Direct Marketing Optimization Case Report

## Data Preprocessing

I have used **Pandas from python** package for reading and processing data file and created dataset by extracting all the required information required for the task given in one dataframe. Transforming data into meaningful arrays is most important as it will improve the model performance and help understand the insights of the data.

## Data Analysis

This problem can be analyzed with linear regression, logistic regression or SVM model. I have selected logistic regression as it is giving better accuracy. Given short time I did not use SVM but DAG SVM for multiclass analysis would have been better model for this case. Also, I have used all the given features corresponding to the model. Although this could have been improved by implementing feature selection methods like Principal component analysis (PCA) or Recursive feature elimination (RFE) limitation of time make forces me to skip it.

## Training and Testing Algorithm

The 75% of data is used to train the classifier. Then the remaining 25% is used to test the classifier. For accuracy and reliability, the best way is using the build in cross validation. The model is implemented for the task of predicting customers most likely to buy credit card mutual fund and loan. Here, the features are the descriptive attributes, and the label is what we are attempting to predict or forecast. I have built 6 models three for predicting probable customers for credit cards, mutual funds and consumer loans among the 645 customers who were not in the training set. The other three models were for predicting the projected revenue to those customers. Many of these customers could actually be clients of all three, two or just one product.

## Result Analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Product | #Clients | Confidence in % | Forecasted revenue | Confidence in % |
| Credit card (CC) | 39 | 74. 48 | 294 | 71.60 |
| Mutual fund (MF) | 20 | 86.41 | 387 | 80.65 |
| Consumer loan (CL) | 23 | 71.19 | 249 | 64.19 |
| Total | ~ 43 potential customers | - | 930 (units 1000‘ euros?) | - |

Table : Analysis results

As we can see the Table 1 confidence in predicting the consumer loan revenue model is less and should be further investigated to increase the chances of success. This is a strong case for PCA or RFE analysis. **Total projected revenue is 930**. There are total of 43 potential new customers projected by the model while few of them are target client 2 or more of MF, CC and CL.

Finally, list of potential customers are presented in Table 2.

|  |  |  |  |
| --- | --- | --- | --- |
| Client ID | Revenue\_CC | Revenue\_CL | Revenue\_MF |
| 196 | 0 | 0 | 73 |
| 1119 | 0 | 0 | 73 |
| 1455 | 25 | 30 | 0 |
| 1007 | 0 | 0 | 54 |
| 766 | 0 | 0 | 54 |
| 19 | 42 | 0 | 0 |
| 352 | 42 | 0 | 0 |
| 1516 | 0 | 0 | 34 |
| 1508 | 0 | 0 | 34 |
| 153 | 14 | 19 | 0 |
| 1218 | 0 | 31 | 0 |
| 1077 | 8 | 22 | 0 |
| 496 | 0 | 30 | 0 |
| 543 | 0 | 27 | 0 |
| 1051 | 0 | 22 | 0 |
| 217 | 0 | 20 | 0 |
| 1443 | 0 | 20 | 0 |
| 1414 | 18 | 0 | 0 |
| 940 | 0 | 0 | 18 |
| 314 | 0 | 17 | 0 |
| 382 | 15 | 0 | 0 |
| 1349 | 15 | 0 | 0 |
| 1076 | 15 | 0 | 0 |
| 587 | 15 | 0 | 0 |
| 535 | 15 | 0 | 0 |
| 373 | 15 | 0 | 0 |
| 359 | 14 | 0 | 0 |
| 1289 | 13 | 1 | 0 |
| 532 | 7 | 4 | 0 |
| 1226 | 0 | 0 | 9 |
| 1435 | 0 | 0 | 9 |
| 583 | 0 | 0 | 9 |
| 161 | 2 | 6 | 0 |
| 1416 | 0 | 0 | 6 |
| 30 | 0 | 0 | 6 |
| 506 | 0 | 0 | 6 |
| 851 | 6 | 0 | 0 |
| 951 | 5 | 0 | 0 |
| 1410 | 4 | 0 | 0 |
| 633 | 3 | 0 | 0 |
| 886 | 1 | 0 | 0 |
| 785 | 0 | 0 | 1 |
| 1008 | 0 | 0 | 1 |

Table : List of customers and their targeted revenue