**Credit limit strategy:**

1. **STRATEGY BASED ON APPROVAL BASIS:**

We only consider the approval status.

We allotted $8000 to everyone who was approved and $1000 to everyone who wasn’t.

Chart, line chart

Description automatically generated

*Fig 1: Variation of parameters with threshold*

|  | @Max revenue | @Max profit | @Optimal |
| --- | --- | --- | --- |
| P value | 0.1 | 0.32 | 0.54 |
| Revenue | 15.26 | 13.33 | 10.86 |
| Actual loss | 13.72 | 6.25 | 2.88 |
| Pseudo Loss | 0.79 | 2.72 | 5.2 |
| Profit | 0.75 | 4.34 | 2.77 |

**Conclusion:**

1. Revenue is very high. But the losses are also very high which results in poor profit.
2. From business point of view, the strategy is incorrect as it only considers the approval status.
3. **STRATEGY BASED ON CREDIBILITY OF APPROVAL (2 Thresholds).**

It considers the credibility of approval of individual.

We set a threshold based on individuals’ credibility.

P1 = 0.45 (Selected to have max profit)

Chart, line chart

Description automatically generated

*Fig 2: Variation of parameters with threshold*

|  | @Max revenue | @Optimal |
| --- | --- | --- |
| P2 value | .44 | .56 |
| Revenue | 11.864 | 10.8 |
| Actual loss | 3.96 | 2.8 |
| Pseudo Loss | 0.62 | 0.52 |
| Profit | 7.43 | 7.49 |

**Iterations:**

1. Direct mapping based on probability

Revenue: 9.8

Actual loss: 2.5

Pseudo loss: 1.1

Profit: 6.2

1. $ 1000 for people with p < 0.5, $ 8000 for people p > 0.8 and mapping others

Revenue: 9.9

Actual loss: 2.3

Pseudo loss: 0.5

Profit: 7.1

1. Creating different bins between 0.5 to 0.8 and allotting credit

Revenue: 9.5

Actual loss: 2.0

Pseudo loss: 0.5

Profit: 7.0

1. For data points in probability limits, considering most correlated features in that range and further binning:

Results: Similar as above

**Conclusion:**

1. Profit is very high as compared to approval-based strategy and revenues are comparable.
2. Credibility of an individual to get approval might not be credibility to return credit and since our model is trained for approval. The results are not reliable.
3. **USING INCOMES AND PROBABILITY COMBINED:**
4. We consider the income distribution based on following features:

Age, Occupation, education, relationship, marital status, work class.

1. As the exact income is unknown, we consider the net average income from these.
2. We consider the weighted sum, based on feature importance coefficients to determine an appropriate salary and then map it to (0,1) as salary score.
3. The credibility of a customer is also given importance.

Net score = w1\*prob + w2\*salary score

1. We map this score to (0,8000)

**Iterations for salary score:**

1. Mean of salaries: Here, the individual with a unsuitable category in a particular feature is not separated from a good individual because, a good feature compensates for bad and the majority of the population lie in the average salary range.

Revenue: 10.6

Actual loss: 4.3

Pseudo loss: 1.9

Profit: 4.4

1. Weighted sum with probability: This helps in compensating the above drawback by assigning the weights for most affecting feature.

Revenue: 9.9

Actual loss: 2.9

Pseudo loss: 0.39

Profit: 6.6

1. Weighted sum of squares: This gives similar results as weighted sum but the importance of weights decreases here. Hence, it is not a suitable strategy.

Revenue: 9.9

Actual loss: 2.8

Pseudo loss: 0.5

Profit: 6.59

1. Square of weighted sum: This enables a very good separation between users with lower salary and higher salary. However, on normalizing, it results in a very less credit distribution amongst less salaried users i.e. a low salary score.

Revenue: 9.9

Actual loss: 2.9

Pseudo loss: 0.5

Profit: 6.5

1. Multiplication of salaries:

This results in a similar problem as square of weighted sum.

Revenue: 6.5

Actual loss: 2

Pseudo loss: 0.46

Profit: 4.07

Best results were obtained for the weighted sum:

Chart, line chart

Description automatically generated

*Fig 3: Variation of parameters with threshold*

|  | Value |
| --- | --- |
| Revenue | 9.9M |
| Actual loss | 2.9M |
| Pseudo Loss | 0.39M |
| Profit | 6.6M |

**Iterations for models improvement and feature engg:**

1. Tried XGB, LGBM, Random Forest, Cat boost, logistic regression
2. Tried creating and adding different features by checking the correlations, statistic value with target variable.
3. Checked for email domain, State, City, Zip code.
4. Created new features for categories with very high and very low approval rates for e.g. Husband and wife in relationship, Exec managerial and prof specialty in occupation, Inquiry purpose code 7, Marital status etc.
5. Tried converting continuous variables to categorical like capital gain group, capital loss group, net capital gain group, working hour group, Age group.
6. Changed mathematical form of continuous variables by using log, sigmoid, squares, exponentials etc.
7. Tried using sigmoid function to separate low salary individual from high salaried individual.
8. Tried creating separate models for male and female.
9. Checked model performance based on different thresholds and different smoothening.
10. Tried different encoding schemes for different variables.