

Lending Hut

Loan application Assistent

Anuraj D. | Sarthak B.| Sameena T.

## Project description

In 2017, Total customer debt was close to $3.7 trillion which has grown approximately 6-7% year on year. This growth is due to increasing disposable incomes, easier loan applications and overall trend in consumerism and increased standards of living. But these outstanding debt carries a direct counterpart risk for the financial institutions which have sanctioned the loans.

Lending hut our a AI backed assistant works on Probability of default model a popular component of credit risk, which helps bank and NBFCs screen loan applications on the bases of how likely they would default.

## Dataset description

For the training purpose we used lending club dataset on Kaggle. For this model data from 2007 to 2014 was used.

**Link:-** [**https://www.kaggle.com/wordsforthewise/lending-club**](https://www.kaggle.com/wordsforthewise/lending-club)

The dataset originally had more than 100 columns and several thousand fields.

## Feature engineering

A probability of default model has to be easy to interpret by any employee of the bank to screen the application that is why all the variables were converted to dummies. There are many continuous variables like annual income, debt-to-income ratio etc. to turn them into discreet variables we used fine classing technique which divides the data into ranges. To find the relevant ranges for the valuables weight of evidence was evaluated for each of the columns with target variable

A weight of evidence , computes to what extent an independent variable would predict the dependent variables

Example

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable Category | Low probability of default | High probability of default | Proportion  Of high default | Proportion of low default | Weight of Evidence |
| Higher Education | 4000 | 600 | 25% | 15% | 0.51 |
| No Higher Education | 12000 | 3400 | 75% | 85% | -0.13 |
|  | 16000 | 4000 | 100% | 100% |  |

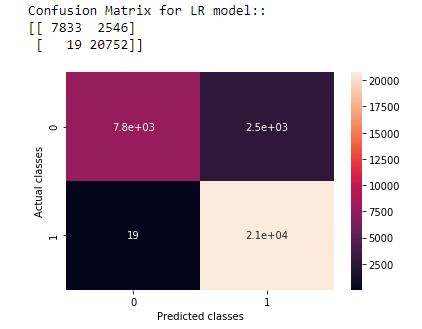
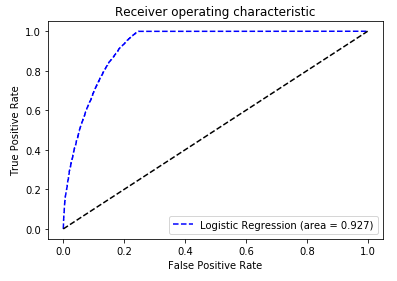
Weight of evidence formula

## Model summary

We applied many classification algorthims like Support vector machine, Logistic regression, Decision tree and after comparing range of metrics like F1 score, ROC etc. Logistic regression came out to be the best overall performing model.

After many iterations we come up with a model with following metric scores:

|  |  |
| --- | --- |
| **Metric** | **Score** |
| Accuracy | 91% |
| Precision | 89% |
| F1 score | 86% |
| AUC | 92.7% |



### Fields for the model

|  |  |
| --- | --- |
| Grade | Loan company’s assigned loan grade |
| Home ownership | The home ownership status provided by the borrower during registration. |
| Interest rate | Interest Rate on the loan |
| Verification status | Indicates if the borrowers income was verified |
| Purpose | A category provided by the borrower for the loan request. |
| Term | The number of payments on the loan. |
| Employment length | Employment length in years. |
| Months since last loan | The number of months since the last loan |
| Total Credit Lines | The total number of credit lines currently in the borrower's credit file |
| Delinquent Accounts | The past-due amount owed for the accounts on which the borrower is now delinquent. |
| Total Credit | Total installment high credit/credit limit |
| Months since earliest credit | Months since oldest revolving account opened |
| Delinquency | The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years |
| Public record | Number of derogatory public records |
| Debt-to-Income(%) | A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower’s self-reported monthly income. |
| Last Public record | Time since last public record |
| Province | The state provided by the borrower in the loan application |
| Annual Income | The self-reported annual income provided by the borrower during registration. |

## Application

The application is created with flask which is a micro web framework. The required libraries required to run the application are:

numpy==1.19.0

pandas==0.25.1

Flask==1.1.1

scikit\_learn==0.23.1

gunicorn==19.9.0

## User Interface

