# MICRO CREDIT LOAN PRIDICTION

## Submitted by:

NEETAL TIWARI

# ACKNOWLEDGMENT

I would like to express my special thanks of gratitude to my SME (Mohd Kashif) as well as my company (flip Robo Technologies) who gave me the golden opportunity to do this wonderful project on the (Micro Credit Loan Use Case), which also helped me to in doing lots of research and I came to know about so many things I am really thank to them.

# INTRODUCTION

## Business Problem Framing

Micro finance plays institutions a major role in economic development in many developing countries. However, many of these microfinance institutions are faced with the problem of default because of the non-formal nature of the business and individuals they lend money to. This study seeks to find the determinants of credit default in microfinance institution.

## Conceptual Background of the Domain Problem

Microfinance is the provision of thrift, credit and other financial services and products of very small amounts to the poor for enabling them to raise their income levels and improve their living standards. It has been recognised that micro finance helps the poor people meets their needs for small credit and other financial service. The informal and flexible service offered to low-income borrowers for meeting their modest consumption and livelihood needs have not only made micro finance movement grow at a rapid pace across the world, but in turn has also impacted the lives of millions of poor positively.

## Review of Literature

Seibel and parhusip (1990) mention that this approach was based on the premise that rural micro- entrepreneurs are unable to organize themselves they need sub need subsidized credit for increasing their income and are too poor to save. Manisha raj, in his research paper entitled “Microfinance Institutions in India and its Legal Aspects” states that microfinance institutions have been proved a very important financial wing to incorporate the poor in the financial sector. Now on the other aspect like the challenges faced by the microfinance institutions Jonathan Morduch and Stuart Rutherford in this study “Microfinance: analytical issues for India” states that the microfinance movement is thus striving to match the convenience and flexibility of the informal sector, while adding reliability and the promise of continuity and in some countries, it is already doing this on a significant scale.

## Motivation for the Problem Undertaken

Microfinance increases the poor’s access to finance and to financial markets and is as such part of a democratization of finance.

The idea behind the concept is to help people at the grassroots level by providing them small loans, with the goal of allowing them to become self- sufficient. Microcredit facilities poor people to borrow money without collateral, which in turn helps them to start their own business.

# Analytical Problem Framing

## Mathematical/ Analytical Modelling of the Problem

In this work we will develop a new approach to solve the non-repayment problem in microfinance due to the problem of asymmetric information. This approach is based on modelling and simulation of ordinary differential system where time remains a primordial component, they thus enable microfinance institution to manage their risk portfolios by a prediction of number of solvent and insolvent borrows ever a period, in order to define or redefine its development strategy, investment and management in an area, where the population is often poor and in need a mechanism of financial inclusion.

## Data Sources and their formats

21 float 10 int and 6 object data types in micro credit project.

```
In [7]: mcp.keys()

Out[7]: Index(['Unnamed: 0', 'label', 'msisdn', 'aon', 'daily_decr30', 'daily_decr90',
       'rental30', 'rental90', 'last_rech_date_ma', 'last_rech_date_da',
       'last_rech_amt_ma', 'cnt_ma_rech30', 'fr_ma_rech30',
       'sumamnt_ma_rech30', 'medianamnt_ma_rech30', 'medianmarechprebal30',
       'cnt_ma_rech90', 'fr_ma_rech90', 'sumamnt_ma_rech90',
       'medianamnt_ma_rech90', 'medianmarechprebal90', 'cnt_da_rech30',
       'fr_da_rech30', 'cnt_da_rech90', 'fr_da_rech90', 'cnt_loans30',
       'amnt_loans30', 'maxamnt_loans30', 'medianamnt_loans30', 'cnt_loans90',
       'amnt_loans90', 'maxamnt_loans90', 'medianamnt_loans90', 'payback30',
       'payback90', 'pcircle', 'pdate'],
      dtype='object')
```

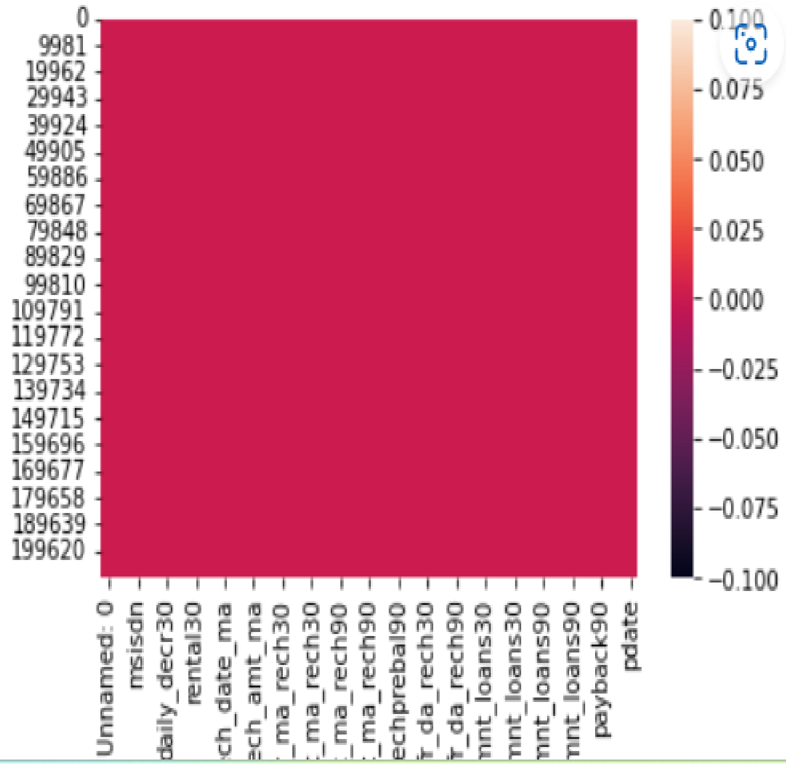
There are 37 columns and 209593 rows in microcredit loan case project. Here is he data description of the features.

| Variable | Definition | Com |
|---|---|---|
| label | Flag indicating whether the user paid back the credit amount within 5 days of issuing the loan{1:success, 0:failure} | |
| msisdn | mobile number of user | |
| aon | age on cellular network in days | |
| daily_decr30 | Daily amount spent from main account, averaged over last 30 days (in Indonesian Rupiah) | |
| daily_decr90 | Daily amount spent from main account, averaged over last 90 days (in Indonesian Rupiah) | |
| rental30 | Average main account balance over last 30 days | Uns |
| rental90 | Average main account balance over last 90 days | Uns |
| last_rech_date_ma | Number of days till last recharge of main account | |
| last_rech_date_da | Number of days till last recharge of data account | |
| last_rech_amt_ma | Amount of last recharge of main account (in Indonesian Rupiah) | |
| cnt_ma_rech30 | Number of times main account got recharged in last 30 days | |
| fr_ma_rech30 | Frequency of main account recharged in last 30 days | Uns |
| sumamnt_ma_rech30 | Total amount of recharge in main account over last 30 days (in Indonesian Rupiah) | |
| medianamnt_ma_rech30 | Median of amount of recharges done in main account over last 30 days at user level (in Indonesian Rupiah) | |
| medianmarechprebal30 | Median of main account balance just before recharge in last 30 days at user level (in Indonesian Rupiah) | |
| cnt_ma_rech90 | Number of times main account got recharged in last 90 days | |
| fr_ma_rech90 | Frequency of main account recharged in last 90 days | Uns |
| sumamnt_ma_rech90 | Total amount of recharge in main account over last 90 days (in Indonasian Rupiah) | |
| medianamnt_ma_rech90 | Median of amount of recharges done in main account over last 90 days at user level (in Indonasian Rupiah) | |
| medianmarechprebal90 | Median of main account balance just before recharge in last 90 days at user level (in Indonasian Rupiah) | |
| cnt_da_rech30 | Number of times data account got recharged in last 30 days | |
| fr_da_rech30 | Frequency of data account recharged in last 30 days | |
| cnt_da_rech90 | Number of times data account got recharged in last 90 days | |
| fr_da_rech90 | Frequency of data account recharged in last 90 days | |
| cnt_loans30 | Number of loans taken by user in last 30 days | |
| amnt_loans30 | Total amount of loans taken by user in last 30 days | |
| maxamnt_loans30 | maximum amount of loan taken by the user in last 30 days | The |
| medianamnt_loans30 | Median of amounts of loan taken by the user in last 30 days | |
| cnt_loans90 | Number of loans taken by user in last 90 days | |
| amnt_loans90 | Total amount of loans taken by user in last 90 days | |
| maxamnt_loans90 | maximum amount of loan taken by the user in last 90 days | |
| medianamnt_loans90 | Median of amounts of loan taken by the user in last 90 days | |
| payback30 | Average payback time in days over last 30 days | |
| payback90 | Average payback time in days over last 90 days | |
| pcircle | telecom circle | |
| pdate | date | |

## **Data Pre-processing Done**

There are no null values in microcredit loan case project.

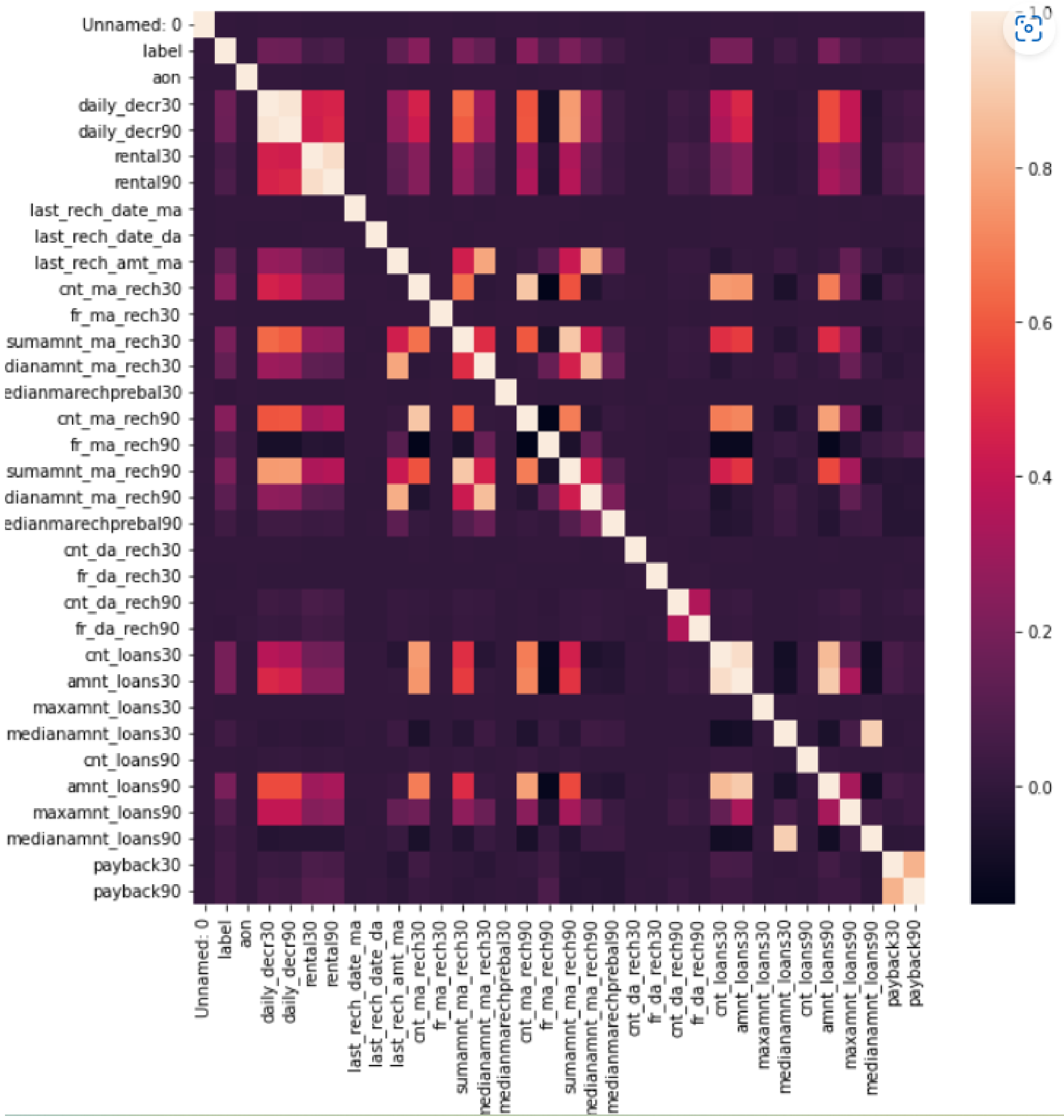
```
In [9]: sns.heatmap(mcp.isnull())

Out[9]: <AxesSubplot: >
```




In microcredit loan case project in many features data are skewed. With the help of sklearn.preprocessing.power_tranform, I had removed the skewness from the dataset.

47335 outliers present in microcredit loan case Dataset, with the help of scipy.states.zscore I had removed the outliers from the Dataset.

# Data Inputs- Logic- Output Relationships




We can see by the graph how all features are correlated with the y label. Some features are positively correlated and some features are very less correlated with the y.

daily_decr30(daily amount spent from main account, over last 30 days), daily_decr90(daily amount spent from main account, over last 90 days), rental30(average main account balance over last 30 days) and rental90(average main account balance over last 90 days) features are positively correlated, Thou we know that if someone take a micro credit loan or any type of loan so it is very important that how much money they are spending and how much money they have in their account respected to pay the loan back,

And another way unnamed, last_rech_date_ma (number of days till last recharge amount), msisdn (mobile number of user) features are negative correlated with y label.

Because mobile number of user and number of days till recharge are not so important respected to pay the loan back,

So those who need the loan and who can pay back the loan, only those people can get micro credit loan.this is how these all positive and negative correlations affect the y label.

# Model/s Development and Evaluation

## Testing of Identified Approaches (Algorithms)

Micro Credit Loan Case prediction is a classifier problem, where we have to predict that people who had taken the micro credit loan have paid the amount or not.

To predict MicroCreditLoanCase project I had used four algorithms such as RandomForestClassifier, DecisionTreeClassifier, KNeighborsClassifier, and SVC.

# Run and evaluate selected models

## 1- Random Forest classifier

In [174]:
```
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
rfcpred=rfc.predict(x_test)
print(accuracy_score(y_test,rfcpred)*100)
print(classification_report(y_test,rfcpred))
print(confusion_matrix(y_test,rfcpred))
```

```
90.79871810674226
              precision    recall  f1-score   support

           0       0.76      0.48      0.59      6701
           1       0.92      0.98      0.95     41977

    accuracy                           0.91     48678
   macro avg       0.84      0.73      0.77     48678
weighted avg       0.90      0.91      0.90     48678

[[ 3244  3457]
 [ 1022 40955]]
```

In Random Forest classifier accuracy score is 90.798%.

## 2- Decision Tree classifier

In [175]:
```
dtc=DecisionTreeClassifier()
dtc.fit(x_train,y_train)
dtcpred=dtc.predict(x_test)
print(accuracy_score(y_test,dtcpred)*100)
print(classification_report(y_test,dtcpred))
print(confusion_matrix(y_test,dtcpred))
```

```
85.80262130736678
              precision    recall  f1-score   support

           0       0.49      0.53      0.50      6701
           1       0.92      0.91      0.92     41977

    accuracy                           0.86     48678
   macro avg       0.70      0.72      0.71     48678
weighted avg       0.86      0.86      0.86     48678

[[ 3521  3180]
 [ 3731 38246]]
```

In Decision Tree classifier accuracy score is 85.802%.

## 3- KNeighbours classifier

```
In [176]: knn=KNeighborsClassifier()
          knn.fit(x_train,y_train)
          predknn=knn.predict(x_test)
          print(accuracy_score(y_test,predknn)*100)
          print(confusion_matrix(y_test,predknn))
          print(classification_report(y_test,predknn))
```

```
87.46456304696166
[[ 2558  4143]
 [ 1959 40018]]
              precision    recall  f1-score   support

           0       0.57      0.38      0.46      6701
           1       0.91      0.95      0.93     41977

    accuracy                           0.87     48678
   macro avg       0.74      0.67      0.69     48678
weighted avg       0.86      0.87      0.86     48678
```

---

In KNeighbors classifier accuracy score is 87.464%.

## 4-SVC

```
In [177]: svc=SVC()
          svc.fit(x_train,y_train)
          svcpred=svc.predict(x_test)
          print(accuracy_score(y_test,svcpred)*100)
          print(classification_report(y_test,svcpred))
          print(confusion_matrix(y_test,svcpred))
```

```
86.43740498787955
              precision    recall  f1-score   support

           0       0.68      0.03      0.05      6701
           1       0.87      1.00      0.93     41977

    accuracy                           0.86     48678
   macro avg       0.77      0.51      0.49     48678
weighted avg       0.84      0.86      0.81     48678

[[  187  6514]
 [   88 41889]]
```

---

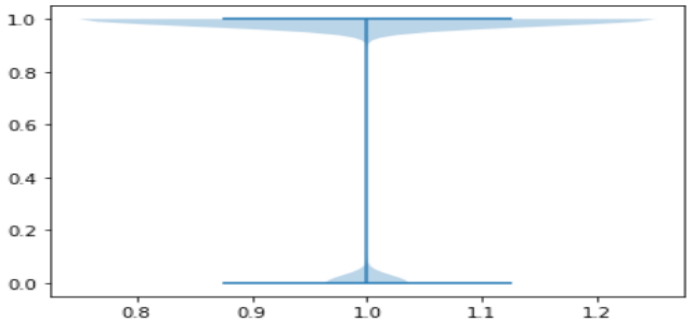
In SVC accuracy score is 86.437%.

## Visualizations

For visualization I had used matplotlib.pyplot and seaborn modules.

```
In [20]: mcp["label"].value_counts()
Out[20]: 1     183431
         0      26162
         Name: label, dtype: int64
```

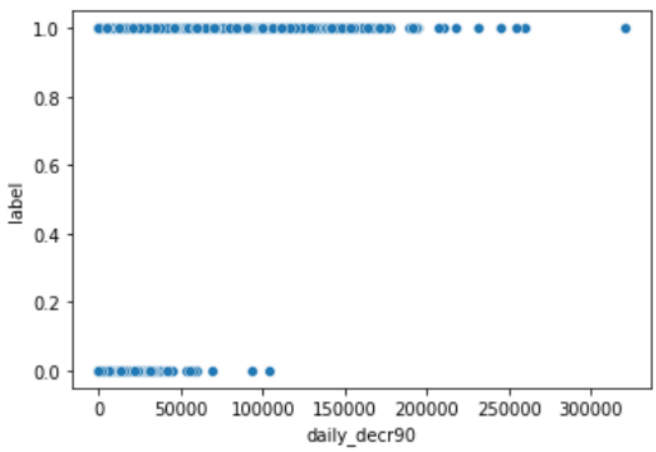
```
In [21]: plt.violinplot(mcp["label"])
Out[21]: {'bodies': [<matplotlib.collections.PolyCollection at 0x213c385b760>],
          'cmaxes': <matplotlib.collections.LineCollection at 0x213c2bda070>,
          'cmins': <matplotlib.collections.LineCollection at 0x213c385bd60>,
          'cbars': <matplotlib.collections.LineCollection at 0x213c3762100>}
```




label means (Flag indicating whether the user paid back the credit amount within 5 days of issuing the loan {1: success, 0: failure})

So out of 209593 people, 183431 people have success to take the credit amount and 26162 people have failure to take the credit amount.

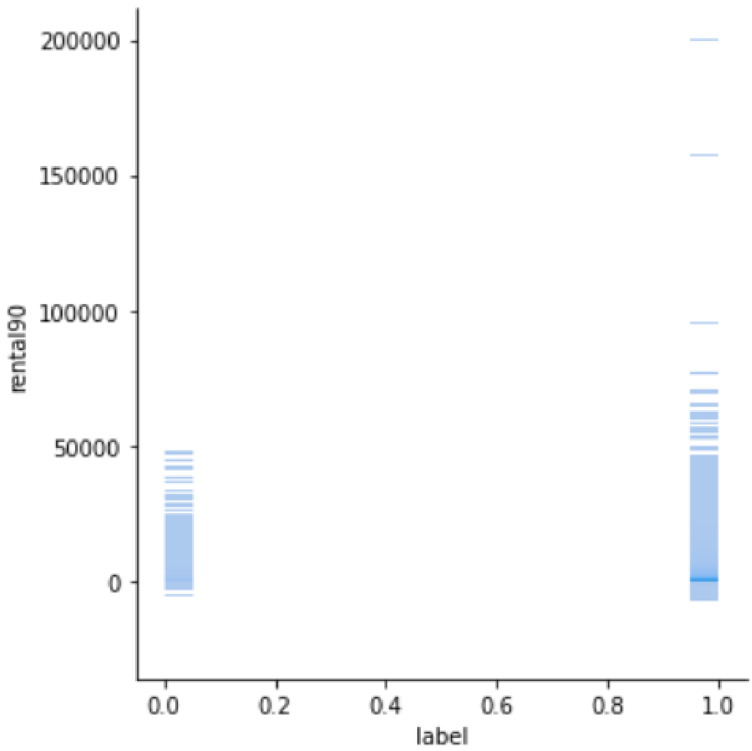
```
In [29]: sns.scatterplot(x="daily_decr90",y="label",data=mcp)
Out[29]: <AxesSubplot: xlabel='daily_decr90', ylabel='label'>
```



1 type of people have 0 to 250000 daily_decr90(Daily amount spent from main account, averaged over last 90 days (in Indonesian Rupiah))

and 0 type of people have 0 to 80000 daily_decr90(Daily amount spent from main account, averaged over last 90 days (in Indonesian Rupiah).

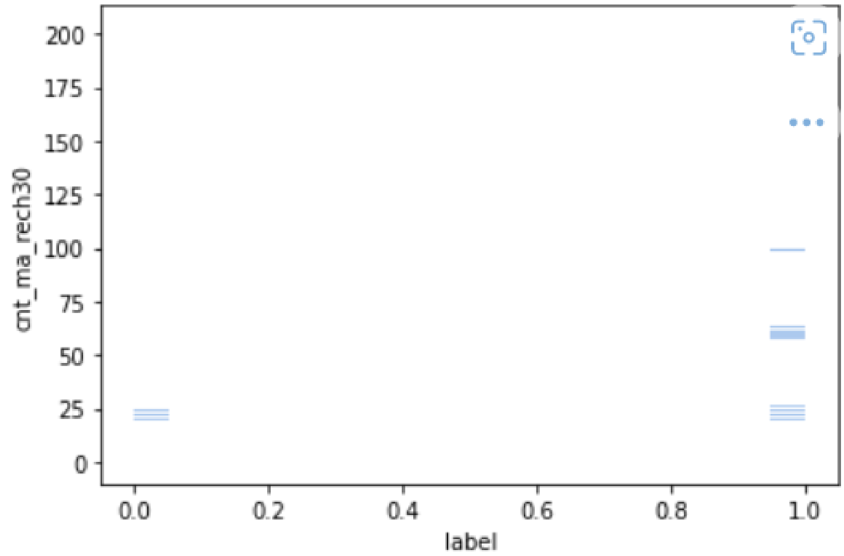
```
In [36]: sns.displot(y="rental90",x="label",data=mcp)

Out[36]: <seaborn.axisgrid.FacetGrid at 0x213cfab93d0>
```




1(success) people have 0 to 80000 rental90(Average main account balance over last 90 days) and 0(failure) people have 0 to 50000 rental90(Average main account balance over last 90 days).

In [46]:
```
sns.histplot(x="label",y="cnt_ma_rech30",data=mcp)
```

Out[46]: <AxesSubplot: xlabel='label', ylabel='cnt_ma_rech30'>

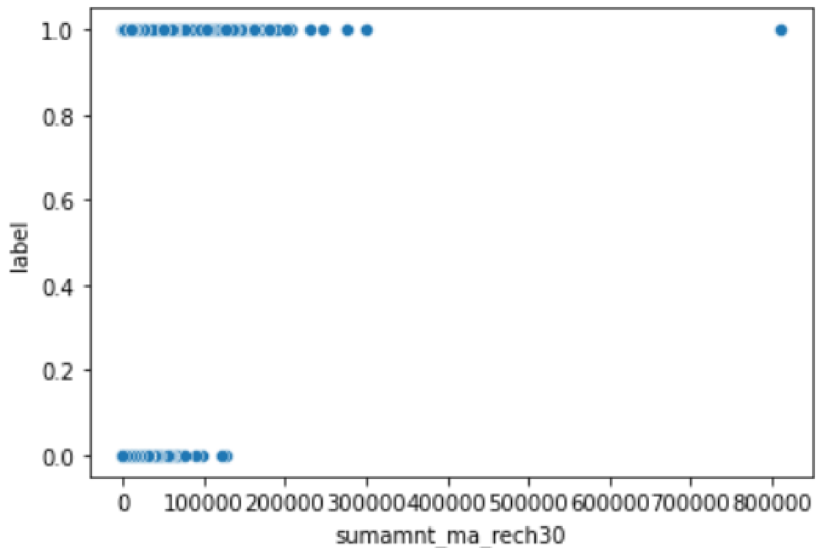



1(success), cnt_ma_rech30(Number of times main account got recharged in last 30 days) data is high in between 0 to 100.

and 0(failure), cnt_ma_rech30(Number of times main account got recharged in last 30 days) data is high in between 0 to 25.

In [52]:
```
sns.scatterplot(x="sumamnt_ma_rech30",y="label",data=mcp)
```

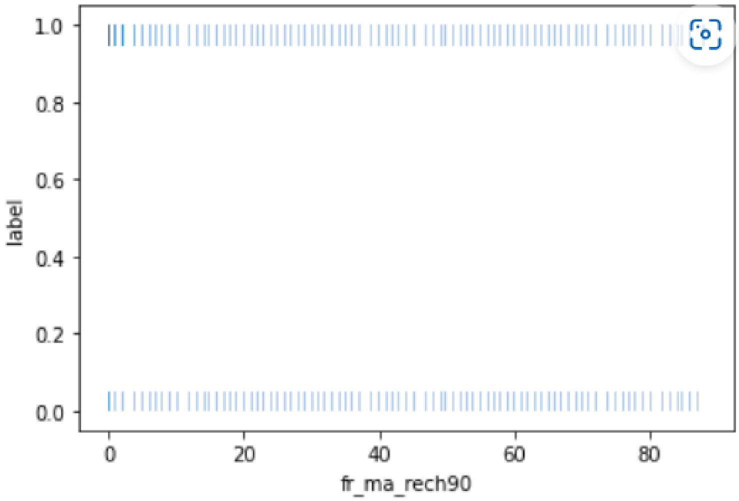
Out[52]: <AxesSubplot: xlabel='sumamnt_ma_rech30', ylabel='label'>



0(failure), sumamnt_ma_rech30(Total amount of recharge in main account over last 30 days (in Indonesian Rupiah)) data is high in between 0 to 10000. 1(success), sumamnt_ma_rech30(Total amount of recharge in main account over last 30 days (in Indonesian Rupiah)) data is high in between 0 to 300000.

```
In [63]: sns.histplot(x="fr_ma_rech90",y="label",data=mcp)
```

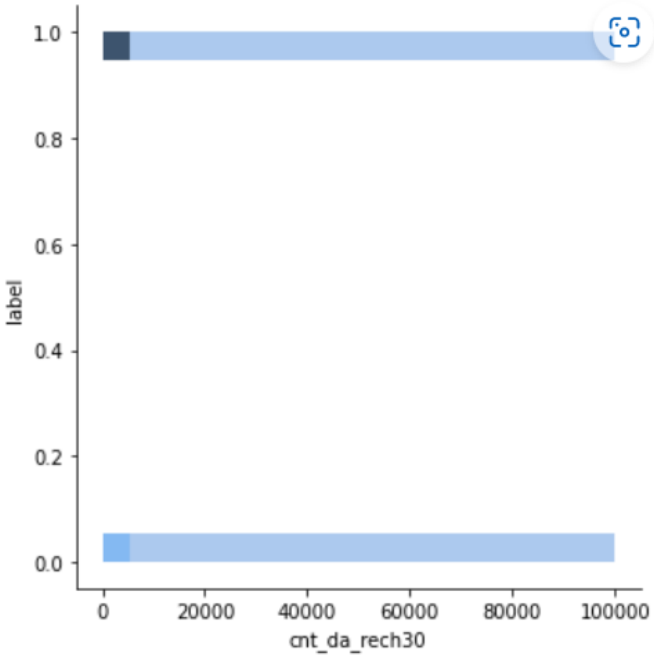
```
Out[63]: <AxesSubplot: xlabel='fr_ma_rech90', ylabel='label'>
```




frequency of fr_ma_rech90(Frequency of main account recharged in last 90 days) data is same in both 1 and 0 type of label.

```
In [75]: sns.displot(x="cnt_da_rech30",y="label",data=mcp)
```

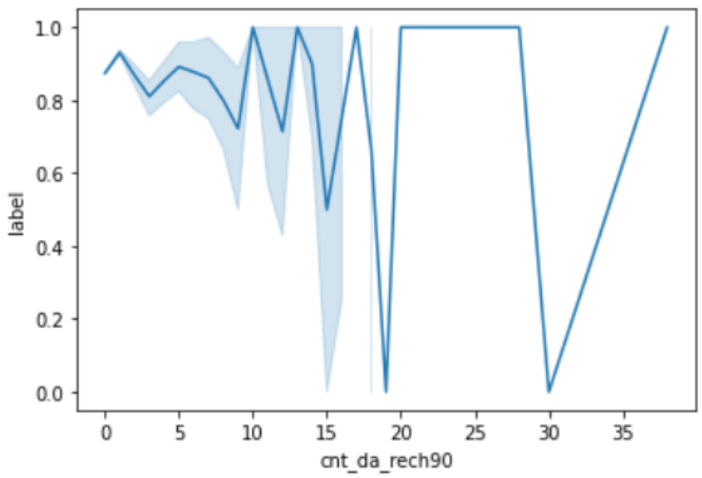
Out[75]: <seaborn.axisgrid.FacetGrid at 0x213fe05c3d0>




1(success) and 0(failure) both have almost same, cnt_da_rech30(Number of times data account got recharged in last 30 days) data.

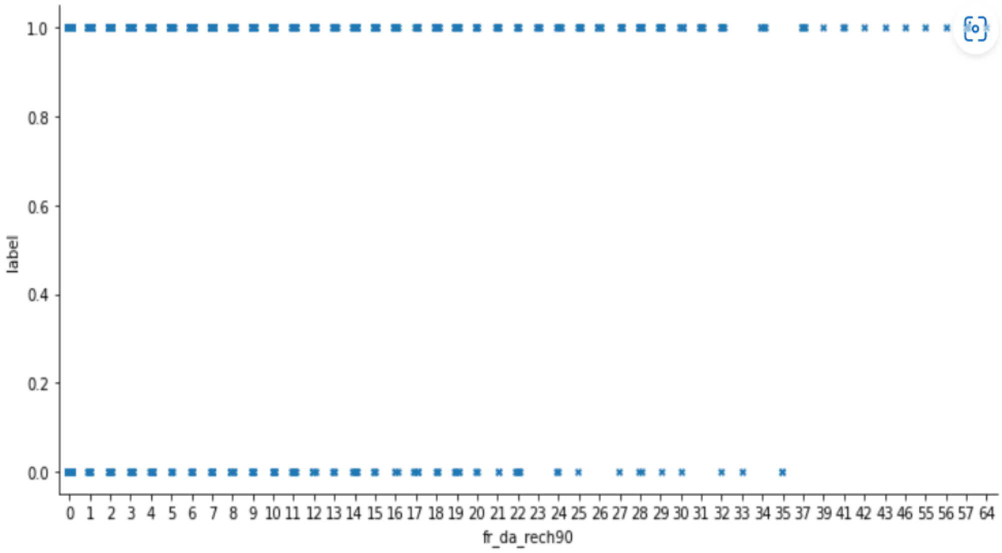
```
In [81]: sns.lineplot(x="cnt_da_rech90",y="label",data=mcp)
```

Out[81]: <AxesSubplot: xlabel='cnt_da_rech90', ylabel='label'>



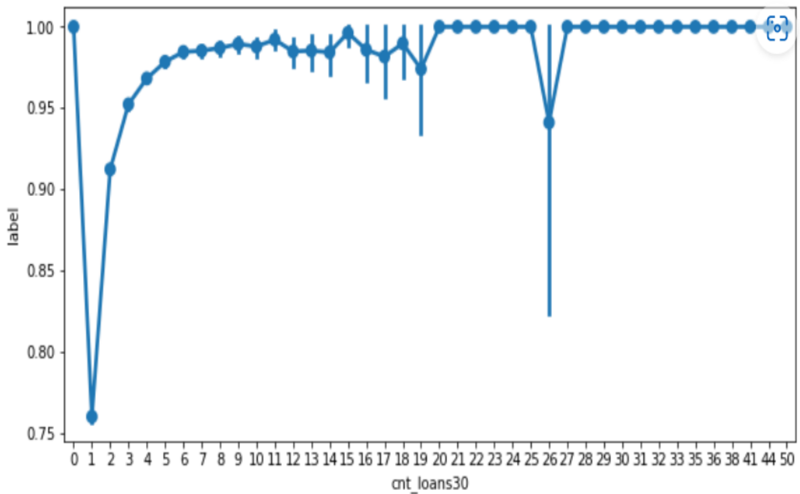

1(success), cnt_da_rech90(Number of times data account got recharged in last 90 days) is higher than 0(failure).

```
In [83]: sns.catplot(x="fr_da_rech90",y="label",data=mcp,aspect=2,marker="X")

Out[83]: <seaborn.axisgrid.FacetGrid at 0x213c2bed3a0>
```



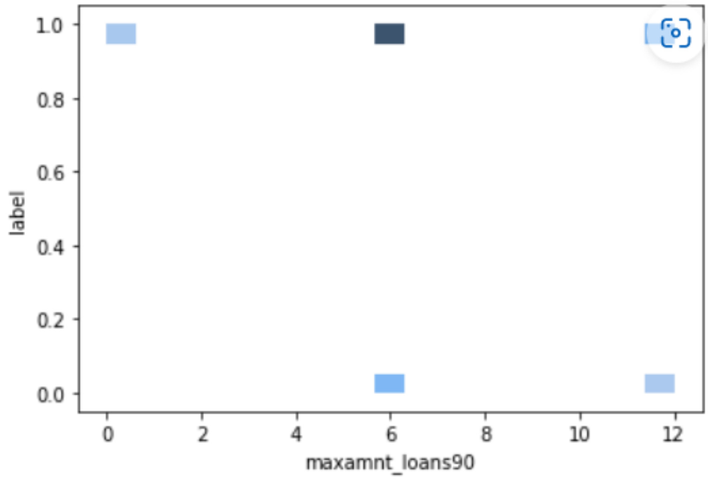

0(failure), fr_da_rech90(Frequency of data account recharged in last 90 days) is in between 0 to 35 and 1(success), fr_da_rech90(Frequency of data account recharged in last 90 days) is in between 0 to 64.

```
In [86]: plt.figure(figsize=(10,5))
         sns.pointplot(x="cnt_loans30",y="label",data=mcp)

Out[86]: <AxesSubplot: xlabel='cnt_loans30', ylabel='label'>
```




cnt_loans30(Number of loans taken by user in last 30 days) data is higher in 1(success) type of label.

In [104]:
```
sns.histplot(x="maxamnt_loans90",y="label",data=mcp)
```
Out[104]: <AxesSubplot: xlabel='maxamnt_loans90', ylabel='label'>




Data of maxamnt_loans90(maximum amount of loan taken by the user in last 90 days) is higher in 1(success).

# CONCLUSION

## Key Findings and Conclusions of the Study

The importance of microfinance in the developing countries like India cannot be undermined it play a vital role for socio-economic upliftment of poor and low-income peoples. Since 1990, poverty reduction has taken priority at both nation and international developments levels. Within this framework, various initiatives have been taken by government. Microfinance has caught the attention as an effective tool for poverty reduction and socio- economic development.

Hence, Microfinance can play a vital role for improving the standard of living of poor. The economic development of any country is severely influenced by the availability of financial service. Microfinance is the form of a board range of financial service such as deposits, loans, payment service, money transfer, insurance, saving, micro credit etc. to the poor and low-income individuals.