjukkan bahwa customer memiliki potensial dank ode 0 menunjukan bahwa customer tidak memiliki potensial.

Berikut adalah variable variable yang terdapat pada data :



1. Apakah tipe dari permasalahan diatas? Dan pendekatan metode maschine learning apakah yang dapat digunakan?

Jawab:

Permasalan diatas adalah untuk mendapatkan model yang akan digunakan perusahaan asuransi tersebut untuk memprediksi customer yang memiliki potensi untuk membeli “cavaran insurance policy”. Dimana pada data tersebut variable respon atau output nya adalah variable “CAVARAN-number of mobile home policy” dimana kode 1 menunjukkan bahwa customer memiliki potensial dan kode 0 menunjukan bahwa customer tidak memiliki potensial. Untuk input atau variable predictor adalah 85 variabel lainnya. Pendekatan machine learning yang digunakan sesuai dengan kasus ini adalah regresi logistic dengan pendekatan machine learning. Karena variable repon atau output memiliki 2 jenis/kode, yaitu 0 untuk customer yang tidak memiliki potensial dan 1 untuk customer yang memiliki potensial.

1. Tunjukkan estimasi model dari kasus tersebut (bagi data menjadi data training dan data testing)

Jawab :

Estimasi model regresi logistic dengan pendekatan machine learing berdasar kasus diatas adalah sebagai berikut

Coefficients:

Values Std. Err.

(Intercept) 2.45777110 11.92583011

X1 0.02207208 0.05151119

X2 -0.21848823 0.22163127

X3 -0.18862236 0.16216414

X4 0.20883146 0.11244944

X5 -0.06571027 0.23201685

X6 -0.07212920 0.11843208

X7 0.01372994 0.13230895

X8 0.04019172 0.11897499

X9 -0.02722062 0.12466773

X10 0.15175306 0.17164930

X11 0.07180354 0.16370931

X12 0.10433185 0.17221043

X13 -0.08735687 0.15011867

X14 -0.13593887 0.15109983

X15 -0.06205278 0.16070852

X16 0.06217727 0.14757342

X17 -0.06304474 0.15333446

X18 -0.16045180 0.15492875

X19 0.12097617 0.10599278

X20 0.05588117 0.11129790

X21 -0.12312235 0.12501903

X22 0.18168162 0.10346935

X23 0.04890552 0.10134569

X24 0.11798757 0.10243382

X25 0.03755221 0.11577044

X26 0.02389082 0.11305372

X27 0.04511185 0.10296726

X28 0.09380490 0.11118715

X29 0.02005064 0.10883064

X30 -0.71524179 1.30331369

X31 -0.69082641 1.30228117

X32 0.22130414 0.17126741

X33 0.16016927 0.15631976

X34 0.06040015 0.16163754

X35 -0.32572227 1.29624721

X36 -0.38031083 1.29421702

X37 -0.04109849 0.11340445

X38 0.02193270 0.10958830

X39 -0.01264803 0.10931966

X40 0.10557621 0.11803302

X41 -0.26066889 0.16735533

X42 -0.01033488 0.10737045

X43 0.02400454 0.05168785

X44 0.49314144 0.42548825

X45 -0.46757494 0.80699103

X46 -1.18050895 1.55914921

X47 0.20844902 0.04679524

X48 0.42738317 1.22949053

X49 -0.06315433 0.12282550

X50 -1.34325685 11.72106913

X51 1.10470070 0.98652141

X52 1.75564810 0.81342456

X53 -1.55871750 6.71886332

X54 0.22290866 0.50087275

X55 -0.19100292 0.13745696

X56 -2.16952927 5.70550326

X57 1.07517083 1.08945662

X58 0.89052588 0.66966357

X59 0.24949042 0.08842046

X60 -0.76176614 4.09149998

X61 0.30907957 0.34697613

X62 -0.16207941 0.92481161

X63 -1.10039899 1.01554662

X64 -0.11445722 0.35008894

X65 -0.80526190 0.85140510

X66 0.14259787 2.14612829

X67 1.87685759 5.01437239

X68 -0.06729330 0.19917413

X69 -2.21795430 7.12670843

X70 0.24240192 0.38968128

X71 -0.30744509 2.45453866

X72 -1.45678774 1.76405761

X73 -6.60656187 3.46079603

X74 -0.78245660 3.04803473

X75 -1.03189241 1.61072330

X76 0.33200742 0.27151392

X77 -0.96793503 3.13133948

X78 -1.86008715 2.89316480

X79 -3.35116021 3.94777400

X80 -0.41531659 0.31671290

X81 1.27739575 5.84042768

X82 0.93143616 1.11815832

X83 0.55482008 0.60141424

X84 2.29036903 1.76976462

X85 0.87389409 1.14727531

Residual Deviance: 1780.836

AIC: 1952.836

Model yang didapatkan berdasarkan nilai koefisien yang didapatkan adalah sebagai berikut :

X86 = 2,457 + 0,02 X1-0,188 X2+…+2,29 X84+0,873 X85

Berikut adalah hasil akurasi dari model tersebut berdasarkan data training dan data testing

* Data training (diambil 80% dari total data)

|  |
| --- |
| target.train  pred.train 0 1  0 4380 263  1 5 9  Dengan akurasi sebesar :  [1] 0.9424522 |
|  |

* Data testing (Diambil 20% dari total data)

|  |
| --- |
| target.test  pred.test 0 1  0 1087 76  1 2 0  Dengan akurasi sebesar :  [1] 0.9330472 |
|  |
| |  | | --- | |  | |

1. Bandingkan dengan metode klasik

Metode klasik yang digunakan adalah metode regresi logistic. Berikut adalah model yang didapatkan :

Coefficients:

Values Std. Err.

(Intercept) 2.1138420604 9.77059240

X1 0.0670403757 0.04632134

X2 -0.1705570569 0.19179858

X3 -0.0317033474 0.14022195

X4 0.2078925650 0.10185866

X5 -0.2812011371 0.20792801

X6 -0.1080357214 0.10691843

X7 -0.0121619221 0.11750560

X8 -0.0070862370 0.10546882

X9 -0.0607260997 0.11102020

X10 0.2482434142 0.15647286

X11 0.1008123464 0.14599478

X12 0.1655043532 0.15599553

X13 -0.0495019017 0.13089533

X14 -0.0856672252 0.13360651

X15 -0.0437624436 0.14162483

X16 0.0103215264 0.13120910

X17 -0.0874389148 0.13661582

X18 -0.1987727113 0.13764142

X19 0.0717739654 0.09343258

X20 0.0333249683 0.09922912

X21 -0.1298829331 0.11075382

X22 0.1197639857 0.09200242

X23 0.0254773825 0.09087334

X24 0.0838343850 0.09146968

X25 0.0359612906 0.10371652

X26 0.0009911826 0.10123204

X27 0.0228173487 0.09104453

X28 0.1062209355 0.09978767

X29 -0.0082631102 0.09750244

X30 -0.9311063625 1.16634889

X31 -0.8949771486 1.16551862

X32 0.1964464941 0.15137528

X33 0.1685125673 0.13691031

X34 0.1074860590 0.14340226

X35 -0.3140487285 1.16847923

X36 -0.3741117150 1.16725475

X37 0.1202393160 0.10093040

X38 0.1404112893 0.09687348

X39 0.1048290540 0.09680303

X40 0.1178892998 0.10284020

X41 -0.1622395595 0.14564133

X42 0.0974358446 0.10039523

X43 0.0693593100 0.04648910

X44 0.5937896592 0.38956926

X45 -0.2482981750 0.45499665

X46 -0.4229858545 1.04450251

X47 0.2315511840 0.04207911

X48 0.2189237687 0.98096772

X49 -0.0853328776 0.11393711

X50 -1.0072866087 4.44658365

X51 1.0766041448 0.93464029

X52 0.7438863302 0.43180131

X53 -1.4390009417 4.06630939

X54 0.2330989080 0.48578192

X55 -0.2424228452 0.11691580

X56 -0.3691349477 2.00726418

X57 1.3818165049 1.02046790

X58 0.7965374485 0.55769261

X59 0.2336176472 0.07671064

X60 -0.6207441432 2.55681189

X61 -0.1715063562 0.32540633

X62 0.3605443317 0.83456738

X63 -1.1834410120 0.90864812

X64 -0.1388154238 0.33291706

X65 -0.9281424806 0.77910554

X66 0.3864391530 1.06973057

X67 0.2393118748 3.56359664

X68 -0.0384357561 0.17780345

X69 -1.6179196744 5.58552414

X70 0.2499318301 0.37113733

X71 -0.2216438461 0.88555053

X72 -1.4088850373 1.69588101

X73 -2.4681357886 1.54829462

X74 -0.6141170188 1.71494790

X75 -1.1120729389 1.55064079

X76 0.4933931976 0.22467962

X77 0.2176797022 4.19796876

X78 -3.0080931396 2.67227113

X79 -3.0421463526 3.19040422

X80 -0.3850179630 0.27636840

X81 1.9484953126 3.96916137

X82 2.4926329411 1.01178353

X83 0.2326136044 0.57177898

X84 2.0926700204 1.43522528

X85 0.9800670153 1.10561126

Residual Deviance: 2248.44

AIC: 2420.44

Model yang didapatkan berdasarkan nilai koefisien yang didapatkan adalah sebagai berikut :

X86 = 2,113 + 0,067 X1- 0,1705 X2+…+2,092 X84+0,980 X85

Dari model yang telah didapatkan diatas, kemudian dicari akurasinya dan didapatkan hasil sebagai berikut :

pred 0 1

0 5466 341

1 8 7

Dengan akurasi sebesar :

[1] 0.940055

Berikut adalah hasil perbandingan akurasi antara penggunaan metode regresi logistik klasik dan regresi logistic dengan pendekatan machine learning :

|  |  |
| --- | --- |
| **Metode** | **Akurasi** |
| Klasik | 94,00% |
| Pendekatan machine learning (Data Training) | 94,24% |
| Pendekatan Machine Learning(Data Testing) | 93,3% |