Credit Risk Model

Lending Club Dataset

Imtiaz Hasan

Version 2.

2007-2018 Kaggle Lending Club Dataset (836 k rows, 3 features)
Model: Tensor Flow Neural Network Model 99% Accuracy Train & Test)

Goal: Predict whether an accepted loan will become a bad loan

Data Preprocessing

Explanatory Features: 151 columns (43 columns with >30% missing , 64 columns were dropped as it is information leakage and would bias the model

Only kept columns which are likely to be available to investors.

Only 3 columns were kept that ultimately were the best predictors of Bad loans

Target Feature: Bad Loans are considered any loans charged off, defaulted, in grace period or late 16-120 days (Bad Loans: 8 % , Good Loans: 88 %)

Key Insights

Important Predictors of Default:

1. Term
2. Loan Amount
3. DTI (Debt to Income Ratio)

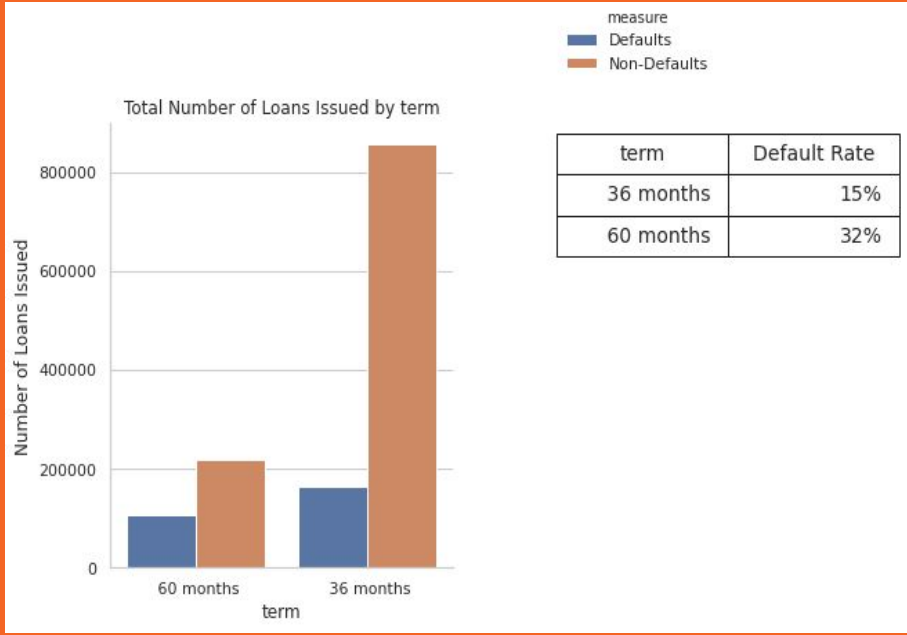
Default Rate Comparison For Important Features

Term Borrowers twice as likely to default on a 60 month loan vs 36 month

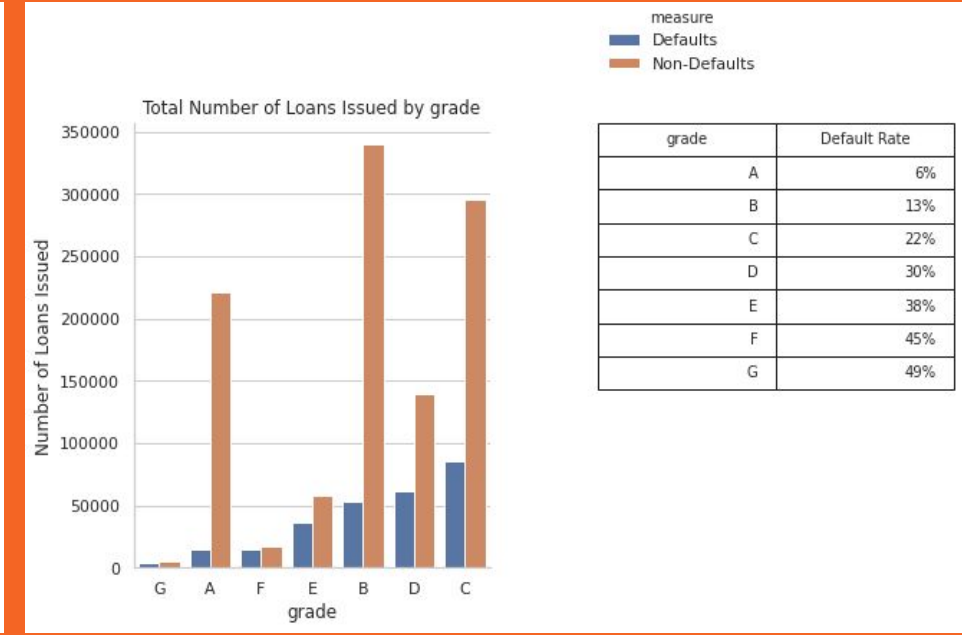
Loan Amount Borrowers would likely default on a higher loan

DTI: Borrowers are likely to default if they have a higher debt to income ratio

Key Insights



term	Default Rate
36 months	15%
60 months	32%



grade	Default Rate
A	6%
B	13%
C	22%
D	30%
E	38%
F	45%
G	49%

Model Summary & Evaluation

Artificial **neural networks** are forecasting methods that are based on simple mathematical **models** of the brain. They allow complex nonlinear relationships between the response variable and its predictors.

The parameters are tuned as follows: 5 epoch, 100 batch size, 20 hidden layers

After the dataset is passed through the neural network the 2nd time, it reaches close to 100% accuracy on both Training & Test Set. Performing this well on unknown dataset shows that the model is able to pick up the patterns from the predictors to determine good/bad loan

```
1 Train accuracy:  1.0 Test accuracy:  0.99999076 Loss:  318.5909639535239
2 Train accuracy:  1.0 Test accuracy:  1.0 Loss:  49.91473986982601
3 Train accuracy:  1.0 Test accuracy:  1.0 Loss:  26.690104481371236
4 Train accuracy:  1.0 Test accuracy:  1.0 Loss:  18.10186760412762
5 Train accuracy:  1.0 Test accuracy:  1.0 Loss:  13.652166416308319
```

Recommendation for implementation

Proposed Solution to test our model with New data

1. Collect Data (Loan Amount, DTI, Term) Loan Amount (numeric), Term(categorical), DTI(aggregated numeric)
2. Ensure size of the dataset is enough to account for all possible real world circumstances (atleast 100,000)

Summary

1. Most of the **loans issued** were in the range of 10,000 to 20,000 USD.The **year**
2. **2015** was the year were most loans were issued
3. Loans that have a **high interest rate**(above 13.23%) are more likely to become a **bad loan** .
4. Loans that have a longer **maturity date (60 months)** are more likely to be a bad loan.
5. The reason that client applied the most for a loan was to consolidate debt
6. Renters are more likely to default on a loan vs Home Owners
7. Feature Reduction and Deep Learning Models on a larger dataset as well as better handling of missing values can improve model performance

Loan Acceptance Model

Lending Club Dataset

Imtiaz Hasan

Kaggle Lending Club Dataset (9 million rows, 5 features)
Model: Tensor Flow Neural Network Model 94% Accuracy Test & 92 % Accuracy Train)

Goal: Predict whether an application will be approved

Accepted & Rejected Loans

Median Scores

accepted	Amount Requested	Debt-To-Income Ratio	Risk_Score
0	10000.0	20.55	636.0
1	13200.0	17.72	699.0

9 % of the applications are approved loans according to the dataset

Accepted & Rejected Loans

The parameters are tuned as follows: 5 epoch, 100 batch size, 20 hidden layers

```
1 Train accuracy:  0.943662 Test accuracy:  0.94817793 Loss:  9414.440069729462
2 Train accuracy:  0.92957747 Test accuracy:  0.9491348 Loss:  8859.027611339465
3 Train accuracy:  0.92957747 Test accuracy:  0.94954914 Loss:  8765.774554077536
4 Train accuracy:  0.92957747 Test accuracy:  0.9497248 Loss:  8713.440914921463
5 Train accuracy:  0.92957747 Test accuracy:  0.94989604 Loss:  8682.39115839079
```

94 % accuracy on the test data was obtained using the following variables

Amount Requested, Employment Length, State, Debt-To-Income Ratio, Risk_Score

Summary

1. 9% of the loans were approved
2. Loans approved had a higher loan amount, higher Fico score and lower debt to income ratio