

Micro Credit Loan Defaulter Project Report



Submitted by:

Benjamin Emmanuel

**ACKNOWLEDGMENT**

I'd like to express my heartfelt gratitude to my SME (Subject Matter Expert) Mohd. Kashif, as well as Flip Robo Technologies, for allowing me to work on this project on Micro Credit Loan Defaulter project, as well as for assisting me in conducting extensive research that allowed me to learn a lot of new things.

In addition, I used a few outside resources to help me finish the project. I made sure to learn from the samples and adjust things to fit my project's needs. The following are all of the external resources that were utilised to create this project:

1) https://www.google.com/

2) https://www.youtube.com/

3) https://scikit-learn.org/stable/user\_guide.html

4) https://github.com/

5) https://www.kaggle.com/

6) https://medium.com/

7) https://towardsdatascience.com/

8) https://www.analyticsvidhya.com/

**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.

* Conceptual Background of the Domain Problem

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).

* Review of Literature

1. What is Microfinance?

Microfinance is frequently thought of as financial services for the poor and low-income. In practise, the term "microfinance institution" is sometimes used more strictly to refer to loans and other services from providers who identify themselves as such (MFIs). Microfinance can also be defined as a collection of different operators working together to meet the needs of the financially underserved in terms of poverty alleviation, social promotion, emancipation, and inclusion. Microfinance organisations use cutting-edge methods to reach and serve their target customers. Microfinance operations are fundamentally different from traditional financial disciplines such as general and entrepreneurial finance. This disparity can be explained by the fact that microcredit loans are often too little to fund growth-oriented business projects. Some unique features of microfinance as follows:

i. Delivery of very small loans to unsalaried workers.

ii. Little or no collateral requirements.

iii. Group lending and liability.

iv. Pre-loan savings requirement.

v. Gradually increasing loan sizes.

If current loans are repaid in whole and on time, there is an implicit assurance of easy access to future loans. Microfinance is seen as a catalyst for poverty alleviation, especially in developing nations, when it is delivered in innovative and sustainable methods to help the underserved poor.

2. Default in Microfinance

In microfinance, default refers to a client's failure to repay a loan. The default could be in terms of the payment amount or the payment schedule.

* Motivation for the Problem Undertaken

The major goal of this research is to create a model that can predict whether or not consumers will pay their loans on time. Machine Learning methods will be used to predict.

We obtained the sample data from our client database. In order to improve the credit selection process, the client requests certain forecasts that will assist them in making future investments and improving customer selection.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

On the basis of accounts that have been recharged in the recent 30 days, I have performed several analytics before moving further with exploratory study. I created a criteria that if a person does not recharge their main account within three months, I just delete their information because they are not important and may be old clients, but there is no revenue rotation. Then I investigated the date columns and discovered that the data is from 2016. I extracted the month and day from the date, separated the data into columns, then attempted to visualise the data using months and days.

I looked at the maximum amount of loan taken by the folks and discovered that there were more outliers in the data. According to the client's description, the loan amount that the customer can pay is either rupiah 6 or 12, so I've removed all the loan amounts that suggest the loan was accepted for more than 12 rupiah.

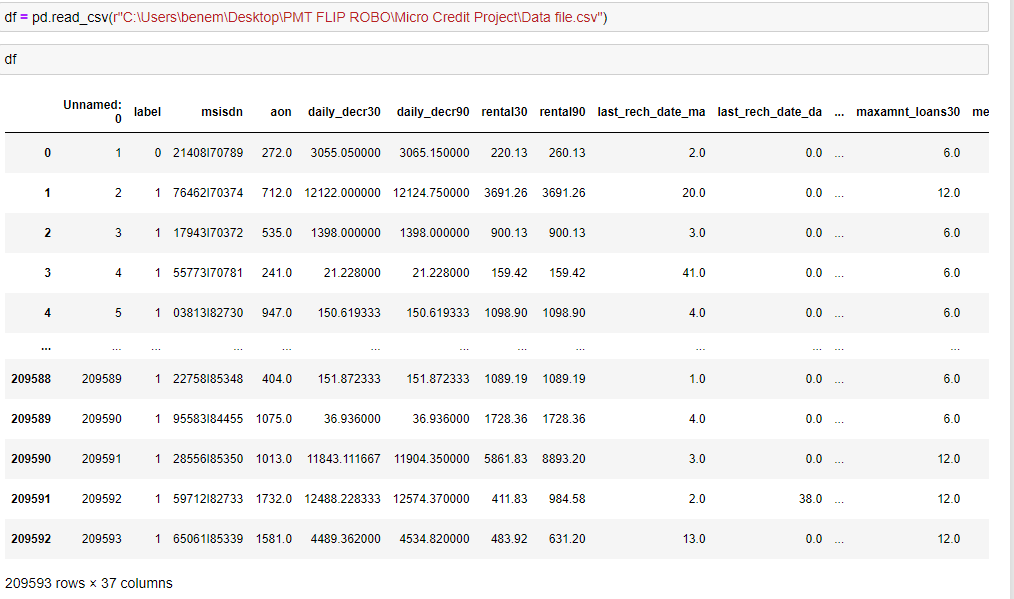
Then I segregated the defaulters' data and verified the network's valuable consumers, discovering that their monthly revenue exceeds 10,000 rupiah. We deleted some columns despite the fact that the data is highly unbalanced and many columns do not have the expected maximum value. We checked for skewed data and attempted to treat it before running the model, which resulted in NaN.

When we tried to remove the undesired data, i.e. the outliers, we discovered that about 40000 records were chopped. Despite the fact that the data provided by the client included about 37 columns and over 2 lakh columns, I did not want to risk losing any valuable data, therefore I skipped the outlier reduction step as well. After scaling my data, I ran it through a number of classification models and discovered that the Extra Trees Classifier Algorithm performed admirably.

* Data Sources and their formats

One of our telecom industry clients gave the information. They are a fixed wireless telecommunications network provider that has launched a variety of products and built a business and organisation around the budget operator model, providing better products at lower prices to all value-conscious customers through a disruptive innovation strategy that focuses on the subscriber.

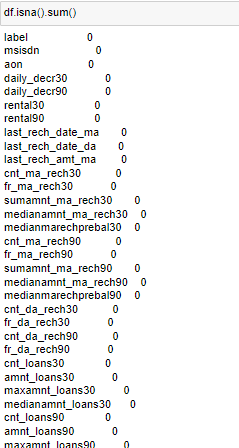
The data was provided by an Indonesian telecom firm in the form of a CSV file with an excel sheet describing the data. They also gave a problem statement, outlining what they expect of us as well as the criteria that must be met.

Let's take a look at the facts right now. I've attached a screenshot to give you an idea of what I'm talking about.

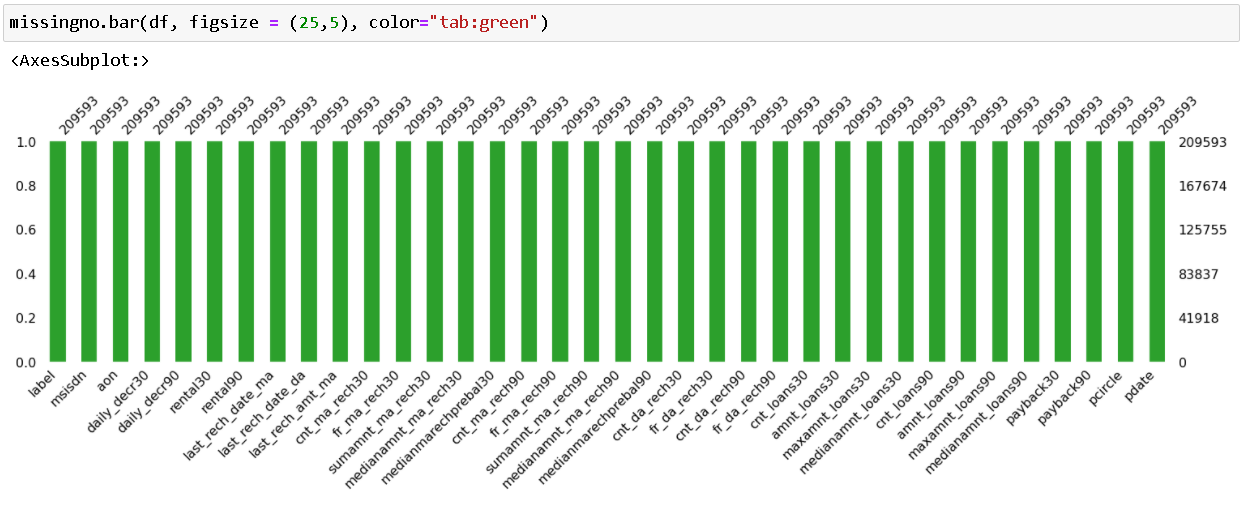
We're looking at the first five and last five rows of our dataset. It reveals that our dataframe contains a total of 209593 rows and 37 columns. This is a Classification challenge because we have the label column that stores the defaulter and non-defaulter values labelled with 0 and 1.

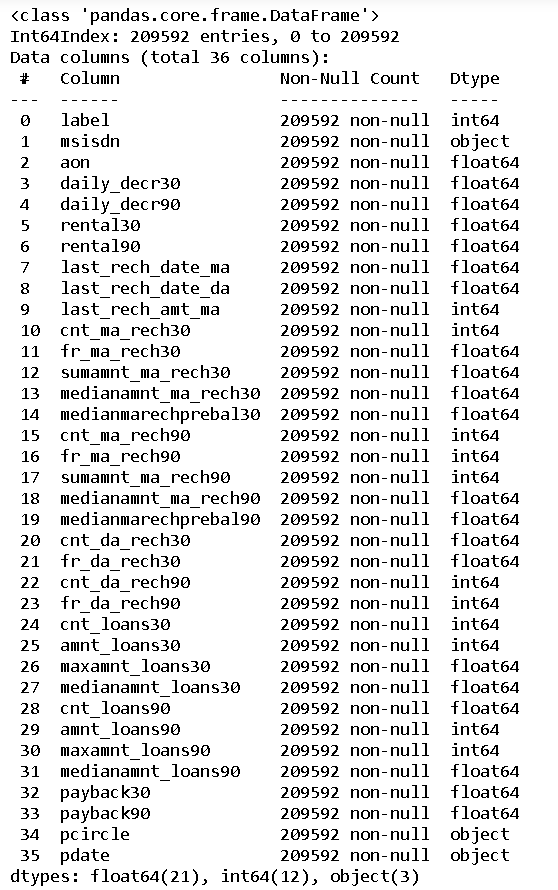
* Data Preprocessing Done

Checked for missing values to confirm the information in the problem statement that there were no null values.



Took a visual on the missing data information as well.



Using the info method, we are able to confirm the non-null count details as well as the datatype information. We have 21 columns with float/decimal datatypes, 12 columns with integer datatypes, and three columns with object/categorical datatypes. Before we can use the information in our machine learning models, we'll need to convert the object datatype columns to numerical data.

* Data Inputs- Logic- Output Relationships

Data description on each column present in our dataset.

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 users

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

rental90 : Average main account balance over last 90 days

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

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

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

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

* State the set of assumptions (if any) related to the problem under consideration

I had assumed that any telecom firm would store customer data for three months, therefore I had cut off my data based on that assumption.

Because the data is from 2016, just the date and months are different, I removed the 2016 year from the pdate columns. Months and days were divided into separate columns.

Then I looked at the data of defaulters separately and discovered that many valuable consumers are defaulters because they may have forgotten to pay or are too preoccupied with their lives. I separated them so that the company could deal with them courteously, as we cannot afford to lose these consumers.

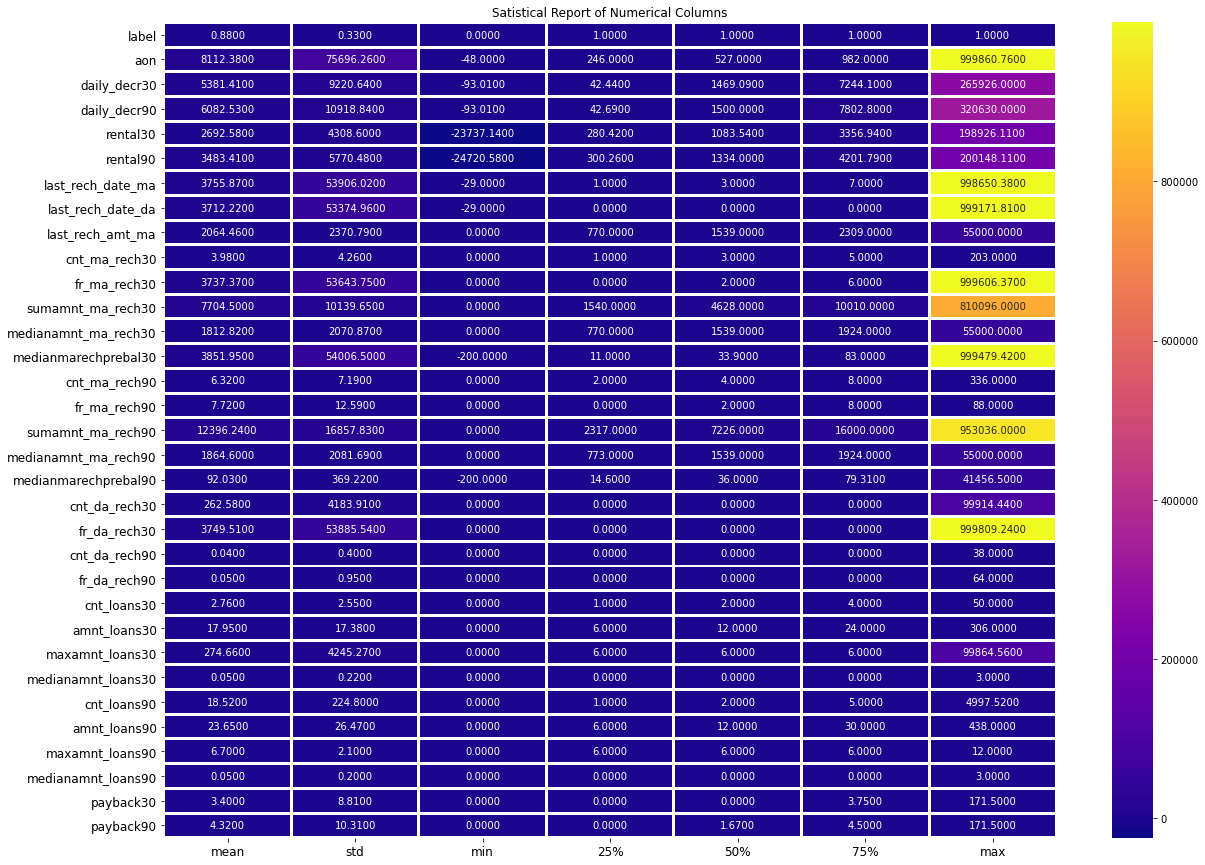
* Hardware and Software Requirements and Tools Used

Hardware technology being used.

* Hardware Used:
  + RAM: 16 GB
  + Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz 2.59
* Software Used:
  + Programming language: Python
  + Distribution: Anaconda Navigator
  + Browser based language shell: Jupyter Notebook
* Libraries/Packages Used:
* Pandas, NumPy, matplotlib, seaborn, scikit-learn

**Model/s Development and Evaluation**

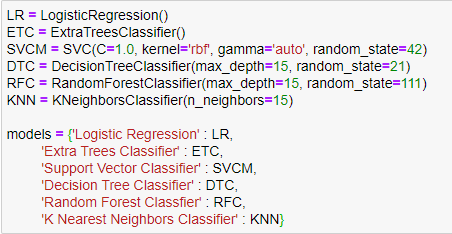
* Identification of possible problem-solving approaches (methods)

I had assumed that any telecom provider would keep the data of its customers. To check the numerical data specifics, we used the describe technique. There are 33 numerical values in the columns, and it appears that the count, mean, standard deviation, minimum value, 25% quartile, 50% quartile, 75% quartile, and maximum value are all mostly properly distributed in terms of data points, but I do see some abnormality that we will confirm with a visual.

The maximum value for columns aon, daily decr30, daily decr90, rental30, rental90, last rech date ma, last rech date da, fr ma rech30, sumamnt ma rech30, medianmarechprebal30, and medianmarechprebal30 is shown in the above report.sumamnt\_ma\_rech90 and fr\_da\_rech30 have quite a high number than the other column values.

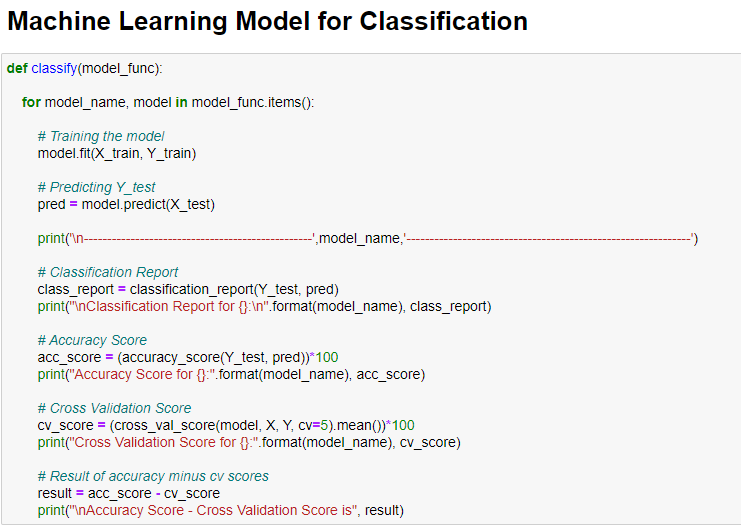
* Testing of Identified Approaches (Algorithms)

Listing down all the 6 classification machine learning algorithms used for the training and testing.



* Run and Evaluate selected models

I made a Classification Model function that includes the evaluation metrics so we can receive the data we need for all of the models.



* Key Metrics for success in solving problem under consideration

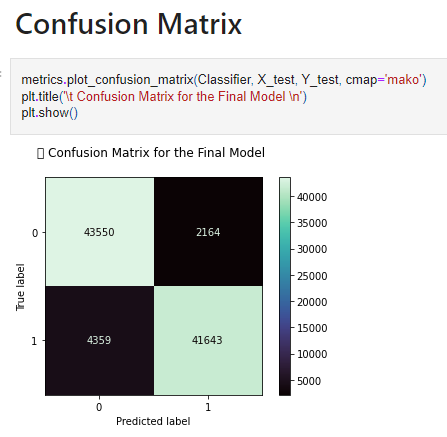
The accuracy score, cross val score, classification report, auc score, and confusion matrix were the main metrics employed in this study. We used Hyperparameter Tuning to find the optimal parameters and to improve our scores, and we'll be using the GridSearchCV method to do it.

1. Cross Validation:

Cross-validation aids in determining the model's overfitting and underfitting. The model is constructed to run on several subsets of the dataset in cross validation, resulting in multiple measurements of the model. If we fold the data five times, it will be separated into five pieces, each representing 20% of the total dataset. During the Cross-validation, the first part (20%) of the 5 parts will be left out as a holdout set for validation, while the rest of the data will be used for training. We'll acquire the initial estimate of the dataset's model quality this way. Further iterations are made in the same way for the second 20% of the dataset, which is kept as a holdout set and remaining 4 parts are used for training data during process. This way we will get the second estimate of the model quality of the dataset. These steps are repeated during the cross-validation process to get the remaining estimate of the model quality.

2. Confusion Matrix: A confusion matrix, sometimes known as an error matrix, is a table structure that permits visualisation of an algorithm's performance, usually a supervised learning algorithm (in unsupervised learning it is usually called a matching matrix). The instances in a predicted class are represented by the rows of the matrix, whereas the instances in an actual class are represented by the columns (or vice versa). The name comes from the fact that it's simple to tell if the system is mixing up two types (i.e., commonly mislabelling one as another).

It's a unique type of contingency table, with two dimensions ("actual" and "predicted") and identical sets of "classes" in each (each combination of dimension and class is a variable in the contingency table).



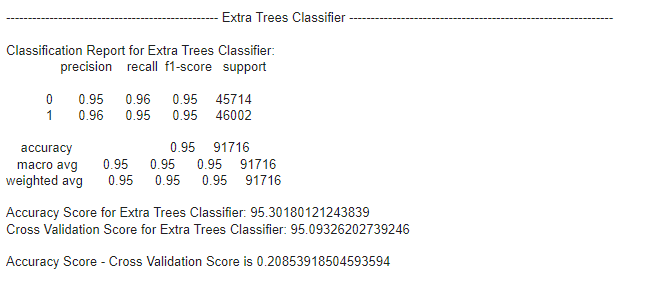
3. Classification Report: The precision, recall, F1, and support scores for the model are displayed in the classification report visualizer. There are four techniques to determine if the forecasts are correct or incorrect: 1. True Negative (TN): the case was negative and was expected to be negative. 2. TP (True Positive): the case was positive and the outcome was expected to be positive. 3. False Negative (FN): the case was positive, but the outcome was projected to be negative. 4. FP / False Positive: the case was negative, but a positive outcome was projected.

Precision refers to a classifier's ability to avoid labelling something positive that is truly negative. It is calculated as the ratio of true positives to the sum of true positives and false positives for each class. It is the precision with which optimistic predictions are made. The precision formula is given below: Precision = TP/ (TP + FP)

Recall: The capacity of a classifier to discover all positive cases is known as recall. It is calculated as the ratio of true positives to the sum of true positives and false negatives for each class. It's also the percentage of positives that were detected correctly. The recall formula is as follows: TP/(TP+FN) = Recall

F1 score: The F1 score is a weighted harmonic mean of precision and recall, with 1.0 being the highest and 0.0 being the lowest. F1 scores are lower than accuracy measurements because they factor in precision and recall. To compare classifier models, utilise the weighted average of F1 rather than global accuracy as a rule of thumb. The formula is: F1 Score = 2\*(Recall \* Precision) / (Recall + Precision)

Support: The number of actual occurrences of the class in the provided dataset is known as support. Imbalanced support in the training data could reveal fundamental problems in the classifier's reported scores, necessitating stratified sampling or rebalancing. Support does not alter depending on the model; instead, it diagnoses the evaluation process.

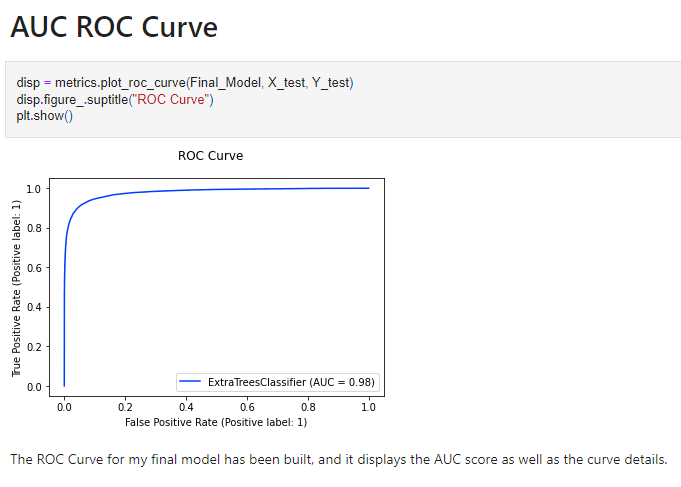


4. AUC-ROC Curve and score:

The AUC - ROC (Receiver Operating Characteristics) curve is a performance evaluation for classification issues at different threshold values. AUC represents the degree or measure of separability, while ROC represents a probability curve. It indicates how well the model can distinguish between classes. The AUC indicates how well the model predicts 0s as 0s and 1s as 1s. By analogy, the higher the AUC, the better the model distinguishes between people who have the condition and those who do not.

The ROC curve is plotted with TPR on the y-axis and FPR on the x-axis, with TPR on the y-axis and FPR on the x-axis.

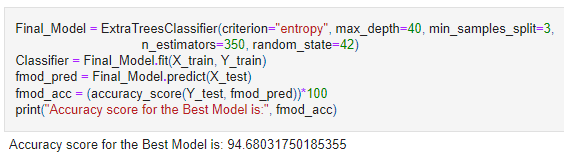
The area under the Receiver Operating Characteristic Curve (ROC AUC) calculated from prediction scores is called score.



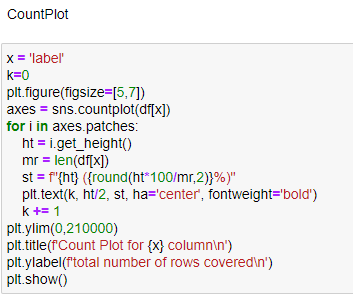
5. Hyperparameter Tuning: There is a list of several machine learning models available. They're all distinct in some way, yet the only thing that distinguishes them is the model's input parameters. Hyperparameters are the name given to these input parameters. These hyperparameters will establish the model's architecture, and the greatest thing is that you get to choose the ones you want for your model. Because the list of hyperparameters for each model differs, you must choose from a distinct list for each model.

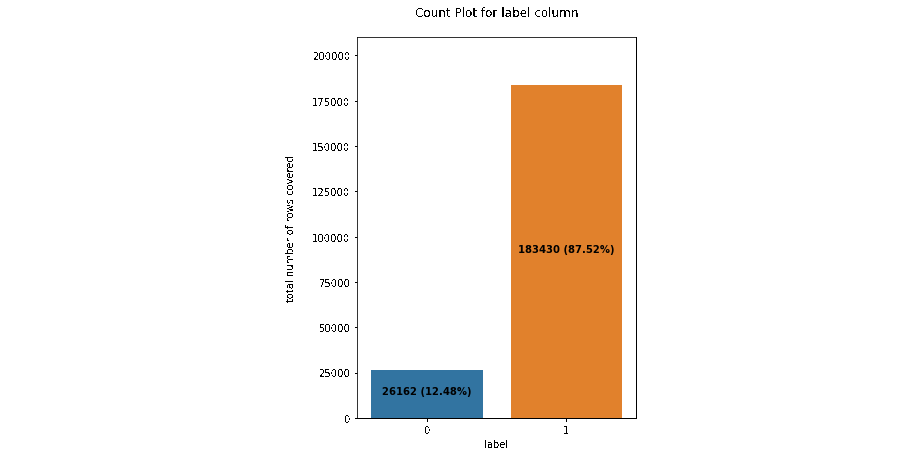
We are unaware of the optimal hyperparameter settings that would produce the best model output. So we tell the model to automatically explore and select the best model architecture. Hyperparameter tuning is the term for the procedure of selecting hyperparameters. GridSearchCV can be used to tune the system, it is a function that comes in Scikit-learn (or SK-learn) model selection package. An important point here to note is that we need to have Scikit-learn library installed on the computer. This function helps to loop through predefined hyperparameters and fit your estimator (model) on your training set. So, in the end, we can select the best parameters from the listed hyperparameters.

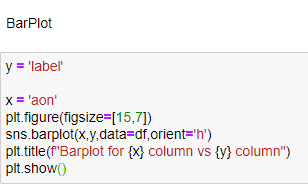


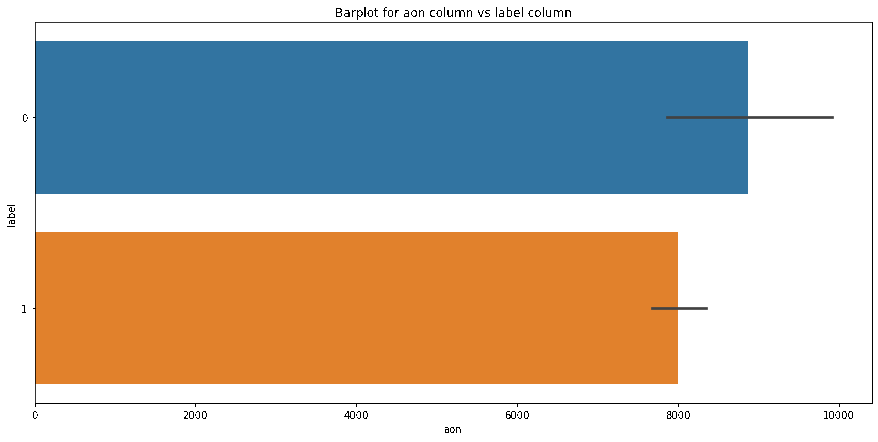


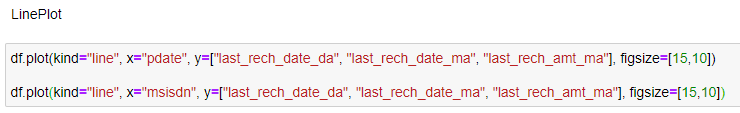
* Visualizations

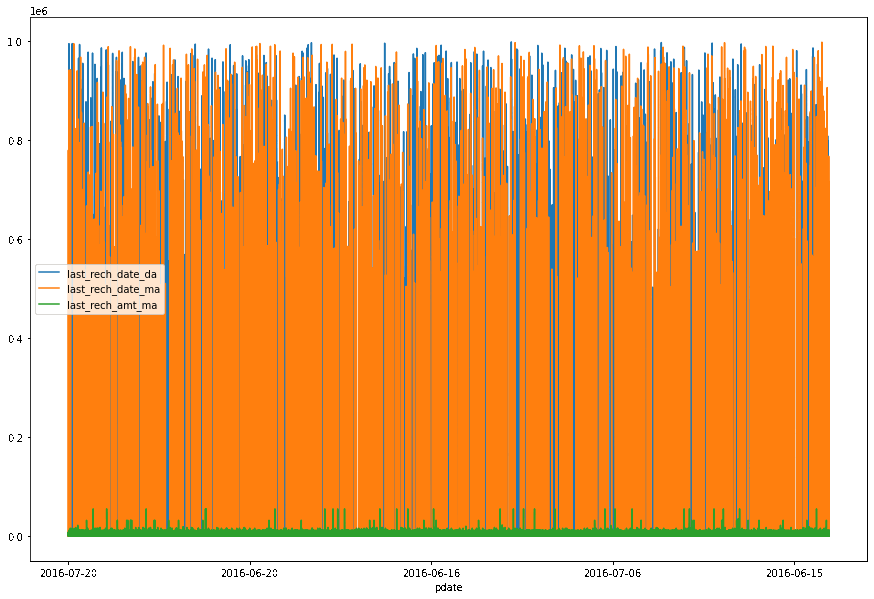
Now we'll look at the various plots created with this dataset in order to gain a better understanding of the data. The plots' codes, as well as the results, are listed below: 



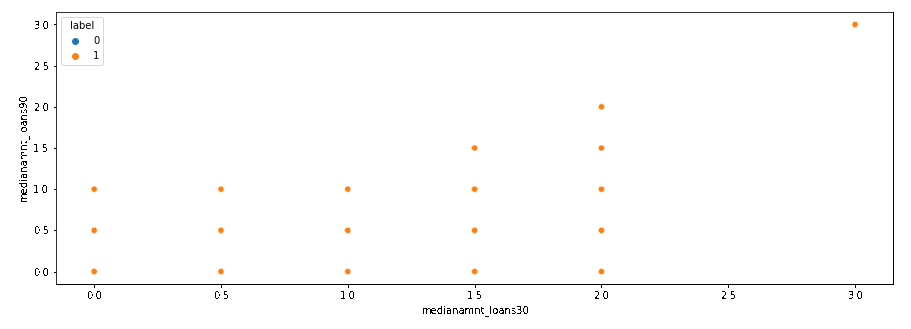












* Interpretation of the Results

**for aon feature:**

* With a mean value of 8112.34, the data spans from -48 to 999860.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature daily\_descr30:**

* With a mean value of 5381.4, the data runs from -93 to 265926.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature daily\_descr90:**

* With a mean value of 6082.52, the data spans from -93 to 320630.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature rental30:**

* With a mean value of 2692.58, the data spans from -23737.14 to 198926.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such

**for feature rental90:**

* With a mean value of 3483.41, the data spans from -24720 to 200148.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such..

**for feature last\_rech\_date\_ma:**

* With a mean value of 3755.85, the data spans from -29 to 998650.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature last\_rech\_date\_da:**

* With a mean value of 3712.2, the data spans from -29 to 999178.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature last\_rech\_amt\_ma:**

* The data is in the range of 0 to 55000, with a mean value of 2064.45.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature cnt\_ma\_rech30:**

* With a mean value of 3.98, the data runs from 0 to 203.
* The data is not dispersed naturally or in a well-defined pattern.
* Data is dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature fr\_ma\_rech30:**

* With a mean value of 3737.36, the data spans from 0 to 999606.
* The data is not dispersed naturally or in a well-defined pattern.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature sumamnt\_ma\_rech30:**

* The data is in the range of 0 to 810096, with a mean value of 7704.5.
* The data is not dispersed naturally or in a well-defined pattern.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature medianamnt\_ma\_rech30:**

* With a mean value of 1812.82, the data runs from 0 to 55000.
* The data is not dispersed naturally or in a well-defined pattern.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature medianmarechprebal30:**

* With a mean value of 3851.93, the data spans from -200 to 999479.
* The data is not dispersed naturally or in a well-defined pattern.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature cnt\_ma\_rech90:**

* With a mean value of 6.32, the data varies from 0 to 336.
* The data is not dispersed naturally or in a well-defined pattern.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature fr\_ma\_rech90:**

* With a mean value of 7.72, the data runs from 0 to 88.
* The data is not dispersed naturally or in a well-defined pattern.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature sumamnt\_ma\_rech90:**

* With a mean value of 12396.22, the data spans from 0 to 953036.
* The data is not dispersed naturally or in a well-defined pattern.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature medianamnt\_ma\_rech90:**

* The data is not distributed normally or in a well curve and ranges from 0 to 55000 with a mean value of 1864.6.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature medianmarechprebal90:**

* With a mean value of 92.03, the data runs from -200 to 41456.
* The data is not dispersed naturally or in a well-defined pattern.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature cnt\_da\_rech30:**

* With a mean value of 262.58, the data spans from 0 to 99914.
* The data is not dispersed naturally or in a well-defined pattern.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature fr\_da\_rech30:**

* With a mean value of 3749.49, the data spans from 0 to 999809.
* The data is not dispersed naturally or in a well-defined pattern.
* Data is widely dispersed and must be treated as such.
* The data is favourably skewed and must be regarded as such.

**for feature cnt\_da\_rech90:**

* The range of data is 0 to 38, with a mean value of 0.04.
* The data is regularly distributed, but not in a well-shaped curve.
* The data is favourably skewed and must be regarded as such.

**for feature fr\_da\_rech90:**

* With a mean value of 0.05, the data ranges from 0 to 64.
* The data is not dispersed naturally or in a well-defined pattern.
* The data is favourably skewed and must be regarded as such.

**for feature cnt\_loans30:**

* The range of data is 0 to 50, with a mean value of 2.76.
* The data is not dispersed naturally or in a well-defined pattern.
* The data is favourably skewed and must be regarded as such.

**for feature amnt\_loans30:**

* With a mean value of 17.95, the data varies from 0 to 306.
* The data is not dispersed naturally or in a well-defined pattern.
* The data is favourably skewed and must be regarded as such.

**for feature maxamnt\_loans30:**

* With a mean value of 274.66, the data spans from 0 to 99864.
* The data is not dispersed naturally or in a well-defined pattern.
* The data is favourably skewed and must be regarded as such.

**for feature medianamnt\_loans30:**

* With a mean value of 0.05, the data spans from 0 to 3.
* Data is not distributed regularly or in a well-defined curve, which is to be expected given that the characteristic has a limited set of values.
* The data is favourably skewed and must be regarded as such.

**for feature cnt\_loans90:**

* With a mean value of 18.52, the data runs from 0 to 4997.52.
* The data is not dispersed naturally or in a well-defined pattern.
* The data is favourably skewed and must be regarded as such.

**for feature amnt\_loans90:**

* With a mean value of 23.65, the data runs from 0 to 438.
* The data is not dispersed naturally or in a well-defined pattern.
* The data is favourably skewed and must be regarded as such.

**for feature maxamnt\_loans90:**

* The range of data is 0 to 12, with a mean value of 6.7.
* Data is not distributed properly or in a well-defined curve, which is logical given that the user has two borrowing options: 5 and 10 for which 6 and 12 must be paid.
* The data is favourably skewed and must be regarded as such.

**for feature medianamnt\_loans90:**

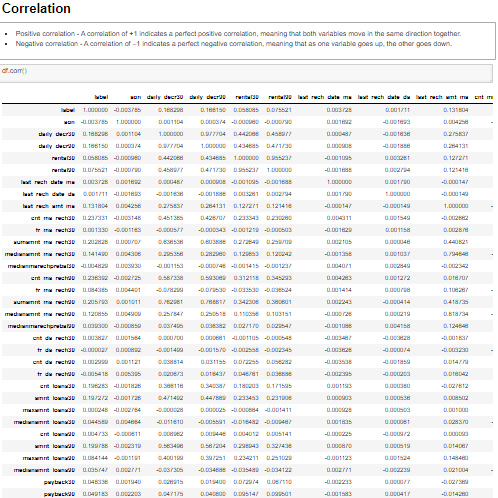
* With a mean value of 0.05, the data spans from 0 to 3.
* The data is not dispersed naturally or in a well-defined pattern.
* The data is favourably skewed and must be regarded as such.

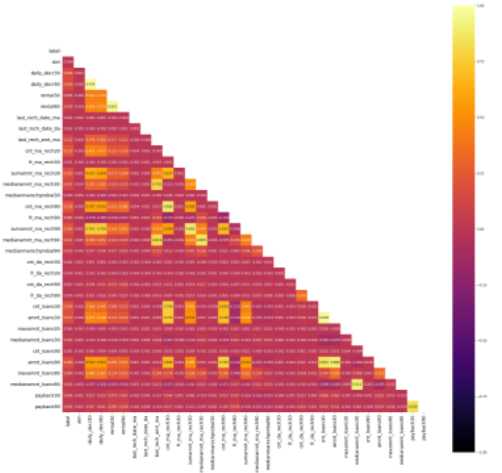
**for feature payback30:**

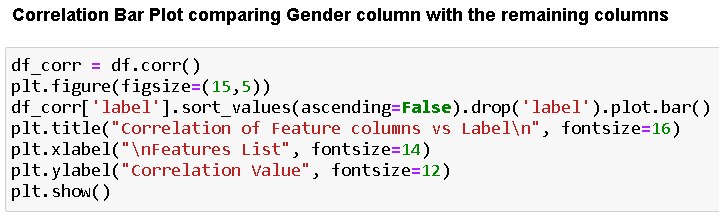
* With a mean value of 3.4, the data ranges from 0 to 171.5.
* The data is not dispersed naturally or in a well-defined pattern.
* The data is favourably skewed and must be regarded as such.

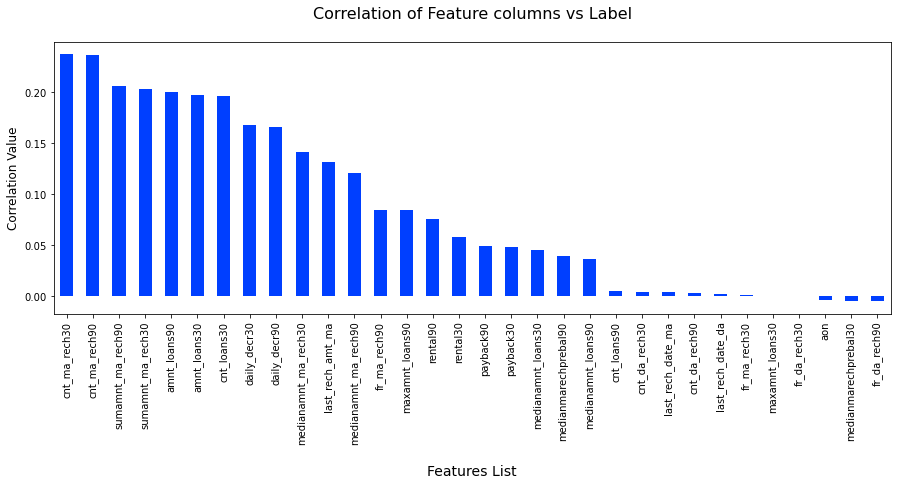
**for feature payback90:**

* With a mean value of 4.32, the data runs from 0 to 171.5.
* The data is not dispersed naturally or in a well-defined pattern.
* The data is favourably skewed and must be regarded as such.









**CONCLUSION**

* Key Findings and Conclusions of the Study

Based on the numerous variables taken into account, an MFI can determine whether or not a person will repay money and whether or not an MFI should issue a load to that individual.

* Learning Outcomes of the Study in respect of Data Science

For greater accuracy, I developed many classification models rather than relying on a single model, and I used cross validation comparison to guarantee that the model did not suffer from overfitting or underfitting. To improve the scores, I chose the best one and did hyper parameter tuning on it.

* Limitations of this work and Scope for Future Work

The limitation is that it will only function for this specific use case, and it will need to be tweaked if used in a new scenario on a similar scale. The scope of this dataset is that we can use it in companies to determine whether we should provide a loan to a person or not, and we can also make predictions about a person purchasing an expensive service based on their personal details that we have in this dataset, such as the number of times their data account has been recharged in the last 30 days and the daily amount spent from their main account, averaged over the last 30 days (in Indonesian Rupiah), so even a marketing firm can use it.

