PROJECT NAME:

MICRO CREDIT LOAN USE CASE

SUBMITTED BY:

LALBIAK ZAUVA

Problem Statement:

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).

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.

Analysis:

- There are no null values in the dataset.
- All the data are int64 or float64 data type, so no need of encoding the data.
- Our target variable "label" is a categorical data.
- There are data which cannot exist and they are handled. Some data has negative values which is not possible. So, they are converted into positive values.

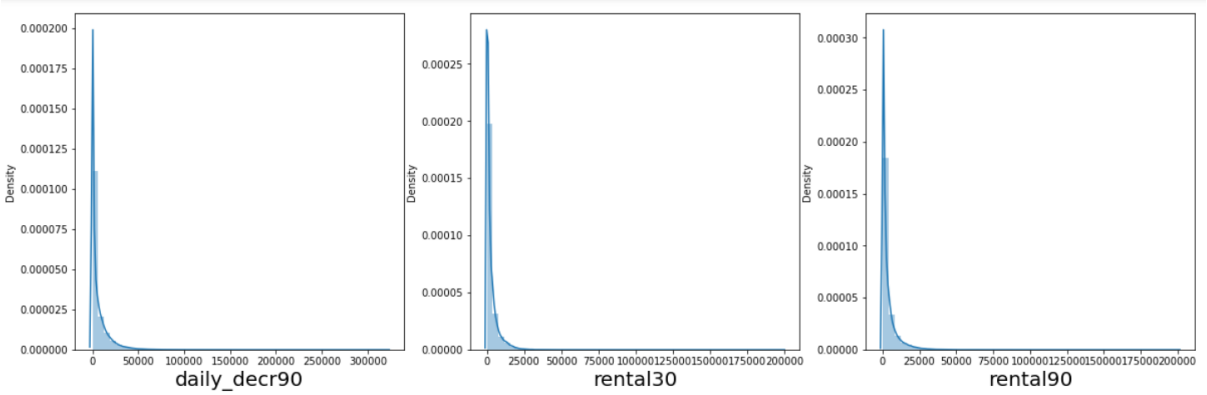
min	0.000000	-48.000000	-93.012667	-93.012667	-23737.140000	-24720.580000	-29.000000	-29.000000	0.000000
25%	1.000000	246.000000	42.440000	42.692000	280.420000	300.260000	1.000000	0.000000	770.000000
50%	1.000000	527.000000	1469.175667	1500.000000	1083.570000	1334.000000	3.000000	0.000000	1539.000000
75%	1.000000	982.000000	7244.000000	7802.790000	3356.940000	4201.790000	7.000000	0.000000	2309.000000
max	1.000000	999860.755168	265926.000000	320630.000000	198926.110000	200148.110000	998650.377733	999171.809410	55000.000000

8 rows × 33 columns

```
data = abs(data)# converting all negative to positive values since -ve value does not make sense.
data.describe()
```

	label	aon	daily_decr30	daily_decr90	rental30	rental90	last_rech_date_ma	last_rech_date_da	last_rech_amt_ma	c
count	209593.000000	209593.000000	209593.000000	209593.000000	209593.000000	209593.000000	209593.000000	209593.000000	209593.000000	2
mean	0.875177	8112.808718	5381.431293	6082.544140	2715.376558	3509.497148	3756.125844	3712.204801	2064.452797	
std	0.330519	75696.032666	9220.606472	10918.796572	4294.257439	5754.630600	53905.872857	53374.833299	2370.786034	
min	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	1.000000	246.000000	42.453000	42.693333	299.700000	330.000000	1.000000	0.000000	770.000000	
50%	1.000000	527.000000	1469.175667	1500.000000	1095.750000	1347.220000	3.000000	0.000000	1539.000000	

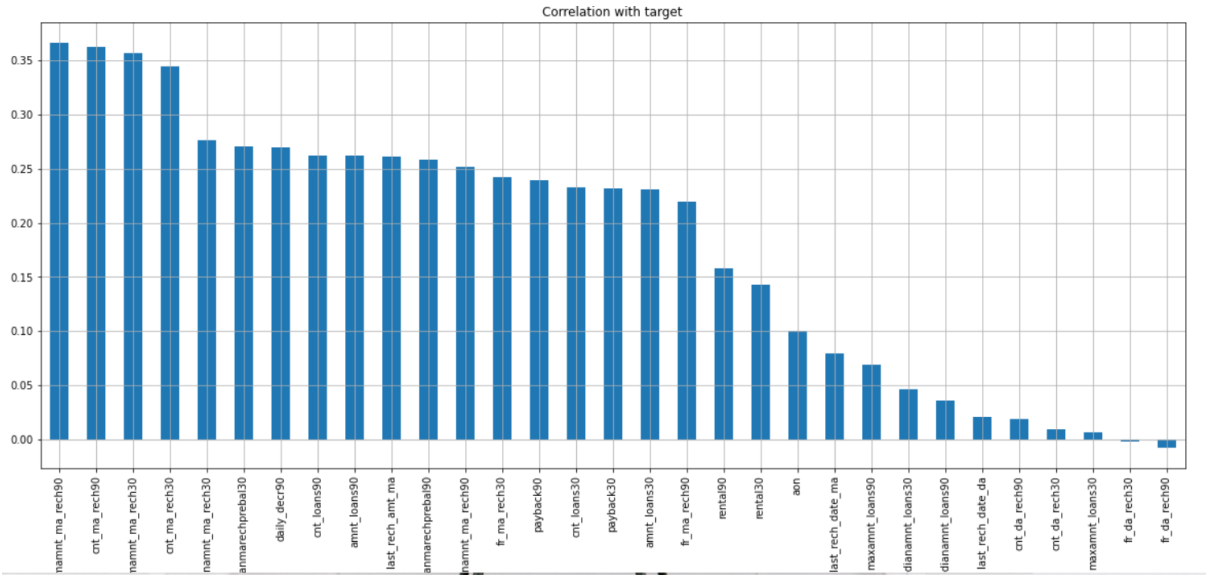
- Then visualization plots are made to see how the data is distributed.
- It turns out all the features has a large skewness and they are treated using Power Transformer. Here are a few plots:



- A heatmap is plotted to see multicollinearity. Turns out that we can remove ‘daily_decr30’ column since there is multicollinearity.

```
plt.figure(figsize=(30,20))
sns.heatmap(data.corr(),annot=True)
plt.show()
```

- A correlation between the features and label is plotted and most of the features has positive correlation with label.



Model building and conclusion:

- Split the data into train and test and import necessary libraries.

```
x = data.drop('label', axis = 1)
y = data.label
```

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import BaggingClassifier, RandomForestClassifier, GradientBoostingClassifier
from sklearn.decomposition import PCA
from sklearn import metrics
from sklearn.svm import SVC
from sklearn.metrics import roc_curve, roc_auc_score, accuracy_score, confusion_matrix, classification_report
from sklearn.model_selection import train_test_split, RandomizedSearchCV, GridSearchCV, cross_val_score
from sklearn.preprocessing import StandardScaler, MinMaxScaler
```

- Out of different algorithms, RandomForestClassifier gives the best accuracy of 91.2%.

```
y_test_pred = rf.predict(x_test)
print(f"The accuracy score is {accuracy_score(y_test,y_test_pred)*100:.2f} %")
```

```
The accuracy score is 91.21 %
```

- Cross Validation score was checked to see if the model is overfitting or not. The cross val score is 91.1% which shows that our model is not overfitting.

```
cv_score = cross_val_score(rf,x_scaled,y,cv = 6)
cv_mean = cv_score.mean()
cv_mean
```

```
0.9117718641022615
```