



MICRO CREDIT LOAN PROJECT

Submitted by:

Kirtarath singh saini

ACKNOWLEDGMENT

All the references, research papers, data sources, professionals and other resources that helped and guided in completion of the project are

1. Google
2. youtube

INTRODUCTION

- **Business Problem Framing**

A Microfinance Institution (MFI) is an organization that offers financial services to low income populations. MFS becomes very useful when targeting especially the unbanked poor families living in remote areas with not much sources of income. The Microfinance services (MFS) provided by MFI are Group Loans, Agricultural Loans, Individual Business Loans and so on.

Many microfinance institutions (MFI), experts and donors are supporting the idea of using mobile financial services (MFS) which they feel are more convenient and efficient, and cost saving, than the traditional high-touch model used since long for the purpose of delivering microfinance services. Though, the MFI industry is primarily focusing on low income families and are very useful in such areas, the implementation of MFS has been uneven with both significant challenges and successes.

Today, microfinance is widely accepted as a poverty-reduction tool, representing \$70 billion in outstanding loans and a global outreach of 200 million clients.

We are working with one such client that is in Telecom Industry. They are a fixed wireless telecommunications network provider. They have launched various products and have developed its business and organization based on the budget operator model, offering better products at Lower Prices to all value conscious customers through a strategy of disruptive innovation that focuses on the subscriber.

They understand the importance of communication and how it affects a person's life, thus, focusing on providing their services

and products to low income families and poor customers that can help them in the need of hour.

They are collaborating with an MFI to provide micro-credit on mobile balances to be paid back in 5 days. The Consumer is believed to be defaulter if he deviates from the path of paying back the loaned amount within the time duration of 5 days. For the loan amount of 5 (in Indonesian Rupiah), payback amount should be 6 (in Indonesian Rupiah), while, for the loan amount of 10 (in Indonesian Rupiah), the payback amount should be 12 (in Indonesian Rupiah).

The sample data is provided to us from our client database. It is hereby given to you for this exercise. In order to improve the selection of customers for the credit, the client wants some predictions that could help them in further investment and improvement in selection of customers.

- **Conceptual Background of the Domain Problem**

Domain related concepts that will be useful for better understanding of the project is the knowledge of micro credit loan and the factors that influence the repayment of loan.

- **Review of Literature**

There are various factors that influence the micro credit loan repayment. frequency of data account recharged in last 90 days

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

micro credit loan defaulter data file was a classification problem as the label which was the target variable was of categorical type we had to build a model which can be used to predict in terms of a probability for each loan transaction, whether the customer will be paying back the loaned amount within 5 days of insurance of.

- Data Sources and their formats

we got data in the micro credit loan csv file which had 209593 rows × 37 columns.

It had 0 duplicate values.

It had no null values present in it.

- Data Preprocessing Done
 - a. Finding null values but were none of them.
 - b. Describing the data.
 - c. Encoding of the data so that is machine understandable.
 - d. Finding correlation of variables with the target one.
 - e. Checking outliers and removing them.
 - f. Checking skewness and removing it using yeo johnson method.
 - g. Balancing the data for creating model and further processing using smote.

- **Hardware and Software Requirements and Tools Used**
various libraries were used like
 1. Pandas – to read the data
 2. Numpy
 3. Seaborn /matplotlib.lib- for visualization and plotting graphs.
 4. Warnings – for unnecessary warnings.
 5. Sklearn preprocessing – for encoding the data we called label encoder from preprocessing.
 6. Power transformation- to deal with the skewness in data with yeo johnson method.
 7. accuracy_score- for the accuracy score.
 8. confusion_matrix,classification_report for classification report.
 9. train_test_split – For creating train and test split.
 10. cross_val_score- For finding cross val score.
 11. GridSearchCV- for hyper parameter tuning
 12. plot_roc_curve – for plotting roc curve.

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

This was a classification problem as the target variable was a categorical one so different type of classification algorithms were tried to and the best suited one was selected.

- Testing of Identified Approaches (Algorithms)

1. Decision tree classifier.
2. Knn classifier.
3. Random forest classifier.

- Run and Evaluate selected models

The screenshot shows a Jupyter Notebook titled 'micro credit loan' running on a local host. The interface includes a top bar with file and kernel management options, and a main area with code cells. The first cell shows the shape of the target variable, and the second cell displays the results of a Decision Tree Classifier, including accuracy and a confusion matrix. The third cell shows the beginning of a Random Forest Classifier implementation.

```
In [98]: tar_res_test.shape
Out[98]: (73373,)
```

Decision Tree Classifier

```
In [99]: from sklearn.tree import DecisionTreeClassifier
dt=DecisionTreeClassifier()
dt.fit(fea_res_train,tar_res_train)
pre_train=dt.predict(fea_res_test)
print('accuracy',accuracy_score(tar_res_test,pre_train)*100)
print(confusion_matrix(tar_res_test,pre_train))
print(classification_report(tar_res_test,pre_train))
```

accuracy 91.91664508742998
[[34023 2733]
 [3198 33419]]

	precision	recall	f1-score	support
0	0.91	0.93	0.92	36756
1	0.92	0.91	0.92	36617
accuracy			0.92	73373
macro avg	0.92	0.92	0.92	73373
weighted avg	0.92	0.92	0.92	73373

Random Forest Classifier

Home Page - Select or create a notebook | micro credit loan - Jupyter Notebook | +

localhost:8888/notebooks/micro%20credit%20loan.ipynb

Gmail YouTube Displaying 2022083... Maps

jupyter micro credit loan Last Checkpoint: Last Wednesday at 7:24 PM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Random Forest Classifier

```
In [100]: from sklearn.ensemble import RandomForestClassifier
rf=RandomForestClassifier(n_estimators=300)
rf.fit(fea_res_train,tar_res_train)
pre_train=rf.predict(fea_res_test)
print('accuracy',accuracy_score(tar_res_test,pre_train)*100)
print(confusion_matrix(tar_res_test,pre_train))
print(classification_report(tar_res_test,pre_train))
```

accuracy 95.5610374388399
[[35314 1442]
 [1815 34802]]

	precision	recall	f1-score	support
0	0.95	0.96	0.96	36756
1	0.96	0.95	0.96	36617
accuracy			0.96	73373
macro avg	0.96	0.96	0.96	73373
weighted avg	0.96	0.96	0.96	73373

K Neighbors classifier

```
In [101]: from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier()
knn.fit(fea_res_train,tar_res_train)
```

Home Page - Select or create a notebook | micro credit loan - Jupyter Notebook | +

localhost:8888/notebooks/micro%20credit%20loan.ipynb

Gmail YouTube Displaying 2022083... Maps

jupyter micro credit loan Last Checkpoint: Last Wednesday at 7:24 PM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

K Neighbors classifier

```
In [101]: from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier()
knn.fit(fea_res_train,tar_res_train)
pre_train=knn.predict(fea_res_test)
print('accuracy',accuracy_score(tar_res_test,pre_train)*100)
print(confusion_matrix(tar_res_test,pre_train))
print(classification_report(tar_res_test,pre_train))
```

accuracy 88.85421067695202
[[36486 270]
 [7908 28709]]

	precision	recall	f1-score	support
0	0.82	0.99	0.90	36756
1	0.99	0.78	0.88	36617
accuracy			0.89	73373
macro avg	0.91	0.89	0.89	73373
weighted avg	0.91	0.89	0.89	73373

cross validation score

```
In [102]: from sklearn.model_selection import cross_val_score
scr=cross_val_score(dt,fea_res,tar_res,cv=5)
print('cross validation score for Decision tree model =', scr.mean())
```

Home Page - Select or create a notebook | micro credit loan - Jupyter Notebook | +

localhost:8888/notebooks/micro%20credit%20loan.ipynb

Gmail YouTube Displaying 2022083... Maps

jupyter micro credit loan Last Checkpoint: Last Wednesday at 7:24 PM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

This is the cross val score of all the three.

The screenshot shows a Jupyter Notebook titled 'micro credit loan' running on a local host. The notebook displays a table of accuracy metrics and three code cells for cross-validation.

	accuracy	macro avg	weighted avg
Decision tree model	0.89	0.91	0.91
Random Forest Classifier	0.89	0.89	0.89
K Neighbors classifier	0.89	0.89	0.89

cross validation score

```
In [102]: from sklearn.model_selection import cross_val_score
scr=cross_val_score(dt,fea_res,tar_res,cv=9)
print('cross validation score for Decision tree model =', scr.mean())
cross validation score for Decision tree model = 0.830061855275453

In [103]: from sklearn.model_selection import cross_val_score
scr=cross_val_score(rf,fea_res,tar_res,cv=9)
print('cross validation score for Random Forest Classifier model =', scr.mean())
cross validation score for Random Forest Classifier model = 0.8787884954086792

In [104]: from sklearn.model_selection import cross_val_score
scr=cross_val_score(knn,fea_res,tar_res,cv=9)
print('cross validation score for K Neighbors classifiermodel =', scr.mean())
cross validation score for K Neighbors classifiermodel = 0.8579139840764978
```

Random Forest Classifier is performing best among all the other models so we'll continue with it

- **Key Metrics for success in solving problem under consideration**
Among all the models Random Forest classifier was considered to be the best one as it was giving the best accuracy score as you can see in the snapshot and the best cross val score too.
- **Interpretation of the Results**
While processing we get to know that highest correlated column is the frequency of recharge in 90 days. The best fit model was random forest classifier with the accuracy score of 95.5610374388399 and cross val score of 0.8787884954086792.

CONCLUSION

- Key Findings and Conclusions of the Study

Here we find that it was a classification problem as the target variable (label) is categorical one and the best While processing we get to know that highest correlated column is the frequency of recharge in 90 days. The best fit model was random forest classifier with the accuracy score of 95.5610374388399 and cross val score of 0.8787884954086792.

- Learning Outcomes of the Study in respect of Data Science

Learnings obtained about the power of visualization, data cleaning and various algorithms used is that it makes the data smooth and easily understandable . visualization helps in understanding the columns or the independent variable more easily that how they influence the target variable or the outcome. Data cleaning cleans the data remove outliers and null values so that the further processing is smooth. Various algorithms of classification were applied as it was a classification problem among them best one was the random forest classifier with the accuracy score of 95.5610374388399 and cross val score of 0.8787884954086792. jupyter notebook or the software wasn't working it was not loading the output it was very difficult to complete the project.