**Name:** Laavanya Ganesh

**UIN:** 654324917

**Profit and ROI Assignment**

**IDS 462 Statistical Software for Business**

**PROFIT VS ROI TRADEOFF**

|  |  |  |
| --- | --- | --- |
|  | **Max Profit /ROI%** | **Max ROI%/Profit** |
| **Logistic Regression** | 24,540/ 96 | 108/19,468 |
| **Neural Nets** | 24,672/104 | 109/25,852 |
| **CHAID** | 26,312/109 | 114/21,852 |

**CALCULATION for: If we choose the maximum profit of 26,312, how much of the maximum ROI do we loose**

(109-114)/114 = -4.4%

**CALCULATION for: If we choose the maximum ROI of 114, how much of the maximum Profit do we loose**

(21,852-26,312)/26,312 = -16.95%

So, from the above calculations, the best alternative would be to choose the CHAID model and maximize the profit at 26,312 with a 4 to 5 % loss in ROI.

**GRAPHS FOR MAIL DEPTH VS PROFITABILITY AND MAIL DEPTH VS ROI% FOR THE THREE MODELS: LOGISTIC REGRESSION, CHAID AND NEURAL NETS**

**Program**

/\*\*Adjust Library Name \*\*/

\*libname Lec7 "C:\Documents and Settings\jspark4\Desktop";

\*libname Lec7 "D:\School\IDS462\Fall2007\Lectures\Lecture7";

libname Lec7 "/folders/myfolders";

data holdout;

set Lec7.ChaidOutput;

if respholdout>.;

run;

proc sort data=holdout;

by descending VR\_RESP;

run;

data RespAnalCHAID (keep=reccount respholdout cumcount cumresp);

set holdout;

reccount=1;

cumcount+reccount;

cumresp+respholdout;

run;

data RespAnalCHAID;

set RespAnalCHAID;

RespPct=cumresp/106;

CountPct=cumcount/1806;

run;

data cutpoint;

set RespAnalCHAID;

lagCountPct=lag(CountPct);

if CountPct ge .75 and lagCountPct lt .75 then output;

run;

proc print data=cutpoint;

run;

data holdout;

set Lec7.NNOutput;

if respholdout>.;

run;

proc sort data=holdout;

by descending VN\_RESP;

run;

data RespAnalNN (keep=reccount respholdout cumcount cumresp);

set holdout;

reccount=1;

cumcount+reccount;

cumresp+respholdout;

run;

data RespAnalNN;

set RespAnalNN;

RespPct=cumresp/106;

CountPct=cumcount/1806;

run;

data cutpoint;

set RespAnalNN;

lagCountPct=lag(CountPct);

if CountPct ge .75 and lagCountPct lt .75 then output;

run;

proc print data=cutpoint;

run;

/\*Assumes that RespAnalLR has been produced from a Logistic Regression model \*/

data compare;

merge Lec7.RespAnalLR (keep=cumresp rename=(cumresp=LRCumResp))

RespAnalNN (keep=cumresp rename=(cumresp=NNCumResp))

RespAnalChaid (keep=cumresp rename=(cumresp=CHCumResp))

;

mailed=12\*\_n\_;

Vcost = mailed;

Fcost=11000;

Tcost=Vcost+Fcost;

RevLR=LRCumResp\*500;

RevCH=CHCumResp\*500;

RevNN=NNCumResp\*500;

PrLR=RevLR-Tcost;

PrCH=RevCH-Tcost;

PrNN=RevNN-Tcost;

RoiLR=PrLR/Tcost;

RoiCH=PrCH/Tcost;

RoiNN=PrNN/Tcost;

\*Profit and ROI calulations here;;

run;

\*\*Stop Here!!;

;

**Log**

1 OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

61

62

63

64

65 /\*\*Adjust Library Name \*\*/

66 \*libname Lec7 "C:\Documents and Settings\jspark4\Desktop";

67 \*libname Lec7 "D:\School\IDS462\Fall2007\Lectures\Lecture7";

68 libname Lec7 "/folders/myfolders";

NOTE: Libref LEC7 was successfully assigned as follows:

Engine: V9

Physical Name: /folders/myfolders

69

70

71 data holdout;

72 set Lec7.ChaidOutput;

NOTE: Data file LEC7.CHAIDOUTPUT.DATA is in a format that is native to another host, or the file encoding does not match the

session encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might reduce

performance.

73 if respholdout>.;

74 run;

NOTE: There were 5822 observations read from the data set LEC7.CHAIDOUTPUT.

NOTE: The data set WORK.HOLDOUT has 1806 observations and 31 variables.

NOTE: DATA statement used (Total process time):

real time 0.04 seconds

cpu time 0.03 seconds

75

76 proc sort data=holdout;

77 by descending VR\_RESP;

78 run;

NOTE: There were 1806 observations read from the data set WORK.HOLDOUT.

NOTE: The data set WORK.HOLDOUT has 1806 observations and 31 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.01 seconds

cpu time 0.00 seconds

79

80 data RespAnalCHAID (keep=reccount respholdout cumcount cumresp);

81 set holdout;

82 reccount=1;

83 cumcount+reccount;

84 cumresp+respholdout;

85 run;

NOTE: There were 1806 observations read from the data set WORK.HOLDOUT.

NOTE: The data set WORK.RESPANALCHAID has 1806 observations and 4 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.02 seconds

86

87 data RespAnalCHAID;

88 set RespAnalCHAID;

89 RespPct=cumresp/106;

90 CountPct=cumcount/1806;

91 run;

NOTE: There were 1806 observations read from the data set WORK.RESPANALCHAID.

NOTE: The data set WORK.RESPANALCHAID has 1806 observations and 6 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

92

93

94 data cutpoint;

95 set RespAnalCHAID;

96 lagCountPct=lag(CountPct);

97 if CountPct ge .75 and lagCountPct lt .75 then output;

98 run;

NOTE: There were 1806 observations read from the data set WORK.RESPANALCHAID.

NOTE: The data set WORK.CUTPOINT has 1 observations and 7 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

99

100 proc print data=cutpoint;

101 run;

NOTE: There were 1 observations read from the data set WORK.CUTPOINT.

NOTE: PROCEDURE PRINT used (Total process time):

real time 0.03 seconds

cpu time 0.04 seconds

102

103 data holdout;

104 set Lec7.NNOutput;

NOTE: Data file LEC7.NNOUTPUT.DATA is in a format that is native to another host, or the file encoding does not match the session

encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might reduce

performance.

105 if respholdout>.;

106 run;

NOTE: There were 5822 observations read from the data set LEC7.NNOUTPUT.

NOTE: The data set WORK.HOLDOUT has 1806 observations and 31 variables.

NOTE: DATA statement used (Total process time):

real time 0.05 seconds

cpu time 0.04 seconds

107

108 proc sort data=holdout;

109 by descending VN\_RESP;

110 run;

NOTE: There were 1806 observations read from the data set WORK.HOLDOUT.

NOTE: The data set WORK.HOLDOUT has 1806 observations and 31 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

111

112 data RespAnalNN (keep=reccount respholdout cumcount cumresp);

113 set holdout;

114 reccount=1;

115 cumcount+reccount;

116 cumresp+respholdout;

117 run;

NOTE: There were 1806 observations read from the data set WORK.HOLDOUT.

NOTE: The data set WORK.RESPANALNN has 1806 observations and 4 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

118

119 data RespAnalNN;

120 set RespAnalNN;

121 RespPct=cumresp/106;

122 CountPct=cumcount/1806;

123 run;

NOTE: There were 1806 observations read from the data set WORK.RESPANALNN.

NOTE: The data set WORK.RESPANALNN has 1806 observations and 6 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

124

125

126 data cutpoint;

127 set RespAnalNN;

128 lagCountPct=lag(CountPct);

129 if CountPct ge .75 and lagCountPct lt .75 then output;

130 run;

NOTE: There were 1806 observations read from the data set WORK.RESPANALNN.

NOTE: The data set WORK.CUTPOINT has 1 observations and 7 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

131

132 proc print data=cutpoint;

133 run;

NOTE: There were 1 observations read from the data set WORK.CUTPOINT.

NOTE: PROCEDURE PRINT used (Total process time):

real time 0.01 seconds

cpu time 0.02 seconds

134

135 /\*Assumes that RespAnalLR has been produced from a Logistic Regression model \*/

136

137 data compare;

138 merge Lec7.RespAnalLR (keep=cumresp rename=(cumresp=LRCumResp))

139 RespAnalNN (keep=cumresp rename=(cumresp=NNCumResp))

140 RespAnalChaid (keep=cumresp rename=(cumresp=CHCumResp))

141 ;

NOTE: Data file LEC7.RESPANALLR.DATA is in a format that is native to another host, or the file encoding does not match the session

encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might reduce

performance.

142 mailed=12\*\_n\_;

143 Vcost = mailed;

144 Fcost=11000;

145 Tcost=Vcost+Fcost;

146 RevLR=LRCumResp\*500;

147 RevCH=CHCumResp\*500;

148 RevNN=NNCumResp\*500;

149 PrLR=RevLR-Tcost;

150 PrCH=RevCH-Tcost;

151 PrNN=RevNN-Tcost;

152 RoiLR=PrLR/Tcost;

153 RoiCH=PrCH/Tcost;

154 RoiNN=PrNN/Tcost;

155 \*Profit and ROI calulations here;;

156 run;

NOTE: There were 1806 observations read from the data set LEC7.RESPANALLR.

NOTE: There were 1806 observations read from the data set WORK.RESPANALNN.

NOTE: There were 1806 observations read from the data set WORK.RESPANALCHAID.

NOTE: The data set WORK.COMPARE has 1806 observations and 16 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

157

158 \*\*Stop Here!!;

159 ;

160

161

162

163 OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

176