Kiva Loans Exploration
Prediction and Visualization of Microfinance Funding Status

Haochen Zhan, Xinhui Yang, Zhuocheng Li

June 2019

Table of Contents

Abstract

"Dream is universal, opportunity is not." this is a motto on the interface page of kiva.org. Kiva is a non-profit organization who help provide an online microfinance platform through which an individual or a group can borrow money for multidimensional categories of purposes which include but not limited to agriculture, education, refugees and IDPs, and arts. On this platform, people can apply for a specific amount of loan which is used to support their goals or purposes. Once the loan goes through the underwriting and approval procedures, their loans will be posted on the kiva.org and wait for lenders to support.

In this report, we are going to explore and analyze factors that may vary the status of the loans based on the factors related to the loaners and lenders aspect. We will use methods such as Generalized Linear Model, Random Forest and Neural Network for regression, classification, and data visualization to display the prediction result of funded status.

Background

Our society has deep poverty amid plenty: 2.8 out of 6 billion people on the planet live less than $2 a day(World development report, 2001). Poverty exists due to social disadvantage and it is affected by various dimensions such as social norms and values. Therefore, to help people who live in poverty promote their empowerment, microfinance, originated by Grameen Bank, is created as one of the products that help alleviate the poverty problem in the third world through providing small loans to poor individuals. As microfinance form of funding gained its popularity, multiple small microfinance institutes founded in the third world countries. However, microfinance institutes often face challenges such as the insufficient amount of available loan, and risks such as bankruptcy, fraud, and portfolio risk which happened when a large number of debtors default. Fortunately, with the advent of the digital age and development of the internet, multiple lenders contributing to one loan become possible, and in this case, the challenge of lacking enough money can be solved.

Kiva is such an international non-profit online platform, founded in 2005 in San Francisco, focusing on expanding financial access to underserved communities and assisting crowdsourcing by building the connections between borrowers and lenders. With loans, individuals or groups are able to create opportunities for themselves such as growing a business. On Kiva platform, there are 3.2 million borrowers looking for loans to start a career, to accept education or for food, and 1.8 million philanthropic lenders who want to contribute in beneficent respect across 78 countries, processing 1.3 billion dollars(kiva.org). Lenders can provide a loan of $25 or more and they will get repaid periodically when borrowers repay the loan. Since lenders can choose where to make an impact and select borrowers by reading their profiles, there is a probability that some borrowers may have a higher rate to be chosen according to his or her profile and the activity to be supported. Once an individual or groups collect full fund, the repayment circle begins, which means that lenders will receive their money back periodically, however, with a default rate 3.4% not receiving the refund, and if the fund is not fully collected by the expiration date, the money will be refunded to the lenders.

The real active connection between kiva's borrowers and lenders are Kiva's field partners—local MFIs who are responsible for screening the borrowers, helping borrowers setting up the profiles online because borrowers normally do not use the internet, and also collecting repayments after fully funded. MPIs sometimes pre-disburse the loan the borrowers before the expiration date, for which MPIs will want to collect the full fund as fast as possible to complete the pre-disbursing goal. Besides, collect a small amount of full fund with a fast speed also place kiva in a competitive role in the micro lending market, which is also consistent with borrowers' demands.

Objective and Purpose

The repayment rate showing on kiva is as high as 96.6%, just like the website said, it is a loan not a donation; so when you're repaid you can use the money again, thus we focus on the prediction of the status of being fully funded, and we are interested if this statistics is as accurate as what the data will say, so our goal is, based on the last ten years of history, to construct a model which predict if a loan can be fully funded before the expiration date. In this paper, we focus on answering the main questions for the borrower:

How can you tell what is the main factor about the loaner such as loaners' personal information, description and image shown on the website, could trigger the lenders' interest? Does geographic location also affect the funded situation? When you do a microfinance research, the websites would allow you to filter the results based on keywords, or main features, such as Woman-Owned Small Businesses, National Community of Supporting Families, Earth Friendly, Sustainable Agriculture, Durable Goods, etc. Furthermore, which group or category of borrowers can have a higher chance to collect full fund? What kind of activities relate to higher full fundraising? For this project, we want to answer the questions and investigate which of these features are more likely for lenders to decide, and if we filter the search results based on specific preferences on the probability for them to get fully funded, would we be able to obtain an overall accurate prediction of the loans' funded status before expiration?

Data Mining

Data description

To answer the questions above, we obtain the raw data from the kiva loan website and this dataset has 1682789 entries and 34 columns. The variables are loan id, loan name, original language, description, description translated, funded amount, loan amount, status, image id, video id, activity name, sector name, loan use, country code, country name, town name, currency policy, currency exchange coverage rate, currency, partner id, posted time, planned expiration time, disburse time, raised time, lender term, num lenders total, num journal entries, num bulk entries, tags, borrower names, borrower genders, borrower pictured, repayment interval, distribution model.

Data Snapshot: https://build.kiva.org/docs/data/snapshots

- Provided by Kiva.org for one simple download, composed of loans, lenders, lenders

Latest Loan Data: http://build.kiva.org/docs/data/loans

- Provided by Kiva.org for cross-validation, composed of the latest information

Data Preprocess

Above is the features we are dealing with from the main dataset in which we remove columns with all missing value and columns with meaningless index. Below are the data that will be used for: funded status as response variable; translated description, description, loaner image, loan amount, activity name, country name, currency policy, borrower genders, currency exchange coverage rate, tags, repayment interval, distribution model and number of lenders as predictor variables. Table 1 shows a summary of the variables we've been using. For treating the text data, we applied Natural Language Processing such as Term-Document, Doc2vec and Word2vec. For the image data, we store each of them as 300*300*3 array in the order of RGB. Except those variables can be expressed as binary form, the rest of the categorical variables were transformed to dummy variables. Overall, we identify our data as three parts: Other variables, Text variable and Image variable. The means we implemented on them will be explained in Methodology.

Preliminary Analysis

Funding Status

Funding status is our response variable, and our goal is to predict whether it would have funded or expired status. Initially, there are four classes in funding status which are funded, expired, fundraising and refunded. However, fundraising is some new cases that are currently gathering the funds from lenders in the most recent time, while the refunded status stands for some rare cases due to some posting errors and violation to Kiva's policy terms. Therefore, the funded and expired classes are the usual cases we determined. The existing concern is that we have too many "funded" cases which stands for 95% of the data, so it will give a highly unreliable classification correct rate. Imagine the machine just place all the response at "funded", and it will give nothing helpful to the upcoming lenders and borrowers to refer. In the following studies, we will focus on the more detailed information in the downsampled dataset. Downsampling method evenly sampled from the variables we are going to use, and balanced the two classes in the meanwhile. After downsampling, we still have adequate 150,767 observations left with 50 to 50 percent of status funded and expired.

Geographic Distribution

We found they are mostly concentrated on the developing countries. They will contribute to a difference because they may have a different currency exchange coverage rate. (Fig.1)

Distribution for Fully Funded Time

Among the funded loans, we found that the fastest funding time takes only 2 minutes, but the longest funding time takes 768562 minutes which is 17.79 days equivalently, and the average funding time is 13.235 minutes. We also found an interesting rule that although people evenly borrowed money during every month, the expired ratio would be higher than usual in summer. (Fig.2)

Sectors and Activities

We found agriculture, food and retail are the most listed sectors in kiva loans dataset. (Fig.3)

Description Insight

Descriptions are crucial for lenders to determine whether or not to submit funds to a specific loan. For reason of consistency of our exploration, we have merged the English descriptions with translated descriptions which are written in other languages. The following two descriptions are randomly selected from the funded group and the expired group in the original data :

Expired

Valens is married to a farmer. He has been a rice farmer for 4 years. He would like to get this loan to buy seeds and support labor costs to get the best harvest. With the profits, he plans to expand *the area he has been using to get more output.The agriculture sector accounts for 37% of Rwanda's gross domestic product, generates 65% of Rwanda's export revenue, and employs approximately 90% of Rwandans (as of 2009). Despite the importance of agriculture to Rwandans and their economy, financial institutions view lending to fund agricultural activities as a high-risk proposition because the profitability of these activities is affected by weather, natural disasters, and price fluctuations. For this reason, farmers in Rwanda remain under-served by financial institutions. Urwego Opportunity Bank is expanding into this market and is happy to provide Kiva lenders with the opportunity to support Rwandan farmers.*

Funded

Elsie and her husband earn well from their banana farm. They have been growing and harvesting bananas for more than two years to make a living. To finance farm maintenance and fertilization costs, she borrowed a new loan from GDMPC micro credit. She is hopeful to gain more return from the farm and wants to improve for a better source of income in the future.

From the above two loan descriptions, we reckon that there are influential factors within the text that eventually led the loan case funded or not. These factors can be the gender of the loaner, individual or group of loaners, the purpose of the loan and some keywords that affect the lenders. To feed the text data into our models, it requires some natural language processing. We predicted the funding status based on the text data only to check its usefulness in prediction. We first transformed the text data to the Term-Document matrix and Term frequency-Inverse document frequency (TF-IDF) form respectively. Both methods are likely to count the frequency of the words in all descriptions and assign different weights on each word, and they have preprocessed words so that the words become uniform. Then we used the regularized logistic regression model on them to observe some keywords patterns that might affect predicting funding status which is in the binary form with 0 stands for expired and 1 stands for funded. After running the function that helps select the best hyper-parameters (penalty type and value) for the model by using the GridSearch function on 80% descriptions as the training set in Python and validated with 5-fold, we fitted the test set which states for 20% of all descriptions. The one with the highest test accuracy is the Term-Document with 0.773 accuracy rate. The confusion matrix and ROC curve for the test results are listed in Fig.4.

Among all the coefficients selected by the logistic model, there are the most negative numbers that contribute to the probability of the loan case being expired (Y = 0), and the most positive coefficients that affect case to be funded(Y = 1). The following keywords are the most influential words which have the largest weight derived by the logistic regression model:

Top 20 words for expired: sep, kilolo, 120, jordan, gegharkunik, represented, jod, colgate, acb, collectively, kopsyah, conservative, texas, septic, metric

Top 20 words for funded: ngn, filter, portoviejo, latrine, milaap, bjs, kadet, ho, murabaha, asasah, widow, dtm, liberia, vahatra, translated

From the above keywords observed some city names like gegharkunik, public facilities like latrine, special tags like widow and many field partners like Vahatra and Asasah. It seems that some words described in descriptions can lead the direction of funding status. Although the prediction result is good to some extent, the possibility of observing any patterns from these two different sets of words is limited. The main reason for that is the indirect relation between description and funding status when we attempted to classify the one based on another. In further exploration, we considered using Doc2vec and Word2vec to process the text data more closely.

Methodology

The overall idea of solving our classification problem is to compare various models and add text data or image data onto other numerical and categorical data which can enhance the eventual prediction result.

1. Models for Other Variables Only

For prediction using other variables excluding all the text data and image data, we used regularized logistic regression (LR), random forest (RF), Multilayer Perceptron (MLP) which is a neural network model with more than one hidden layer. Note that the neural network without any hidden layers is exactly the logistic regression model. The reason for using multiple layers in our model is to add more trainable parameters systematically such that produces a better result.

2. Models for Other Variables + Text Data

For analyzing all variables including the text data, we implement the same MLP model introduced from above, plus Recurrent Neural Network (RNN) which is an advanced deep learning model that takes the original word embedding vectors we trained from the Word2Vec model. For instance, RNN takes word vector one by one from a description, then use forward propagation to output, backward propagation to adjust weights based on gradient descent. The key point for RNN is to consider the previous words when process the current word. By coordinating RNN with Long-short term memory (LSTM) cells which extended the memory of RNN that also solved the issue of vanishing gradients. The comparison between the feedforward neural network and RNN is as Fig.5.

2.1 Two Ways of Processing Text Data

2.1.1 Using Doc2vec and Word2vec mean

The most intuitive way of transforming text data to structured data is to make each description a vector under the same dimension. To achieve this goal, we can use the pretrained Doc2vec model from Wiki database (1.51G) to make each description a 300-dimensional vector. This method is nothing special but makes similar documents closer in the vector space. Another alternative way is to get 100-dimensional word vectors for all words appeared in descriptions by Google News pre-trained Word2vec model (3.6G). Then add all word vectors in each description and find the mean of each. Word2vec mean not only consumed one day to run but also lost the context of the sentence within the description.

2.1.2 Explain Word2vec Vectors and Embedding Layer in RNN

Since Word2Vec is what we used to make text data structured. Word2Vec is a group of related models that are used to produce word embeddings. These models are shallow, two-layer neural networks that are trained to reconstruct linguistic contexts of words. Word2vec takes as its input a large corpus of text and produces a vector space, with each unique word in the corpus being assigned a corresponding vector in the space. Word vectors are positioned in the vector space such that words that share common contexts in the corpus are located closely. Fig.7 shows the spatial relationship among some words after we projected the word vector space on 2-dimensional space by using PCA dimension reduction technique. Basically, each vocabulary in our trained Word2vec model has 100 dimensions. Words that share common meaning has a high similarity value. For example, vectors for pig, fattening, vitamins, feeds, and piglets have close distance which implies the context of the description can be kept. Word2vec mean method introduced above is time consumable and reluctant. The other safer and advanced way is to form a high-dimensional matrix which has (number of keywords within the descriptions) * (word vector dimension (in this case 100)) dimensions. Then we can transform descriptions as sequences of discrete numbers from 0 to number of all words and put them into embedding layer cooperating with the matrix we formed before as weights, and let RNN to compile them without losing any information.

2.1.3 Fail on Dimensional Reduction for Text Data in Term-Document Form

We once tried to reduce the dimension of descriptions in Term-Document representation and failed because the number of features are over 50,000 thus is impossible to do even on the server of STATS department.

3. Models for Other Variables + Text Data + Image Data

The reason of combining three different kinds of data is to boost our prediction accuracy. We believe information of all sorts would affect lenders' decision. Besides the other variables and text variable, loaners' image plays an important role for displaying on the crowdfunding website. Fig.8 has shown 50 images each randomly chosen from expired and funded. It can be discovered that there are two missing photos in funded and four missing photos in expired cases. The way they shot the photos are completely different. We will collect the deep features in the photos by applying Convolutional Neural Network (CNN) which is suitable for image processing and eventually concatenate them together in the neuron forms. In this research, we set two layers of kernel and use technique called max pooling and strides to accelerate the CNN process on our images. The whole process is shown in Fig.6

Result

1. Other Variables Only

Generalized Linear Model: Logistic with Ridge Regularization (L1 Penalty) (LR)

The accuracy rate of L1 Logistic Regression is better than L2 LR which results in 82.85%. Since the matrix consists of many dummy variables, which give very large coefficients, we shrank the regularization term to find the best fit. From the accuracy curve, we found that the best alpha choice for using euclidean distance regularization is between 1.5 and 2 in this case (by the stochastic solver). Furthermore, using the fit intercept is not helpful in this case, which will cost more accuracy than usual. Fig.9 shows the accuracy changes as we adjusted the regularization term.

Random Forest Classification (RF)

Compared with the linear model, random forest could give a relatively higher prediction rate. The prediction score of each single Decision Tree is between 79% and 81%. The prediction score of Random Forest is about 85.24%. Random forest is the dominant model for classification to some extent. The importance of coefficients given by RF is in Fig.12. As Rf selected, currency policy, number of lenders, repayment interval, distribution model, loan amount, some countries like Zimbabwe and China, some tags like widowed, job creator and schooling are indicated as crucial factors for funding status. Fig.10 has shown the ROC curves between LR and RF model. The area under curve from RF is much bigger than LR.

Neural Network Classification

We set up a two-layer neural network with each layer contains 512 neurons and activated them by relu. The training accuracy became better as we iterate more times. It eventually gives 83% accuracy rate which is worse than random forest. The keras network result visualizes the accuracy on cross validation has a significant increase with each epoch, in contrast to the loss in decrease.Fig.11

2. Other Variables + Text Data & Other Variables + Text Data + Image Data

Since the neural networks work implicitly, there is not much to tell comparing with the baseline models for other variables only. Best accuracy after tuning is reported as following:
MLP for Other variables + Word2vec mean : 82.97% (see code details) MLP for Other variables + Doc2vec: 83.74%(see code details) RNN for Other variables + Word2vec vectors: 77.25% (See Fig.13) MLP for Other Variables + Doc2vec + Image Data in CNN deep features: 83.22%(See Fig.14). From the above results, we found that the RNN is a bad strategy for us to deal with descriptions since it generated the worst accuracy rate. Using MLP for combinations of two and three types of data works pretty decent. This can be further developed and served as a tester to predict the funding status. However, it's still hard to explain what types of photos are preferable to lenders since we simply processed images by CNN.

Conclusion

Combined the results from the linear and non-linear methods above, we can conclude that

- High education costs and primary/secondary school costs are the most popular activities. They have a very significant positive influence to be successfully fully funded.
- Lenders prefer to contribute their investment to female education, but they do not like women owned business. Furthermore, lenders prefer to invest manufacturing sectors, such as Sewing, Embroidery and Carpentry.
- Lenders are mostly friendly usually prefer eco-friendly, health and sanitation projects, but they don't like interesting photos, which has a significant negative impact on the tag information.
- The country region also makes a difference, because of its exchange coverage rate.
- Lenders don't like interesting or inspiring photo, neither the powerful stories. It will only give you a negative impact on the funded amount.

Discussion

We did not use some columns such as Video ID, because these columns either contain all missing values or are not related to the variables we are dealing with. Besides, as we know that the field partners (MFIs) would disburse the loans to the borrowers ahead of time, and then they will begin to collect the loans online after post time. Thus we expect a small time difference between posted_time and raised_time for that disburse_time is earlier than posted_time. The reason for this fact can be easily interpreted: to make money while yielding data, which is invaluable. Due to Kiva's non-profit trait, its field partners will be charged zero interest rate when they obtain the loans from Kiva. Since the field partners have a direct connection with the borrowers, they can schedule an interest rate of repayment under Kiva’s supervision, which ensures that field partners provide practical outcomes for borrowers. When the field partners repay the loan to Kiva, they only need to pay the principal of the loan and make the interest from the borrowers as their profit. A shorter period fundraising time with small amount of loan means less risk since the 4% of default risk rate is due to the borrowers’ default.

References

“Attacking Poverty: Opportunity, Empowerment, and Security.” *World Development Report 2000/2001*, 2000, pp. 1-12

Thorpe, Devin. “Kiva Is Really A Crowdfunded Bank For Refugees And Other 'Unbankables'.” *Forbes*, Forbes Magazine, 24 Sept. 2018,www.forbes.com/sites/devinthorpe/2018/09/24/kiva-is-really-a-crowdfunded-bank-for-refugees-and-other-unbankables/.

Hartley, S. E. (2010). Kiva. org: Crowd-sourced microfinance and cooperation in group lending. *Available at SSRN 1572182*.

Moleskis, Melina and Canela, Miguel Angel, Crowdfunding Success: The Case of Kiva.Org (March 3, 2016). IESE Business School Working Paper No. 1137-E. Available at SSRN: https://ssrn.com/abstract=2769841 or http://dx.doi.org/10.2139/ssrn.2769841

Donges, N. (2018). Recurrent neural networks and LSTM. Towards Data Science.

https://towardsdatascience.com/recurrent-neural-networks-and-lstm-4b601dd822a5

Li, Y., & Yuan, Y. (2017). Convergence analysis of two-layer neural networks with relu activation. In Advances in Neural Information Processing Systems (pp. 597-607).

Pierre Baldi & Roman Vershynin (2019). The Capacity of Feedforward Neural Network.

https://www.math.uci.edu/~rvershyn/papers/bv-capacity-neural-networks.pdf

Shen Yuanyuan, and Thai T. Pham. “A Deep Causal Inference Approach to Measuring the Effects of Forming Group Loans in Online Non-Profit Microfinance Platform.” *ArXiv.org*, 8 June 2017, arxiv.org/abs/1706.02795

Supplementary

Table 1. Summary for used data

Feature	Description
Description (text data)	Personal profile including borrowers' age, marital status, and other information
Translated Description (text data)	These descriptions were written in other languages and later translated to English
Loaners image (300*300 RGB)	The loaners' image shown in the front page
Loan Amount	The total amount that borrowers want to collect, which is the only numerical data taken account
Funded Status	The status of the fund, including funded, expired, fundraising and refunded
Activity Name	The description of the utilization of the loan
Borrower Gender	It contains mixed group of males and females
Country Name	The countries that the borrowers are from
Currency Policy	Whether the lenders need to share the currency exchange risk or not (Share or Not Share)
Currency Exchange Coverage Rate	Exchange rate of the currencies at the time the loan case was posted
Tags	The additional information about borrower's personal info
Repayment Interval	The way the borrower repays part of the loan and interest. (Monthly, Bullet or Irregular)
Distribution Model	Process through KIVA directly or field partner

Figure 1 Geographic Distribution of Countries

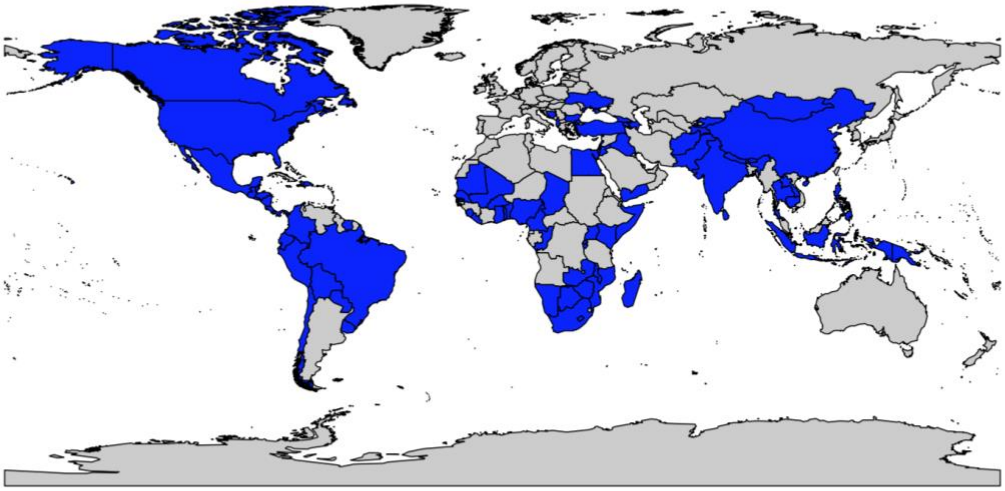


Figure 2 Distribution for Fully Funded Time

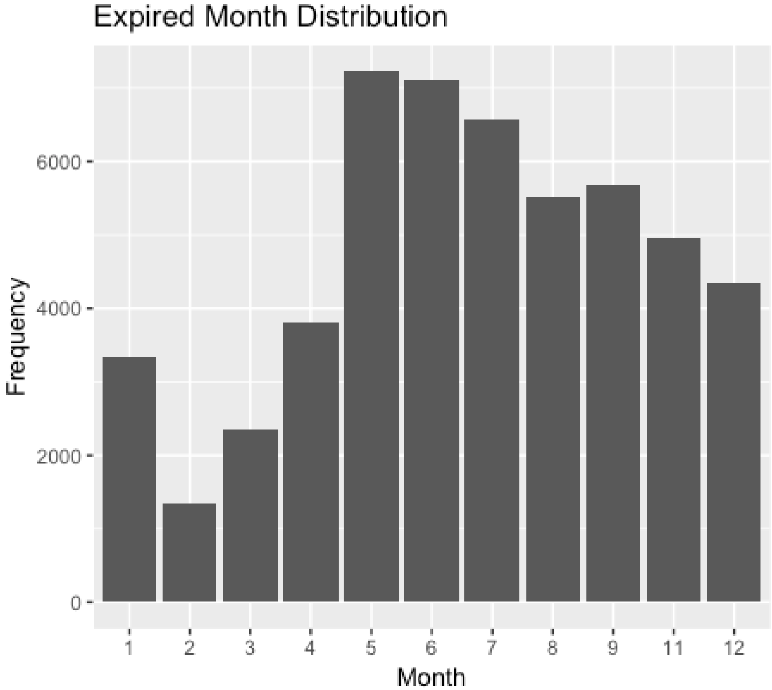


Figure 3 Exploratory Visualization of Activity and Sector

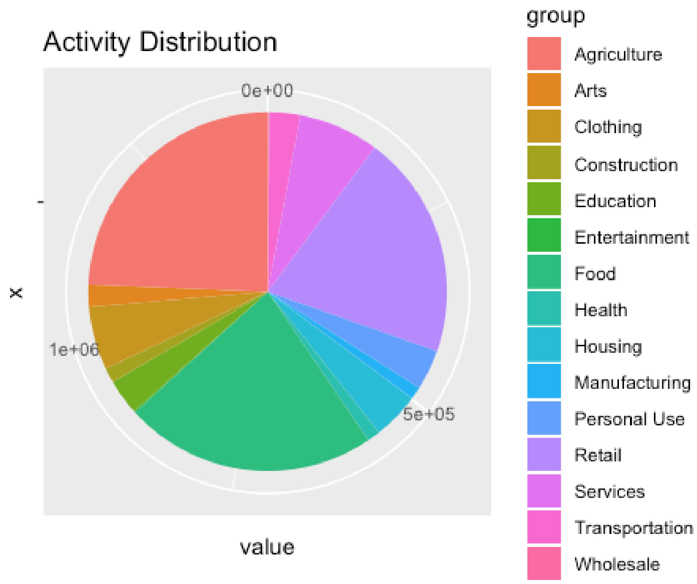


Figure 4 Confusion matrix and ROC Curve for LR on text data in the Term-Doc form

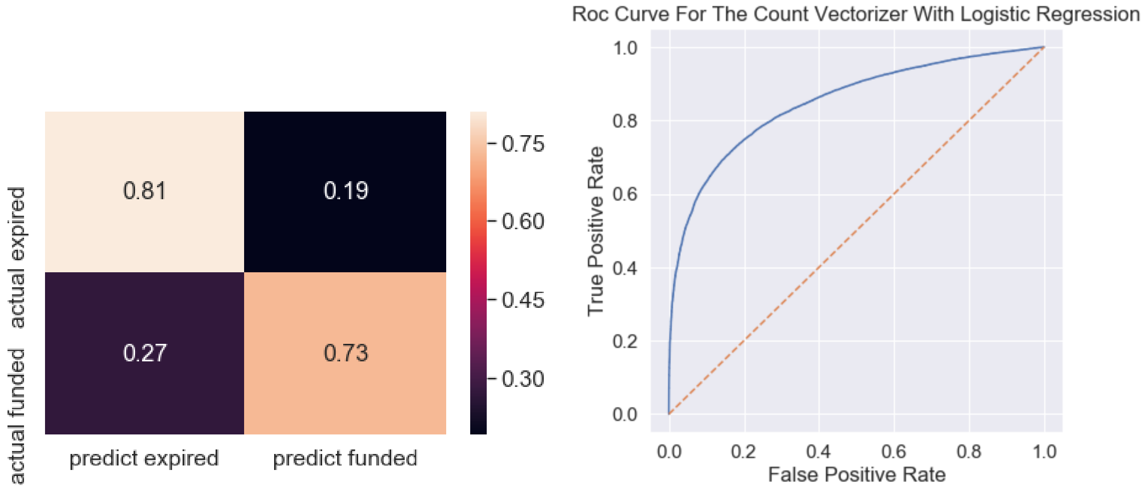


Figure 5 RNN vs Feed_Forward Neural Network

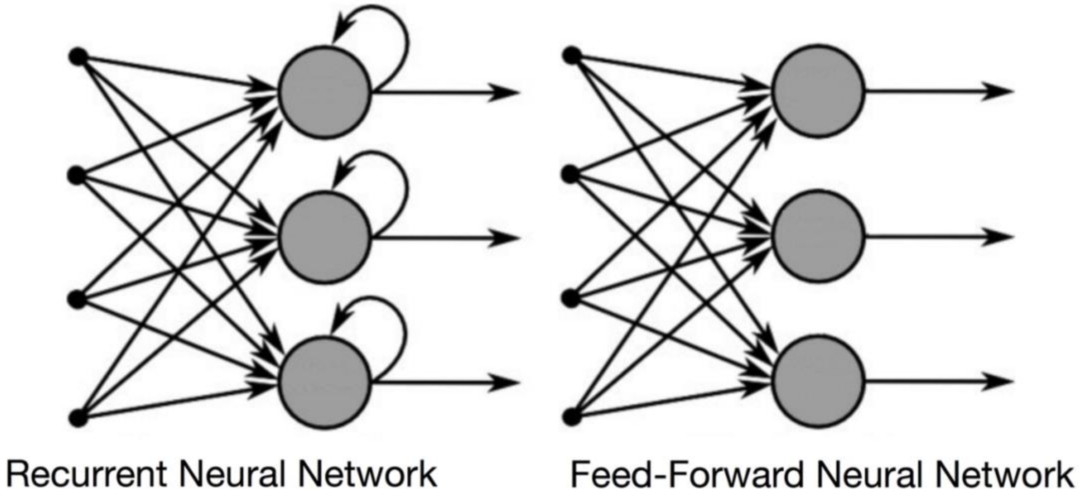


Figure 6. Other variables + Text data + Image Data

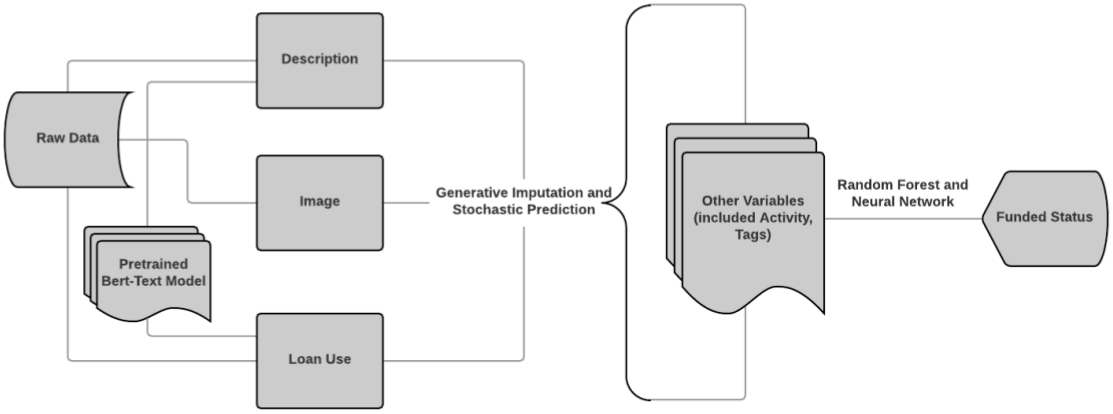

Description
Raw Data
Image
Pretrained Bert-Text Model
Loan Use
Generative Imputation and Stochastic Prediction
Other Variables (included Activity, Tags)
Random Forest and Neural Network
Funded Status

Figure 7 Sampled Word2vec vectors on 2D by PCA

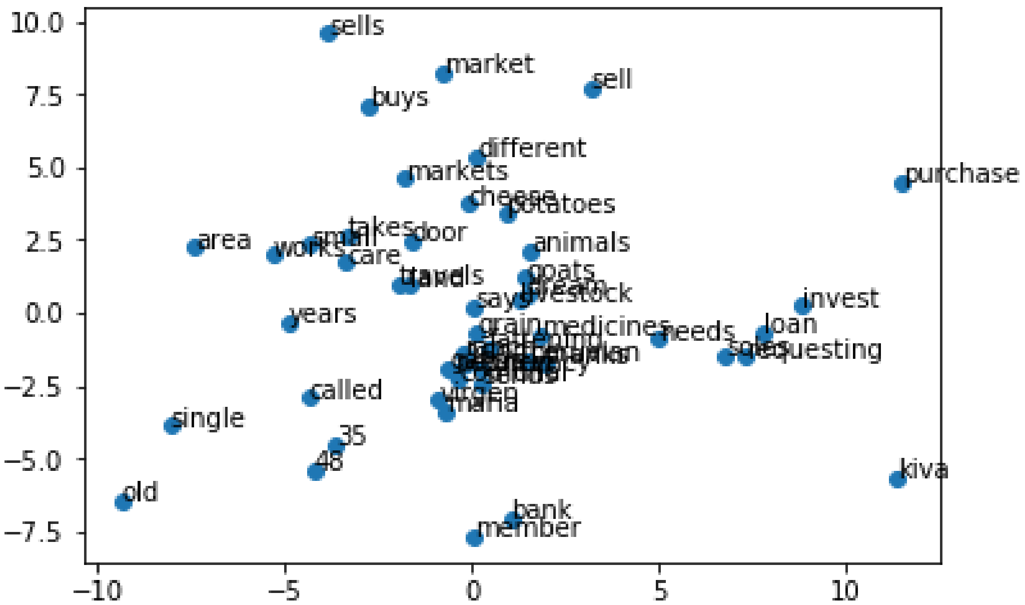


Figure 8 50 images each for expired and funded cases.

Expired:

LARSEN
softwire
101 WAYS
YOUR KID$
CAN MAKE MONEY
The Secrets About Money that You Don't Learn In School!

Funded:

ONE
ACRE
FUND

Figure 9 Accuracy Curve with L1 Logistic Regression

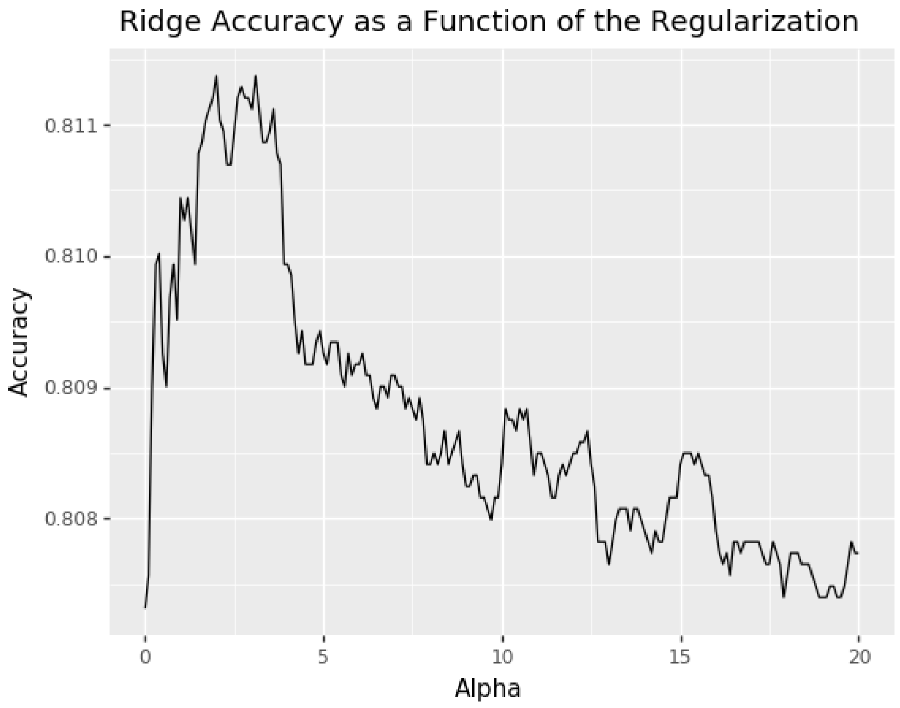


Figure 10 ROC curve between LR and RF

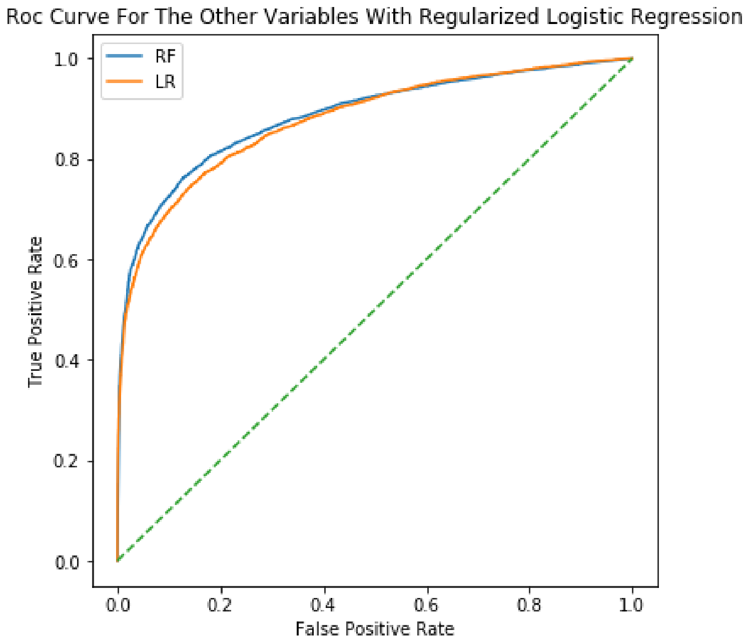
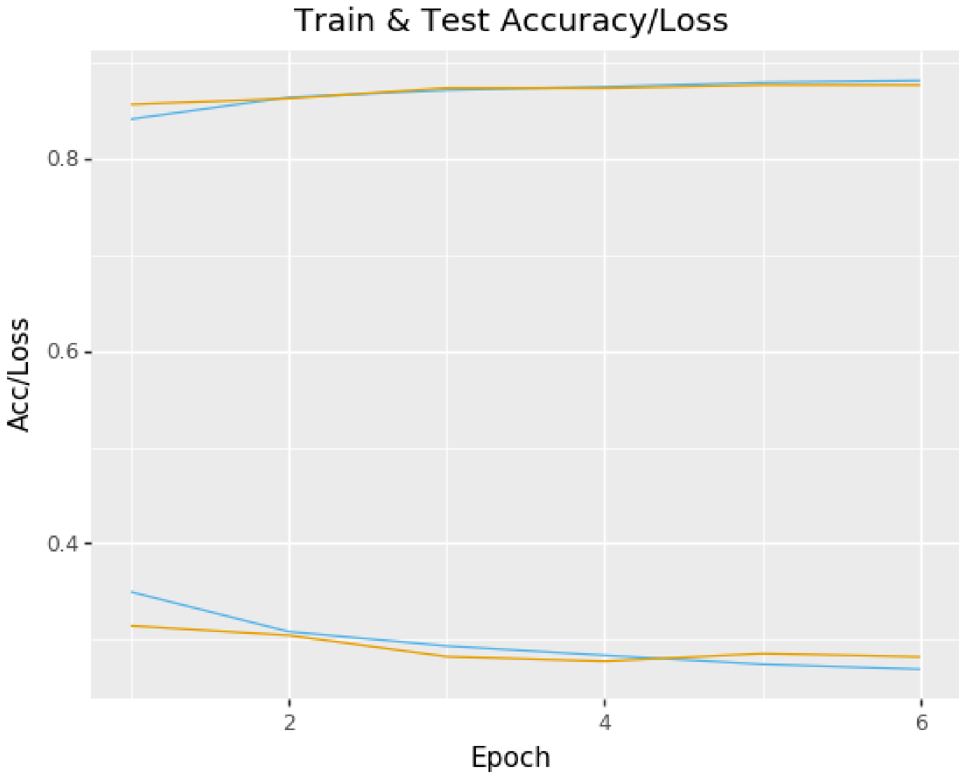
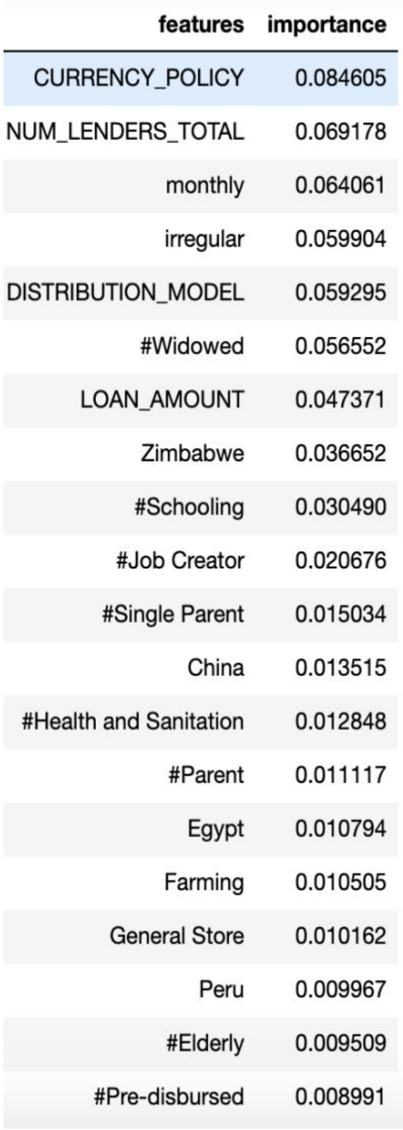


Figure 11 Neural network Loss & Accuracy curve



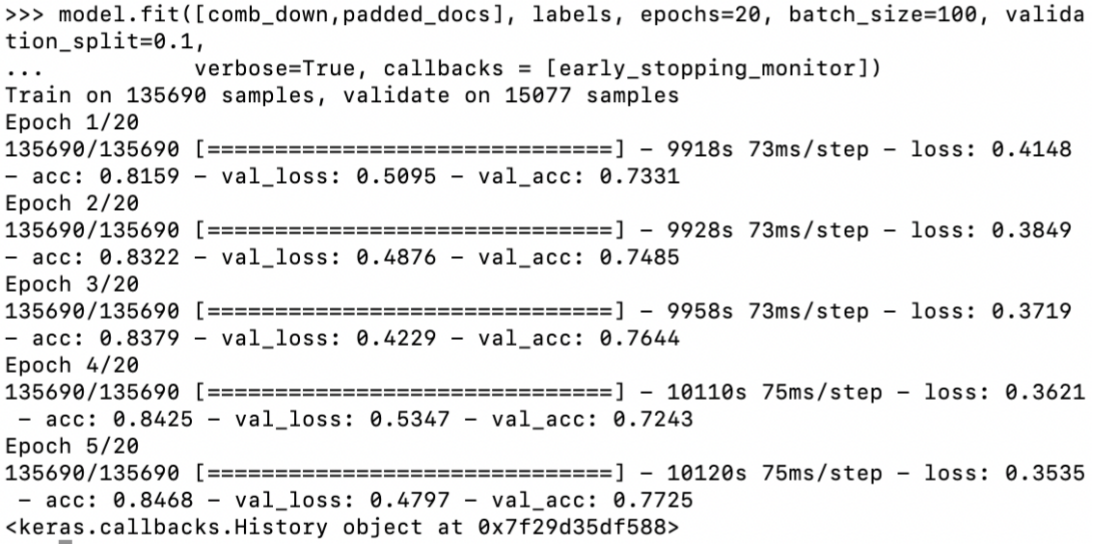
where blue is the training line, and yellow is the testing line

Figure 12 Random Forest Coefficients Importance Ranking



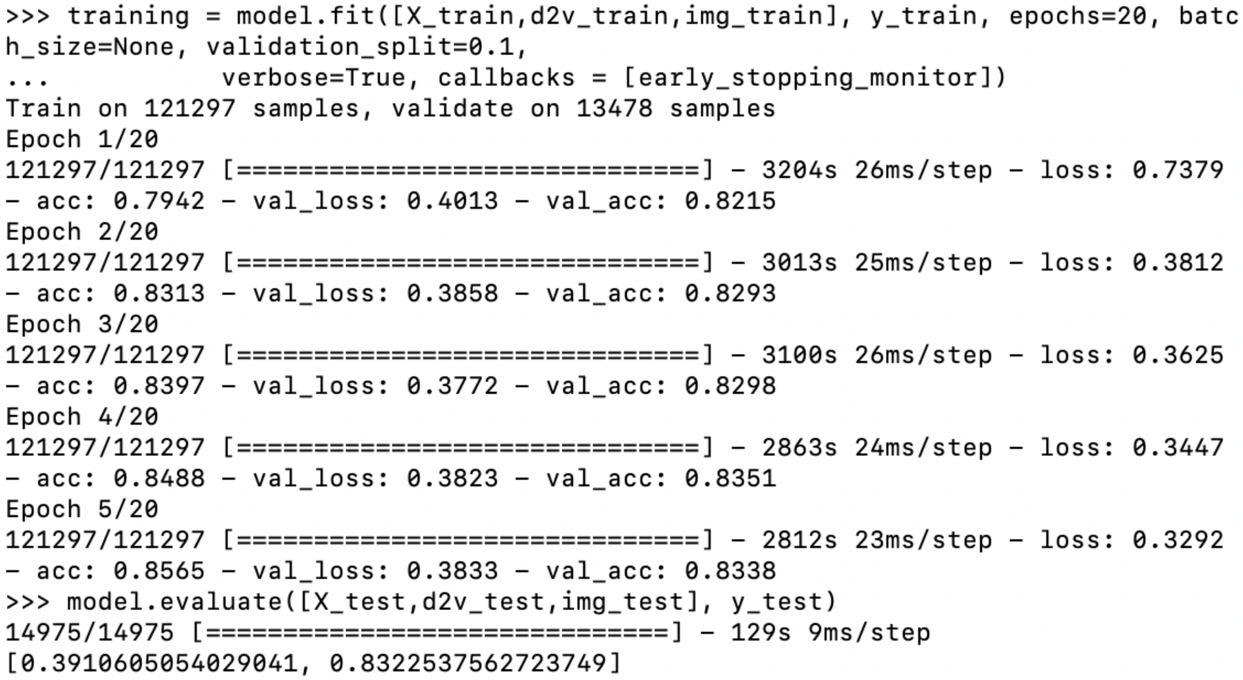
features	importance
CURRENCY_POLICY	0.084605
NUM_LENDERS_TOTAL	0.069178
monthly	0.064061
irregular	0.059904
DISTRIBUTION_MODEL	0.059295
#Widowed	0.056552
LOAN_AMOUNT	0.047371
Zimbabwe	0.036652
#Schooling	0.030490
#Job Creator	0.020676
#Single Parent	0.015034
China	0.013515
#Health and Sanitation	0.012848
#Parent	0.011117
Egypt	0.010794
Farming	0.010505
General Store	0.010162
Peru	0.009967
#Elderly	0.009509
#Pre-disbursed	0.008991

Figure 13 RNN for Other variables + Word2vec vectors



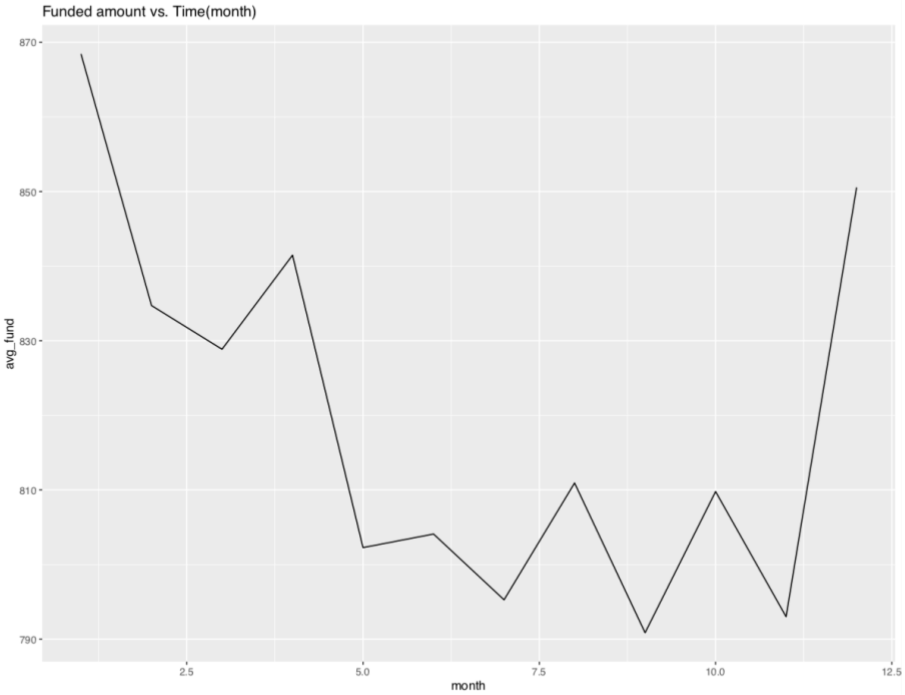
```
>>> model.fit([comb_down,padded_docs], labels, epochs=20, batch_size=100, valida
tion_split=0.1,
...           verbose=True, callbacks = [early_stopping_monitor])
Train on 135690 samples, validate on 15077 samples
Epoch 1/20
135690/135690 [==============================] - 9918s 73ms/step - loss: 0.4148
- acc: 0.8159 - val_loss: 0.5095 - val_acc: 0.7331
Epoch 2/20
135690/135690 [==============================] - 9928s 73ms/step - loss: 0.3849
- acc: 0.8322 - val_loss: 0.4876 - val_acc: 0.7485
Epoch 3/20
135690/135690 [==============================] - 9958s 73ms/step - loss: 0.3719
- acc: 0.8379 - val_loss: 0.4229 - val_acc: 0.7644
Epoch 4/20
135690/135690 [==============================] - 10110s 75ms/step - loss: 0.3621
 - acc: 0.8425 - val_loss: 0.5347 - val_acc: 0.7243
Epoch 5/20
135690/135690 [==============================] - 10120s 75ms/step - loss: 0.3535
 - acc: 0.8468 - val_loss: 0.4797 - val_acc: 0.7725
<keras.callbacks.History object at 0x7f29d35df588>
```

Figure 14 MLP for Other Variables + Doc2vec + Image Data in CNN deep features



```
>>> training = model.fit([X_train,d2v_train,img_train], y_train, epochs=20, batc
h_size=None, validation_split=0.1,
...           verbose=True, callbacks = [early_stopping_monitor])
Train on 121297 samples, validate on 13478 samples
Epoch 1/20
121297/121297 [==============================] - 3204s 26ms/step - loss: 0.7379
- acc: 0.7942 - val_loss: 0.4013 - val_acc: 0.8215
Epoch 2/20
121297/121297 [==============================] - 3013s 25ms/step - loss: 0.3812
- acc: 0.8313 - val_loss: 0.3858 - val_acc: 0.8293
Epoch 3/20
121297/121297 [==============================] - 3100s 26ms/step - loss: 0.3625
- acc: 0.8397 - val_loss: 0.3772 - val_acc: 0.8298
Epoch 4/20
121297/121297 [==============================] - 2863s 24ms/step - loss: 0.3447
- acc: 0.8488 - val_loss: 0.3823 - val_acc: 0.8351
Epoch 5/20
121297/121297 [==============================] - 2812s 23ms/step - loss: 0.3292
- acc: 0.8565 - val_loss: 0.3833 - val_acc: 0.8338
>>> model.evaluate([X_test,d2v_test,img_test], y_test)
14975/14975 [==============================] - 129s 9ms/step
[0.3910605054029041, 0.8322537562723749]
```

Average fund amount versus month



This the plot of average funded amount versus time in one year period, and there is a clear downward trend between the first month and September, which indicates that borrowers are more likely to obtain a larger amount of fund at the beginning of the year, and receive less amount of money during the summer.