Credit Risk Analysis

Ahmet E. Tekden and Enes Özipek

Abstract—**It's important for finance companies to know the risk of giving credit to their clients. Credit scores are essential for companies to decide on whether giving credit to someone will be a loss or not. In this project, we make an experiment on features of clients to decide on which ones are more important. SVM, Logistic Regression, Gaussian Naive Bayes(GNB), Multi-layer Perceptron(MLP) and Decision Trees are utilized and compared to each other. At last, an interface that calculates credit score by considering given features is provided.**

I. INTRODUCTION

Banks play a crucial role in market economies. They decide who can get finance and on what terms and can make or break investment decisions. For markets and society to function, individuals and companies need access to credit.

Due to high risks associated with inappropriate credit decisions that may result in huge amount of losses, financial institutions experience more serious challenges as time goes by. Credit coring algorithms are required to solve these challenges. These algorithms, which make a guess at the probability of default, are the methods banks use to determine whether or not a loan should be granted to a specific borrower.

Due to the nature of the problem, a bank should pay attention not to give a loan to risky customers. Therefore, we tried to maximize true negative rate while maintaining a healthy balance of accuracy.

II. RELATED WORK

Since credit scoring is a crucial issue, many people have some work and progress made on it. In recent papers, SVM, ANN, Logistic Regression and Decision Trees are utilized and compared with respect to their accuracies and Type I,II errors. These algorithms, which are used in these papers, are tested mostly on the datasets of Australian, China and German credit datasets, yet not all papers utilized these three ones. One of the results is as following:

Table 1 [1]

		Australian credit dataset		
		Average accuracy (%)	Type I error (%)	Type II error (%)
LRA	Mean	86.56	12.68	14.05
	SD	2.51	4.16	4.14
DT	Mean	84.39	18.00	13.70
	SD	2.75	5.67	3.97
ANN	Mean	83.28	19.27	14.68
	SD	3.03	5.33	4.37
SVM	Mean	85.67	7.20	20.04
	SD	2.71	2.94	4.41
		China credit dataset		
		Average accuracy (%)	Type I error (%)	Type II error (%)
LRA	Mean	72.07	18.49	43.23
	SD	5.40	7.10	10.39
DT	Mean	77.85	16.56	31.23
	SD	6.07	7.56	12.91
ANN	Mean	71.12	17.61	47.20
	SD	6.71	11.59	15.24
SVM	Mean	67.63	3.24	79.76
	SD	4.10	4.75	12.34
		German credit dataset		
		Average accuracy (%)	Type I error (%)	Type II error (%)
LRA	Mean	76.14	11.70	52.23
	SD	2.31	2.64	5.20
DT	Mean	72.10	17.06	53.20
	SD	2.76	3.35	6.94
ANN	Mean	71.43	19.32	50.17
	SD	2.59	3.41	6.73
SVM	Mean	76.28	10.57	54.40
	SD	2.19	2.66	5.42

Observing the results of these algorithms, we can see that type-II errors are higher on German and Chinese credit datasets. This is because of the fact that, in German and Chinese datasets, the frequency of reliable customers is higher.

Examining papers, we can see that they reached $70-80\%$ accuracy on their models, even though true negative rate fluctuates a lot between models.

Since we used *Give Me Some Credit Competition* dataset, we also reviewed interviews of winners and their accuracies. Winners reached about 86% accuracy in their works. However, generally, it might be concluded that preprocessing of the dataset was one of the significant aspects of training part.

III. Methods

Due to the nature/definiion of a problem, various algorithms might lead to different accuracy rates, we utilized several methods during this experiment.

A. Logistic Regression

Logistic regression is the naive approach someone might come up with when the dependent variable is dichotomous. It is used to describe data and to examine the relationship between one dependent binary variable and several ratio-level independent variables.

B. Decision Tree

Decision tree is a classification tool in the form of a tree structure which basically It divides dataset into smaller and smaller subsets until high purity level is obtained. The eventual result will be a tree with decisive level nodes and leaf nodes. Lastly, this tree is used to classify instances.

C. Gaussian Naive Bayes Classifier

Naive Bayes classifiers are simplistic probabilistic models. GNB is Naive Bayes classifier based on Gaussian distribution.

GNB works by calculating PDF of classes based on Gaussian distribution that is trained, then comparing these probabilities.

D. Support Vector Machines

Support vector machines are hyperplanes that define a decision boundary. Optimally, they try to maximize minimum distance between training data. Then, using these models, instances will be classified as one of the target classes.

E. Multilayer Perceptron

Multilayer perceptron is a feed-forward neural network with many hidden layers between input and output layer. It's a supervised algorithm. It is used to classify classes that are not linearly separable.

F. Hybrid classifier

This is the model that uses multiple learners at the same time. We used voting classifier for our purpose.

Voting classifier is an approach of labeling that takes probability values from all learners. Afterwards, it makes a classification based on ensemble of these probabilistic values.

IV. Experiments

In order to train our models and verify effectiveness of them, we used the dataset given by *Kaggle*. After preprocessing of data, algorithms mentioned in *Methods* section are used to measure the performance of these algorithms.

Higher cost is given to type-II errors, due to the fact that misclassification of risky borrower has more costly consequences then misclassification of good borrower does.

A. Dataset

We used the dataset of *Kaggle-Give Me Credit* competition. It has the following features.

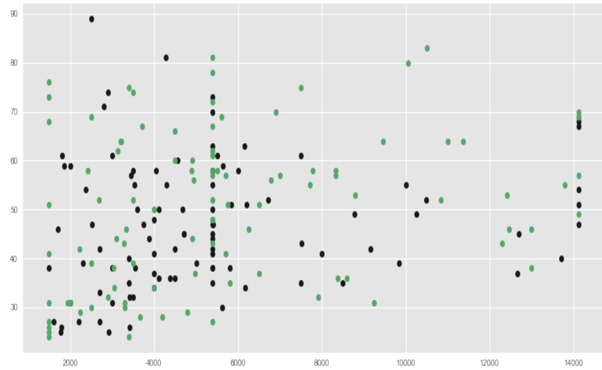
Data Dictionary

Variable Name	Type
RevolvingUtilizationOfUnsecuredLines	percentage
age	integer
NumberOfTime30-59DaysPastDueNotWorse	integer
DebtRatio	percentage
MonthlyIncome	real
NumberOfOpenCreditLinesAndLoans	integer
NumberOfTimes90DaysLate	integer
NumberRealEstateLoansOrLines	integer
NumberOfTime60-89DaysPastDueNotWorse	integer
NumberOfDependents	integer
SeriousDlqin2yrs	Y/N

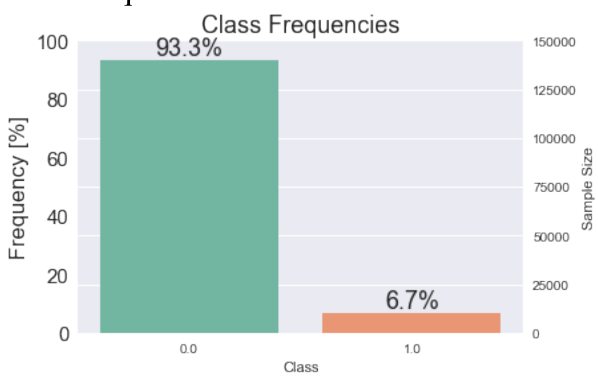
We tried to predict people who experienced 90 day delinquency. We created our models according to features given in the data dictionary.

B. *Visualization of Data*

Age vs Income graph:



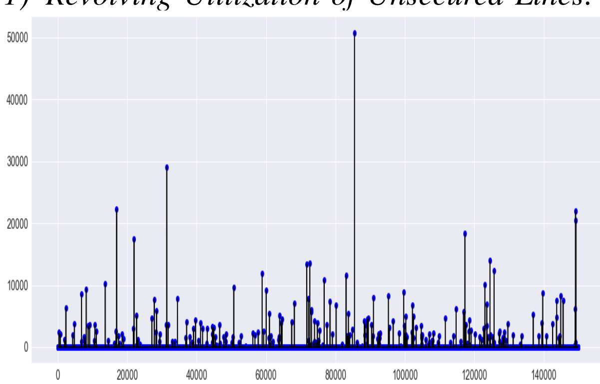
Class frequencies:



C. *Preprocessing*

Generally, in the real-life cases, data are not pure enough to start training process immediately. First of all, data should be preprocessed to eliminate outliers to relive the effect of them to accuracy rate.

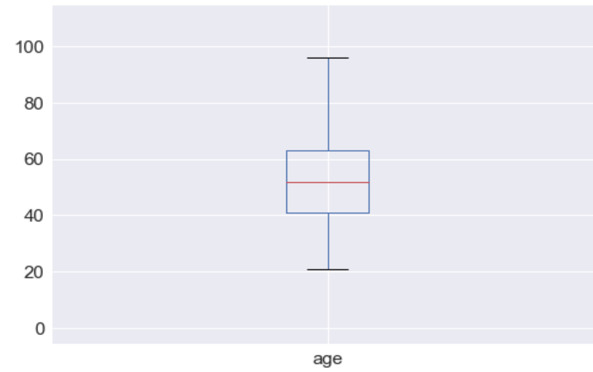
1) *Revolving Utilization of Unsecured Lines:*



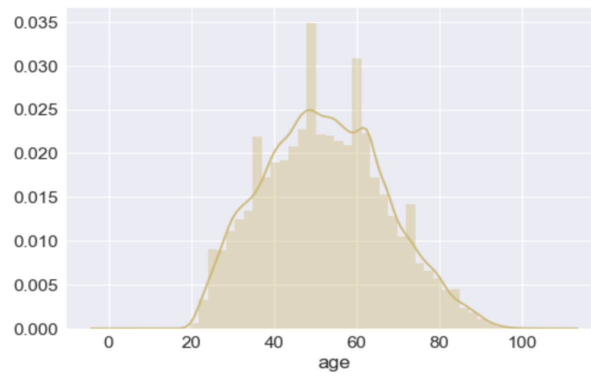
When we analyze the feature, we see that there are some outliers affecting our model adversely. Related statistics given below:

Median: 0.1541807
Mean: 6.0484381
Values less than 2(Ratio): 99.75267%

2) *Age:*



Age is also another important feature for credit risk analysis. In our dataset there are some ages like 0,1,105. These kinds of values are not normal in real life cases. Therefore, we tried to make them in a reasonable range by eliminating outliers or converging them to closest values.

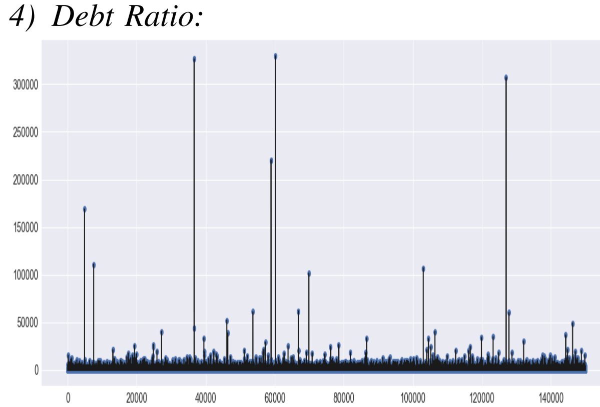


Median: 52

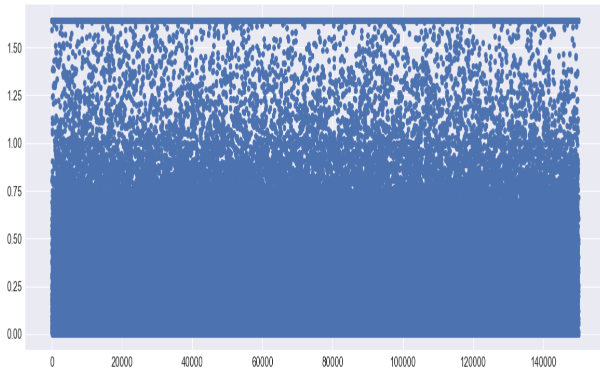
3) *Number of Time 30-59 Days Past Due:*

Another natural decisive factor of this experiment is to consider whether a borrower has paid her/his debt on time. When we analyze the feature, we encounter that there are two outliers '96 and 98' since the closest value to them is 13. We again eliminate the outliers.

4) *Debt Ratio:*



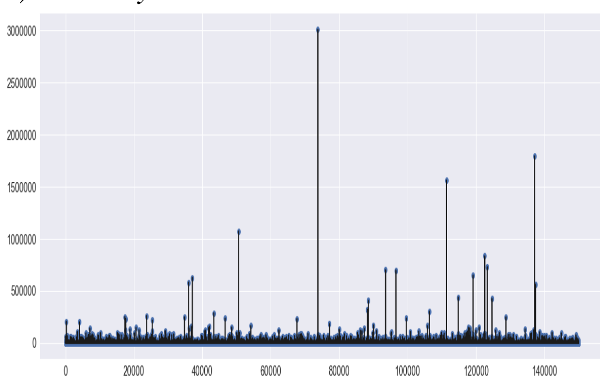
Debt ratio is also significant feature of credit risk analysis. Therefore, we need to eliminate outliers again.



Related statistics given below:

count 150000
mean 0.607359
std 0.588294
min 0.000000
25% 0.175074
50% 0.366508
75% 0.868254
max 1.641791

5) Monthly Income:



Monthly income is another decisive factor used by financial institutions. After examining the feature and eliminating its outliers, we obtained the following statistics for new income set.

count 150000
mean 5993.580100
std 3164.949126
min 1500.000000
25% 3903.000000
50% 5400.000000
75% 7400.000000
max 14128.000000

6) Number of Dependents:

In this experiment, we also considered the number of people borrower look after. The more the number is the harder she/he is provided with the loan. When we analyze the feature, there are some outliers. We replaced them with 10 which is maximum value for the new feature.

7) Other Features:

There are other features that we will not cover in details. They are *Number of Times 90 Days Late*, *Number Real Estate Loans or Lines*, *Number of Time 60-89 Days Past Due*. They are filtered with the approach we used for *Number of Time 30-59 Days Past Due*.

D. Comparison of Models

We used many machine learning algorithms to train our model.

1) Logistic Regression:

First algorithm we tried on our dataset was logistic regression. When we try to fit data by using equal prior probabilities, yet, we couldn't get a satisfying result due to the over-fitting data. Therefore, we adjusted class weights. Result we get is summarized as confusion matrix below:

	Class 0	Class 1
Class 0	61853	8164
Class 1	936	4047

2) Decision Tree: Second algorithm we tried on our dataset was decision tree.

At first tries, we trained our model with no specification. Here, We get really low FN rate. After analyzing our model, we saw that our data was over-fitting sample.

To overcome this, we lowered impurity threshold. Furthermore, we also limited max depth of our model.

Our decision tree can be seen in figure 1

	Class 0	Class 1
Class 0	50585	19432
Class 1	936	4047

3) Gaussian Naive Bayes Classifier: After preprocessing, it was possible to use gaussian naive bayes classifier on our datasets.

Since our dataset was imbalanced, we put custom weights on our classes.

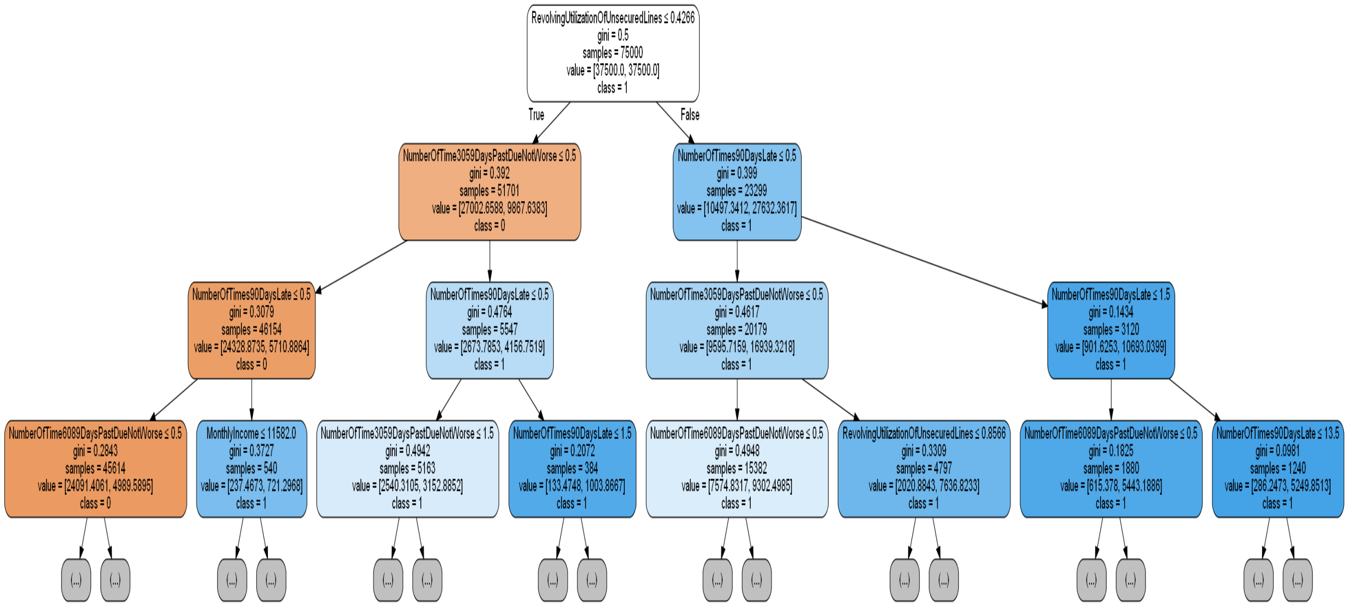


Fig. 1. Figure 1, First three Layer of Decision Tree

Overall, we got good accuracy from these model.

	Class 0	Class 1
Class 0	56333	13684
Class 1	1219	3764

4) Support Vector Machines: This is fourth algorithm we run on our dataset. Since frequency of our data is imbalanced. We couldn't get high false positive rate. Yet it still give high true positive rate.

	Class 0	Class 1
Class 0	60033	9984
Class 1	3432	1551

5) Multi Layer Perceptron: Multi layer perceptron gave different results on different runs. It's performance fluctuated, but nonetheless, it gave not a bad result.

We run our multi layer perceptron with identity activation function and 100 hidden nodes.

	Class 0	Class 1
Class 0	68239	1778
Class 1	3358	1625

6) Hybrid classifier: Since we created many algorithms with at least above average performance, we also wanted to use an hybrid algorithm to see its performance.

Since some of these algorithms performed better, we gave them relatively higher weight. Results we obtained were as followed.

	Class 0	Class 1
Class 0	57103	12914
Class 1	1287	3696

V. Conclusion/discussion

Credit risk analysis is an important challenge in machine learning. Many people works on it and they try to improve the state of art. Machine learning algorithms becomes vital and attracts more attention than it has attracted ever.

We analyzed different machine learning algorithms to compare how they perform. Saying exactly which algorithms is best is hard in this context since cost of type 1 error and type 2 error are different. But knowing these costs, one can decide on which algorithms to use.

Comprehensive results of our algorithms are on table 1. We only included algorithms that performed above average.

Apart from these, we also tried algorithms like Knn, boosting etc. But their performance was not so promising to make us mention about them in *Methods* section. Especially, we expected KNN

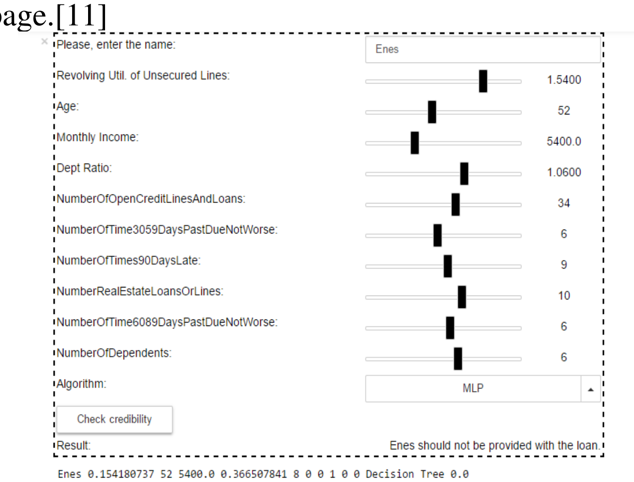
Algorithm	Accuracy(%)	Type-I Error(%)	Type-II Error(%)
LogReg	87.86	11.66	18.78
DT	72.84	27.75	18.78
GNB	80.13	19.54	24.46
SVM	82.11	14.25	68.87
MLP	93.15	2.53	67.38
Hybrid	81.06	18.44	25.82

TABLE I

COMPREHENSIVE RESULTS

to perform better, but probably since our data was imbalanced, it didn't perform as good as we anticipated.

We prepared a GUI for users to test different instances to check if an instance denotes a risky customer or not. It can be reached in our Github page.[11]



Some Results(0: give loan, 1: don't give loan):

Enes 0.15418 52 5400 0.3665 8 0 0 1 0 0 Decision Tree 0.0

Enes 0.15418 52 5400 0.3665 8 0 0 1 0 0 Logistic Regression 0.0

Enes 0.15418 52 5400 0.3665 8 0 0 1 0 0 GaussianNB 0.0

Enes 0.15418 52 5400 0.3665 8 0 0 1 0 0 MLP 0.0

Enes 1.54 52 5400.0 1.06 34 6 9 10 6 6 Decision Tree 1.0

Enes 1.54 52 5400.0 1.06 34 6 9 10 6 6 Logistic Regression 1.0

Enes 1.54 52 5400.0 1.06 34 6 9 10 6 6 MLP 1.0

Enes 1.54 52 5400.0 1.06 34 6 9 10 6 6 GaussianNB 1.0

REFERENCES

[1] Gang Wang, Jinxing Hao, Jian Ma, Hongbing Jiang, A comparative assessment of ensemble learning for credit scoring, Expert Systems with Applications, Volume 38, Issue 1, January 2011, Pages 223-230, ISSN 0957-4174,https://doi.org/10.1016/j.eswa.2010.06.048.

[2] Tian-Shyug Lee, I-Fei Chen, A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines, Expert Systems with Applications, Volume 28, Issue 4, May 2005, Pages 743-752, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2004.12.031.

[3] Baesens, B., Van Gestel, T., Viaene, S. et al. J Oper Res Soc (2003) 54: 627. doi:10.1057/palgrave.jors.2601545

[4] Chih-Fong Tsai, Jhen-Wei Wu, Using neural network ensembles for bankruptcy prediction and credit scoring, Expert Systems with Applications, Volume 34, Issue 4, May 2008, Pages 2639-2649, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2007.05.019.

[5] David Martens, Bart Baesens, Tony Van Gestel, Jan Vanthienen, Comprehensible credit scoring models using rule extraction from support vector machines, European Journal of Operational Research, Volume 183, Issue 3, 16 December 2007, Pages 1466-1476, ISSN 0377-2217, https://doi.org/10.1016/j.ejor.2006.04.051.

[6] David West, Neural network credit scoring models, Computers and Operations Research, Volume 27, Issues 1112, September 2000, Pages 1131-1152, ISSN 0305-0548, https://doi.org/10.1016/S0305-0548(99)00149-5.

[7] Hand, D. J. and Henley, W. E. (1997), Statistical Classification Methods in Consumer Credit Scoring: a Review. Journal of the Royal Statistical Society: Series A (Statistics in Society), 160: 523541. doi:10.1111/j.1467-985X.1997.00078.x

[8] http://blog.kaggle.com/2012/01/03/the-perfect-storm-meet-the-winners-of-give-me-some-credit/

[9] http://blog.kaggle.com/2011/12/21/score-xavier-conort-on-coming-second-in-give-me-some-credit/

[10] http://blog.kaggle.com/2011/12/20/credit-where-credits-due-joe-malicki-on-placing-third-in-give-me-some-credit-2/

[11] https://github.com/enesozi/Credit-Risk-Analysis