**Decision maker for Credit card issuance**

**Question 1 : What is your predictive model?**

1. **Describe the arithmetic clearly so that another learner could implement your model on new standardized input data if they wished**.
2. **Give an example of the score you would assign the following applicant, whether they would be approved or rejected for a credit card and why.**

Age: -0.06

Years at employer: 0.23

Years at address: -0.58

Income: -0.38

Credit card debt: 0.14

Auto debt: -0.06

**Answer 1(a) :** Model is as follows :

1. I have used Linest formula on the “**Training Set**” sheet of

“**NewData\_Final-Project.xls** “, and calculated the beta values using all the six parameters as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Age | Years at Employer | Years at Address | Income | Credit Card Debt | Automobile Debt |
| -0.02 | -0.19 | 0.03 | -0.08 | -0.21 | -0.08 |

1. We know that

Y=beta1\*x1 +beta2\*x2+…..beta6\*x6+y-intercept.

1. Using the above formula I have calculated the predicted y value for each data in col “S”of “**Training Set**” sheet of “**NewData\_Final-Project.xls** “.

**Answer 1(b)**

Z- Score for the customer is -0.05369 . Since it is below the threshold of 0.1 **, the customer can be given the credit card .** Any customer above .1 is predicted to be default.

**Question 2 : What would the bank’s *average profit per applicant* be (net profits divided by 200) when using your predictive model on the Training Set?**

**Answer 2 :**

1. To calculate the model’s AUC on the Training Set,
2. I have copied the

predicted y value (col “S”of “**Training Set**” sheet of “**NewData\_Final-Project.xls** “)

and pasted in

cells “C23” to “C222” of sheet “**AUC Calculator training**” of “**FP\_AUC\_Calculator-and-Review-of-AUC-Curve.xls**”.

1. Changed the value of cell “I4” from $5000 to $4900 as given in assignment
2. The optimum threshold on the training set to minimize the average cost per test is **0.1**
3. The average cost-per-event at the Training Set optimum threshold is **$729**
4. Using the same threshold of “0.1”, used on the training set, the cost per event on Test Set is **$658**
5. AUC is 0.84 for both the training set and test set
6. Given

Loss per unprofitable customer(FP) = -$4900

Profit per profitable customer is (TP)= $ 4000

Proportion of unprofitable customer is 25%

Perfect model excluding all unprofitable customers will give 0.75\*$4000 = $3000 per event

At a cost per default of $4900 (False Negative) doing no forecasting results in a cost-per-event of (.25\*$4900) = $1,225.

The **base rate cost-per-event is $1,225**.

Money bank saves, per event, using my model and its data-inputs, instead of issuing credit cards to everyone who asks at 25% default rate is

($2500\*0+$4900\*50)/200-$658

**Saving =$1225-$658 = $567**

Average profit per event while using no model is : 0.75\*$4000-0.25\*$4900 **= $1775**

At 0.1 threshold using test data

TP= 40 (which will not be given credit card)

FP=33 (which will not be given credit card)

FN=10 (which will be given credit card) and will default

So, 135 (which will be given credit card)

125\*4000-10\*4900= $451000

**Profit Per event=$407800/200= $2255**

**Question 3 : What is the incremental financial value per applicant of your model over no model on the Training Set?**

**Answer 3:**

**Incremental value =-$2255-$1775 =$480**

**Question 4 : Evaluate your model on the Test Set data. How confident are you that your model does not over-fit the Training Set data?**

1. Choose between three broad degrees of confidence: “very” “somewhat” or “not at all.” (Note that “not at all” is still an acceptable answer if you give persuasive reasons for why you chose this answer).
2. Explain the evidence your degree of confidence is based upon.
3. The beta values on training set are

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Age | Years at Employer | Years at Address | Income | Credit Card Debt | Automobile Debt |
| -0.02 | -0.19 | 0.03 | -0.08 | -0.21 | -0.08 |

The root-mean-square residual (the standard deviation of model error) on Standardized output for the Test Set is 0.9599

1. The observed correlation R on the Test Set 0.2803
2. The Standard deviation of model error, in Dollars, for the Test Set

= 0.9599\*$5755.91 = 5525.23

1. the 90% confidence interval, in dollars, for the Test Set is $9088.19 (above or below the estimate)
2. Percentage Information Gain (P.I.G.) on the Test Set = 2.9%
3. The Correlation, R, of my model on the Training Set is 0.8455
4. The standard deviation of model error in dollars on the Test Set is $ 9088.19
5. The standard deviation of my model error on the standardized Training Set output is 0.8602
6. The standard deviation of model error in dollars on the Training Set is $ 6100.60
7. the 90% confidence interval, in dollars, for the Training Set is $ 10034.60 (above or below the estimate)
8. the Percentage Information Gain (P.I.G.) on the Training Set? -28.6%
9. The point estimate of profitability, in dollars is 1596.45 (-0.06\*5755.91+1905.51)
10. With 50% confidence, the range of profitability

The mean of the interval is $1596.45 (from pt.12).

This is a left-sided 25% confidence interval.

The normsinv(p = .25) = -0.67448975.

The interval from p= .25 to p = .75 is +- plus or minus

(the standard deviation of error as a fraction of the standard deviation of profitability)\*(normsinv(.25))\*(standard deviation of profits)

= (0.9599)\*(0.67448975)\* ($5755.91) = $3726.62

The 50% confidence interval range is from ($3726.62+1596.45 ) to ( $3726.62-1596.45 ).

**$5323.07 to $2130.17**

1. With 50% confidence, the range of profitability

The mean of the interval is $1596.45 (from pt.12).

This is a left-sided 25% confidence interval.

The normsinv(p = .25) = -0.67448975.

The interval from p= .25 to p = .75 is

+- plus or minus (the standard deviation of error as a fraction of the standard deviation of profitability)\*(normsinv(.25))\*(standard deviation of profits)

= (0.9599)\*(0.67448975)\* ($5755.91) = $3726.62

The 50% confidence interval range is from ($ 1596.45 +$3726.62 ) to ( $1596.45- $3726.62 ).

**$5323.07 to - $2130.17**

1. With 99% confidence, the range of profitability?

The mean of the interval is $1596.45 (from pt.12).

The left-sided confidence interval is .5% or .005. The normsinv(p = .005) = -2.575829304.

The interval from p= .005 to p = .995 is:

+- plus or minus(the standard deviation of error as a fraction of the standard deviation of profits)\*(normsinv(p = .005))\*(standard deviation of profits)

= (0.9599)\*(2.575829304)\* ($5755.91)

= +- $14231.71

The range is from ($15828.16 to -$12635.26)

1. With 24% confidence, the range of profitability is

The mean of the interval is $1596.45 (from pt.12).

The left-sided confidence interval is 38% or 0.38. The normsinv(p = .38) = -0.305480788

The interval from p= .38 to p = .62 is:

+- plus or minus(the standard deviation of error as a fraction of the standard deviation of profits)\*(normsinv(p = .005))\*(standard deviation of profits)

= (0.9599)\*( 0.305480788)\* ($5755.91)

= +- $1620.168765

The range is from ($3216.62 to -$23.72)

1. Between Training Set and the Test Set, the dollar value of the standard deviation of model error

**decreased by 32%**

**From $9088.190to $6100.60**

As can be seen from point no. 16 the confidence level is at the max 24% beyond which negative values for range of profitability starts.

Thus this is not a good model, and we cannot solely rely on Linest function for regression analysis , and some other regression model needs to be applied , or drop some of the parameters etc to get better beta values.

With a 32% decrease between Training Set and the Test Set I think the model does not over fit. (Point 17)

Answer 4(a) I would say “Somewhat” (Point 16)

Answer 4(b) Since the confidence interval is 24% only, I say some what.

Answer 4 . Not overfitting.

Thanks.

|  |
| --- |
|  |
| - |

*Profitability = mean + std. profitability \* std. deviation*