Paper 484-2017

Using Text Analysis to Improve the Quality of Scoring Models with SAS® Enterprise Miner™

Piotr Małaszek, Warsaw University of Life Science

ABSTRACT

Transformation of raw data into sensible and useful information for prediction purposes is a priceless skill nowadays. Vast amounts of data, easily accessible at each step in a process, gives us a great opportunity to use it for countless applications. Unfortunately, not all of the valuable data is available for processing using classical data mining techniques. What happens if textual data is also used to create the analytical base table (ABT)?

The goal of this study is to investigate whether scoring models that also use textual data are significantly better than models that include only quantitative data. This thesis is focused on estimating the probability of default (PD) for the social lending platform kokos.pl. The same methods used in banks are used to evaluate the accuracy of reported PDs. Data used for analysis is gathered directly from the platform via the API. This paper describes in detail the steps of the data mining process that is built using SAS® Enterprise Miner™.

The results of the study support the thesis that models with a properly conducted text-mining process have better classification quality than models without text variables. Therefore, the use of this data mining approach is recommended when input data includes text variables.

INTRODUCTION

There is no doubt that these days the loan market is important for everybody due to the omnipresent money need both for investments and daily necessities. Hardly anyone at a young age can take the liberty to buy property or a new car. Also, small and medium-sized enterprises have problems with funding for their ideas. In both cases the biggest barrier is usually the credit history of the borrower. Since the last financial crisis of 2007–08, it is observable that from year-to-year banks have less proclivity to lend money to such customers. This fact had doubtless a huge impact in the development and popularization of non-bank financial institutions.

Thus, worth mentioning is the emerging alternative to banks: the social lending market (P2P Lending), which allows individuals to lend or borrow money to each other directly without the participation of traditional financial intermediaries (Chen, D. Chaodong, H. 2012). It is a great opportunity for borrowers to easily lend money without unnecessary formalities, but is it really a safe place to invest money? On the Internet it is easy to find entries about painful experiences with social lending platforms. The first social lending platform in Poland was kokos.pl. On their home page, they inform that the average return on investment is 16,94% and during 8 years over 141 million PLN was borrowed through this portal on 125100 auctions (data from 1 SEP 2016). Unfortunately, like most investments with a high return rate, it's connected also with high risk. After analyzing data from this platform, it emerged that up to 17,5% of loans had problems with timely repayment.

Data mining, which improves every year, can be used in this situation. Thanks to the development of technology which allows for faster and more accurate decision-making, there is the possibility to build a decision support system which will help the decision-maker use the data to solve unstructured problems. The most often applied credit risk assessment models use only quantitative data. Very few models used for this purpose a text mining approach. In the case of P2P loans, the success of the transaction of course depends on identity verification or high income, but also it seems that investors should take care to review the description of loan's purpose. Thereupon, it sounds like a good idea to include in the input data variables from text mining (Tkach, D. 1997) and investigate whether such an approach impacts or improves the predictive ability of the model.

In this paper is shown step by step how to effectively use data made available by the platform to increase the probability of successful investments by not investing in loans if the likelihood that an obligor will default

is high. The models were trained on data which were known at the moment of decision making so it is worthwhile to say that such models can be used in real time. The approach described in this work is quite universal and allows the use of this method for other applications like fraud detection or in marketing. To assess the creditworthiness of borrowers at kokos.pl, three of the most popular models were used: logistic regression, neural network and decision tree.

In my project, SAS software was used to explore data, prepare it for analysis, and finally to estimate and compare data mining models.

Exploration and preparation of the data were done with the SAS® Enterprise Guide® 7.1 using procedures like PROC IMPORT, PROC SORT and others.

All transformations of variables, text mining with %TEXTSYN macro and classification models were built in SAS Enterprise Miner version 14.1.

Finally, comparison of the models and chart creation has been done in Base SAS® version 9.4 using PROC LOGISTIC including ROCCONTRAST.

SOCIAL LENDING

The first company to offer P2P loans in the world was Zopa from London launched in 2005. Less than two years later the first in the US was founded, PROSPER, which after 9 months from release had over 100 thousand users. Social lending came to Poland in 2008, 3 years after Zopa was created. Today on the Polish market together with kokos.pl are lendico.pl, zakra.pl, and finansowo.pl, so it is not difficult to observe the steady growth of this market. It owes its popularity to the fact that it offers a more cost-effective alternative to banks (Galloway, I. 2009). Credit risk is shared in project-specific pools of lenders and each member funds a small share of the financed amount. As compensation for taking risk, interest is paid to the lenders, whereas platform operators typically charge fixed fees (Bohme, R. Potzsch, S. 2010).

PROBABILITY OF DEFAULT

Platform owners claim that peer-to-peer lending is beneficial for both borrowers and investors. They argue the fact of eliminating expensive financial intermediaries which reduces the cost of loans. But it is good to ask yourself - are unknown borrowers sufficiently reliable to lend them your money?

Under Basel II (Engelmann, B. Rauhmeier, R. 2006), a person is classified as a default when:

- the obligor is more than 90 days past due on a material credit obligation;
- it is unlikely that the obligor will be able to repay its debt to the bank without giving up any pledged collateral.

On the other hand, kokos.pl platform gives following definition of default:

If the borrower does not repay its obligations, Kokos.pl issues notifications requesting repayment. In the case of non-payment of two installments in a row, the borrower is sent a final request for payment within 7 days. After that, the term loan gives a value in foreclosure and can be transferred by investors to debt collection on Kokos.pl. The investor can also work alone to recover the investment or sell the debt in the aftermarket.

In the studied data set with 54719 analyzed auctions, 9550 of them were not repaid according to the agreement meaning almost in 1 of 5 loans customers defaulted. This means that a model detecting auctions which probably will have problems with repayment can be very useful for every investor.

DATA PREPARATION

The most time-consuming part of the majority of analytical projects is data preparation. Word is that it consumes about 80% of a data scientist's time. The following section describes this process starting from downloading data, through data pre-processing and ending with computing new variables.

Data for analysis were gathered directly from the kokos.pl platform via the API. Collected data came from the period since establishment of the service in 2008 to 6 December 2015. This means that the results of

the models tend to be unaffected by changes in the business cycle (Engelmann, B. Rauhmeier, R 2006). This approach to measure obligor's PDs is called through-the-cycle philosophy (TTC). Downloaded data in csv format was imported using SAS Enterprise Guide and then stored in SAS datasets and processed by 4GL code.

After cleaning the data, the crucial task was to choose relevant observations, because the downloaded data set contained auctions with different statuses. On the platform 11 various statuses of auctions are defined. For example, "New loan. Auction in progress", "0% of the investments have been collected", "Money collected. Money has been sent. Repayment in progress" or "Loan repaid". From those 11 possible statuses, only two of them represent a situation when the money actually has been sent to the borrower. It is: "Money collected. Money has been sent. Repayment in progress" or "Loan repaid". For the other 9 statuses, a loan may not be in default, so such auctions have been omitted in further analysis. It is analogical to credit risk assessment by the Bank where the scoring model is built only on the sample of customers who the bank lends money.

After that, the next task was to compute new variables. Some of variables needed only binary transformation, but some additional variables were created which were very likely to be useful for the classification model. Examples of such variables are: age of borrower, number of days from the moment of registration or how many loans were taken by the borrower.

TEXT MINING

Text mining is the field of data science having tremendous growth over the last few years. However, most of the data mining models still use only structured quantitative data. Regardless of that, during the analysis of the data available on the site kokos.pl one of the first ideas was to analyze the loan's description. The aim of such an approach was to check whether plain text provided by the borrower in the auction has a significant impact on the repayment of such loan. In other words, whether emotions, phrases, and terms used by the borrower in the loan description contain important information that could be used by the model to better PD detection.

This approach uses text mining in order to transform unstructured text data into variables which will be acceptable for data mining models and use them together with the structured data. With this end in view, the following nodes provided SAS Enterprise Miner were used: Text Parsing, Text Filter, Text Cluster and Text Topic.

MODEL ASSESMENT

In total, 6 models have been estimated. Logistic regression, neural network and decision tree based on data that included only quantitative variables and three the same models which additionally used variables obtained in the text mining process. The aim of this chapter is to describe how the model with the best discriminatory power was chosen.

One of the most popular measures of model fit is the area under the Receiver Operating Characteristic (ROC) curve (AUC). The ROC curve is a visual tool that can be easily constructed if two representative samples of scores for defaulted and non-defaulted borrowers are available (Basel Committee on Banking Supervision 2005). Before the ROC curve will be presented it is worth taking a look at the exemplary model scores distribution of defaulted and non-defaulted loans in Figure 1. In the ideal world, the score distribution for defaulted and non-defaulted loans should be disjoint. The result of the earlier mentioned 6 models, is a probability that particular loan will not be repaid. This is called the score and has a value from 0 to 1. Therefore, the important part of modeling is to choose the appropriate cut-off point (C) which is a value from 0 to 1. If $y < C$ then the observation is marked as non-defaulted and if $y > C$ then it is marked as a defaulted. Based on such an approach we can get four different scenarios. If the score of a particular loan is lower than the cut-off point and in fact this loan was repaid contractually, then the decision is correct. However, if the score is higher than the cut-off point then decision is incorrect. Analogously if the score is higher than the cut-off point and the loan was not repaid then the decision was correct. Otherwise, if the score is lower than the cut-off point the decision is incorrect. This means that an inevitable part of making any decision is the risk of error.

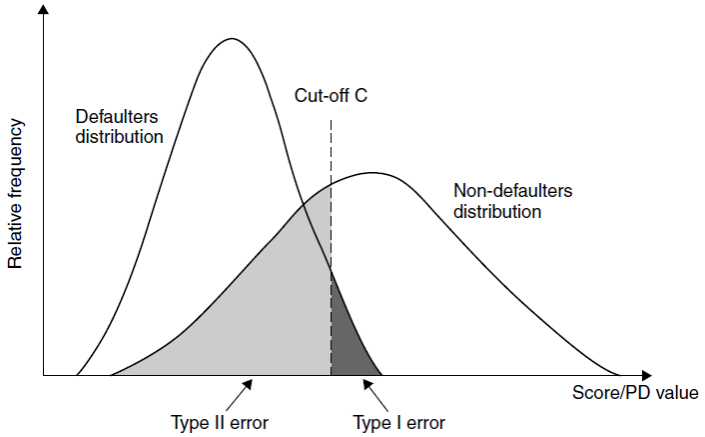


Figure 1. Distribution of rating scores for defaulting and non-defaulting debtors
Source: Izzi, L. Oricchio, G. Vitale, L. (2012)

For the situation described above two types of errors are possible:

- Type I error - percent of "bad" borrowers wrongly foreseen as non-defaulted. This error usually generates losses related to costs incurred to recover money owed.
- Type II error – percent of "good" borrowers misclassified as default. This conversely, produces more limited losses originating from the failed earnings in terms of fees and interest margin caused by the wrong classification of a good borrower as a future insolvent (Izzi, L. Oricchio, L. Vitale L. 2012).

The ROC curve presented in Figure 2 has the following structure. For each of the possible cut-off points a confusion matrix is created. Next a sensitivity (hit rate or true positive rate), and 1-specificity (false alarm rate or false positive rate) is calculated for a lot off cut-off points and then in such a way are then placed on the graph and linked together. The more points, the ROC curve is more smooth. Based on such an ROC curve, the area under curve (AUC) is calculated. Area A is an area for a random model without discrimination power and it has a 0.5.rating. A model's performance is better the steeper the ROC curve is at the left end and the closer the ROC curve's position is to the point (0,1). An ideal model has an AUC equal to 1 (Basel Committee on Banking Supervision 2005).

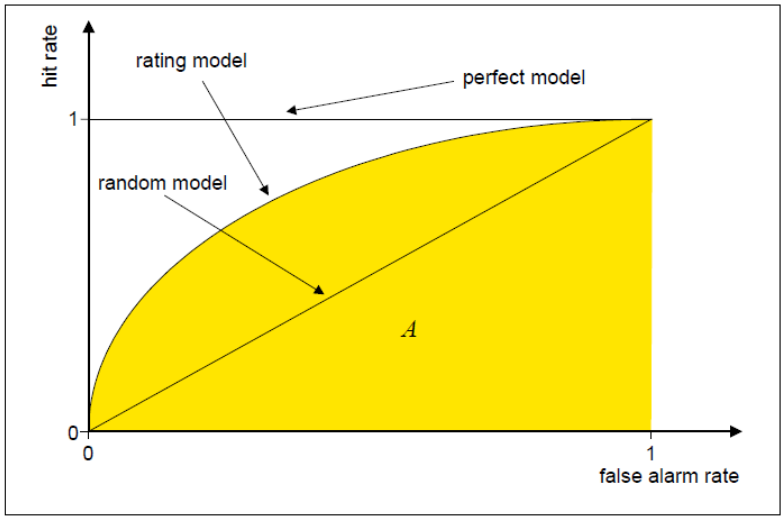
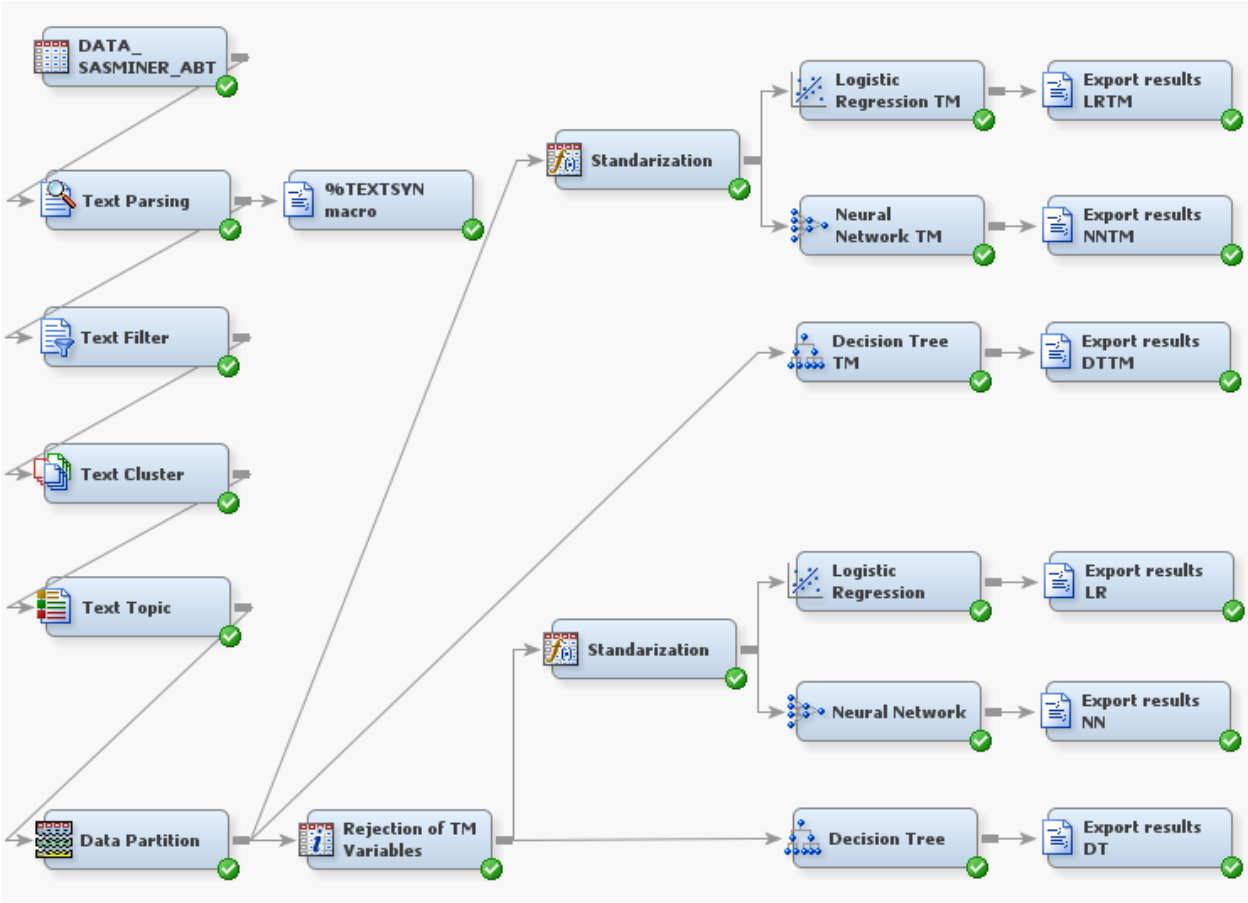


Figure 2. Receiver Operating Characteristic curves
Source: Basel Committee on Banking Supervision

RESULTS

This section presents the results of the text mining process and model assessment. The Display 1 below graphically represents the diagram built in SAS Enterprise Miner. In the top left corner is placed the data set with input data. After that were added text mining nodes in order to create additional variables which allow a model for classification with better quality. In meantime was also run the node with SAS code creating set of synonyms which was used by Text Parsing node. After creation of text mining variables, the data was partitioned into training and validation sets in proportions 70:30 in order for unbiased assessment. To estimate models with and without text variables, it was necessary to reject text mining variables for three of the models using a metadata node. The last step before modelling was to standardize the variables using a Transform Variables node. After model estimation the results were exported to SAS tables using SAS Code nodes and were compared in Base SAS using PROC LOGISTIC procedure including ROCCONTRAST.



Display 1. Diagram in SAS® Enterprise Miner™ 14.1

TEXT MINING

The following subsections describe the approach to the text mining process. In order to extract meaning hidden in unstructured textual data the following nodes were used in the analysis: Text Parsing, Text Filter, Text Cluster, Text Topic. Each subsection contains information about the node's properties, values and description of benefits of using nodes.

Text Parsing

The first node used in the text mining process is Text Parsing which enables the ability to: identify different part of speech, define a stop list with terms to drop, remove punctuation, treat different terms with the same root as equivalent, define list of synonyms, and so forth.

For the proper analysis of text variables it is important to choose only the part of speech which contains important information. In the analyzed data set the best results were obtained when the Text Parsing node analyzed only nouns and adjectives. Additionally, a positive effect on the model accuracy was to turn on the *Stem Term* option and adding as a stop list a set of low-information polish words.

The Text Parsing node enables you to add a Data Set with synonyms. In this project a synonym data set was created using the %TEXTSYN macro which corrects misspellings that appear in the input data source by using the code below:

```
%TEXTSYN(docds=emws.TEXTparsing_train, termds= emws.TEXTparsing_TERMS,
         outds=emws.TEXTparsing_tmOUT, synds=bib.textsyn,
         textvar=description1, mnpardoc=8, mxchddoc=10);
```

Part of created table with synonyms is presented in Table 1. Each row represents a document where a value from the column “Term” will be treated as a value from column “parent”. The example below shows only the proper correction of words: ring, first, beautiful, but every user should be aware of the need to verify this table because sometimes suggestions might be not fully correct.

	example1	Term	parent	childndocs	numdocs	minsped
2290	... guzik. Brakuje mi na !!pierscionek!! i ogolni...	pierscionek	pierścionek	6	67	9
2291	... przeznaczone beda na zakup !!pierscionka!...	pierscionka	pierścionka	2	20	9
2292	... sprawdzić jak działa kokos !!piersza!! pożyc...	piersza	pierwsza	1	1443	7
2293	... pare dni deklaruję spłłate !!pierszej!! raty za...	pierszej	pierwszej	3	412	6
2294	... Swieta mojego dziecka , !!pierszy!! raz orga...	pierszy	pierwszy	3	2570	7
2295	... się tak jak w !!pierwse!! aukcji o szybsza sp...	pierwse	pierwsze	1	681	7
2296	... szybsza splata jak przy !!pierwsej!! pozyczc...	pierwsej	pierwszej	2	412	6
2297	... wcześniejszej spłaty.Drodzy Inwestorzy, ni...	pierwsz	pierwszy	3	2570	4
2298	... miała być na zakup !!pierściona!! tak też się...	pierściona	pierścionka	1	20	5
2299	Jest to moja !!piewsza!! pożyczka w tym serw...	piewsza	pierwsza	1	1443	7
2300	... pożyczenie od kogoś tych !!pieńiędzy!!. Poż...	pieńiędzy	pieniędzy	1	1816	11
2301	... git..ara, 5 zielonych tyż !!pikne!!, a na piero...	pikne	piękne	1	14	10
2302	... załować. I taki byłby !!pikny!! świat, lecz nie...	pikny	piękny	1	49	10
2303	... o wyrozumiałość.Sprawa jest dość !!pilnaZ!...	pilnaz	pilna	1	419	6
2304	... betonu, !!pily!! do ciecia betonu na...	pily	pilny	2	655	12

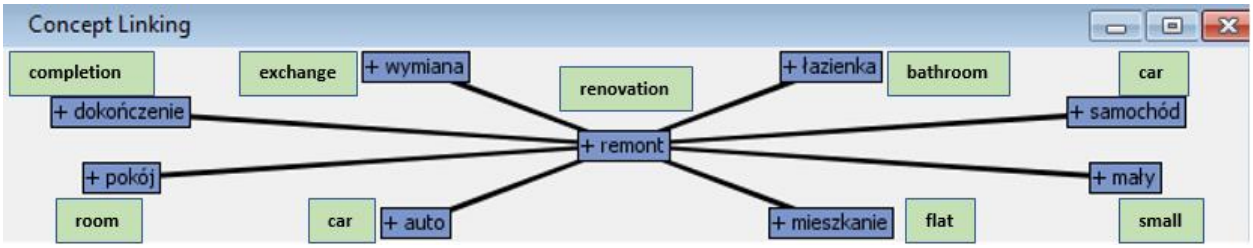
Table 1. %TEXTSYN macro results.
Source: Output from SAS® Enterprise Guide® 7.1

Text Filter

The main objective of this node is to filter terms that will be irrelevant in further analysis. According to Zipf’s law a relationship between the rank of the word and the frequency of occurrence can be observed (Chakraborty, G. Pagolu, M. Garla, S. 2013). Therefore, the *Maximum Number of Documents* was set to 10 and several dozen terms were detected on the *Number of Documents by Frequency* plot which occurred to many times have been manually added to stop list.

This node provides an Interactive Filter Viewer which allows you to add synonyms like auto – samochód (in English: auto - car) which mean exactly the same thing, but this synonym was not created by %TEXTSYN macro due to the completely different construction of the word. Synonyms can be easily added by selecting two or more terms, click RMB and select Treat as Synonyms.

Another useful feature of the Interactive Filter Viewer in the Text Filter node is the view Concept Linking option as is shown in Display 2. It allows you to detect the relationship between terms that are highly associated with the selected term and identify if there is a need add new synonyms.



Display 2 Concept Linking window for word renovation in SAS® Enterprise Miner™ 14.1

Text Cluster

The first node creating additional variables in the analyzed data set which are subsequently used by the model is the Text Cluster node. The aim of this node is to create clusters that will help with identifying the desired value of the target variable. Documents are assigned to mutually exclusive clusters so each document can belong to only one cluster which is described by a set of terms.

Text Cluster gives the possibility to use two algorithms: *Expectation Maximization* and *Hierarchical clustering*. Both approaches rely on the Singular Value decomposition (SVD) to organize terms and documents into a common semantic space based upon term co-occurrence (Berry, M. Kogan, J. 2010).

When documents are parsed, a frequency matrix is generated. Depending on the application, the user can define the number of dimensions. For the purpose of text segmentation, a recommended number of dimensions ranges from 2 to 50, but for prediction and classification higher values from 30 to 200 are used. The best results were obtained with the settings as in Table 2.

PROPERTY	VALUE
Transform	
SVD Resolution	High
Max SVD Dimensions	100
Cluster	
Exact or Maximum Number	Maximum
Number of Clusters	40
Cluster Algorithm	Expectation-Maximization
Descriptive Terms	5

Table 2. Text Cluster node settings.

Text Topic

The Text Topic node enables a user to explore the document collection by automatically associating terms and documents according to both discovered and user-defined topics. Topics are collection of terms that describe and characterize a main theme or idea (SAS® Institute Inc.). In contrast to clustering which assigns each document to a unique group, the Text Topic node assigns a score for each document and term to each topic. It is worth mentioning that one document or term can belong to more than on topic. The *Number of created Single-term Topics* and *Number of Multi-term Topics* can be parametrized. In this work those values were set consecutively to 25 and 50.

In Display 3 is shown *Interactive Topic Viewer* content. The topics were sorted in descending order by documents frequency. Selected topic consists of the following terms: repayment, auction, rating, investment, month, earlier, subsequent etc., so it is easy to notice that terms are generally connected with documents where the borrower tried to show prospective investors that the repayment will be quick and without any problems. In the Document section are listed all documents with assigned weights which show how correlated each document is with the selected topic. Additionally, in yellow are highlighted documents that have a topic weight higher than the cut-off level.

Topics

+repayment,+auction,+rating,+investment,+month

Topic	# Docs	Term Cutoff	Document Cutoff
+spłata,+aukcja,+rating,+inwestycja,+miesiąc	9024	**0.014**	**0.1**
+remont	3859	**0.001**	**0.001**
+remont,+łazienka,+dokończenie,+kuchnia,+mieszkanie	3823	**0.012**	**0.191**
+pozyczke,+prosze,sie,+pozyczki,juz	3594	**0.013**	**0.092**
+praca	3556	**0.001**	**0.001**
+praca,+zbędny,+samochód,+auto,+naprawa	3399	**0.013**	**0.099**
+potrzebny	2713	**0.001**	**0.001**

Terms

Topic Weight	+	Term	Role	# Docs	Freq
0.672	+	spłata	Rzeczownik	12496	15557
0.372	+	aukcja	Rzeczownik	7414	10051
0.333	+	rating	Rzeczownik	4134	4281
0.189	+	inwestycja	Rzeczownik	5089	5712
0.146	+	miesiąc	Rzeczownik	7006	8921
0.116	+	wcześniejszy	Przymiotnik	2608	2804
0.094	+	kolejny	Przymiotnik	2919	3196

Documents

Topic Weight	description1	TextCluster_SVD1	TextCluster_SVD2
0.1	Druga pozyczka w Kokos ,budowanie ratingu ,splata dwoch rat od razu	0.01131775338510572	-0.13285671923581305
0.1	Prawdopodobnie pożyczka zostanie spłacona do końca roku. Niezbędne do	0.008954525823045817	-0.19130312108341277
0.1	Spłata pozostałych rat co zostanie to na remont samochodu przed zimą.	0.005886773711469865	-0.35026172725013716
0.1	Pożyczka dobra opinie i rating	0.01217944432114565	-0.11244010354566271
0.099	Podniesienie ratingu w kokos.pl, test systemu i cele inwestycyjne.	0.013515132035336547	-0.11504030185958322
0.099	Chciałabym pożyczyć 1000 zł na spłatę debetu w rachunku bankowym.	0.01034338301192577	-0.10193606048448377

Display 3. Output from Topic Viewer in SAS® Enterprise Miner™ 14.1

CLASSIFICATION MODELS ASSESSMENT

Results of the 6 estimated models are compared to assess if models using the text mining variables performed better. Values of chosen statistical measures are presented in Table 3.

Measures \ Model	Logistic Regression TM	Neural networks TM	Decision tree TM	Logistic Regression	Neural networks	Decision tree
AUC (area under ROC curve)	0,795	**0,797**	0,781	0,779	0,789	0,775
ACC (accuracy)	0,843	0,843	**0,849**	0,840	0,845	0,848
MR (misclassification rate)	0,157	0,157	**0,151**	0,160	0,155	0,152
TPR (true positive rate)	**0,297**	0,256	0,257	0,288	0,251	0,289
FPR (false positive rate)	0,042	0,032	**0,026**	0,044	0,029	0,031

Table 3. Models comparison

Results from the table above unambiguously show that including text variables into the input data set had an influence on improving the results of each type of model. This may confirm the fact that each indicator has a better value for models with textual variables.

The biggest difference in AUC value could be observed for the logistic regression model. In order to compere those two curves and check whether the difference between them was significant the code below was used:

```
PROC LOGISTIC DATA=results.all;
      MODEL isDefaulted(event='1')=lr_isDefaulted1 lrtm_isDefaulted1/nofit;
      ROC 'lr' lr_isDefaulted1;
      ROC 'lrtm' lrt_isDefaulted1;
      ROCCONTRAST reference('lrtm') / estimate e;
RUN;
```

In Figure 3 are graphically compared two ROC curves. The dashed red curve which is an effect of including results from textual responses, has a larger area under the curve compared to the solid blue curve which represents the model with numerical inputs alone.

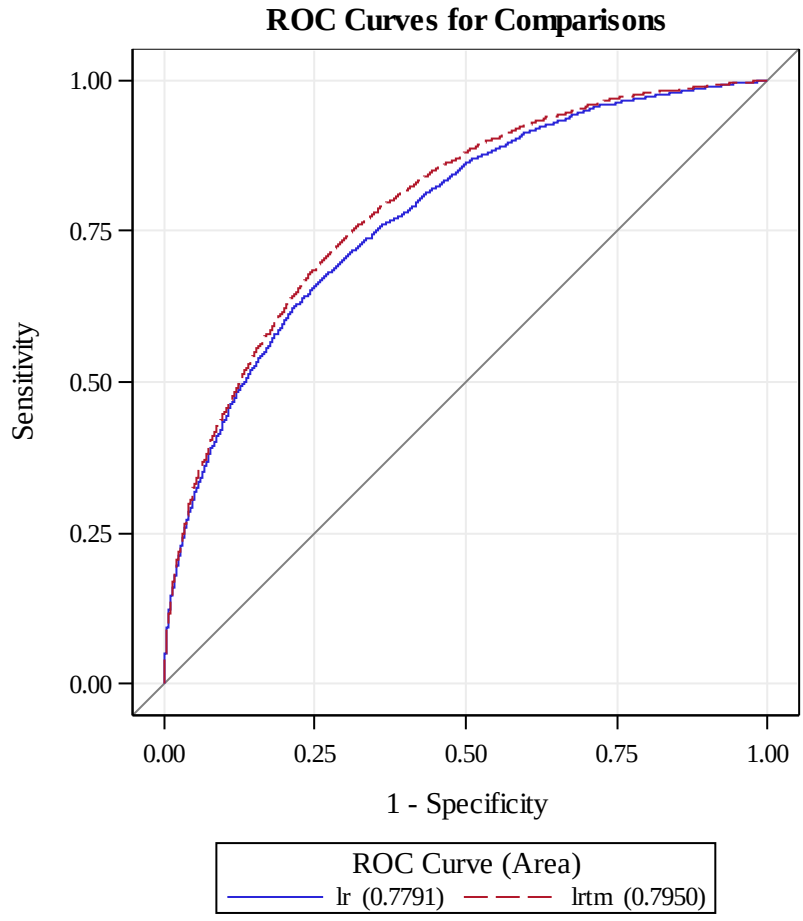


Figure 3. ROC Curves Comparison
Source: Output from Base SAS® version 9.4

PROC LOGISTIC procedure returns also Table 4, which allows you to test whether the difference between those two curves was significant.

ROC Contrast Estimation and Testing Results by Row						
Contrast	Estimate	Standard Error	95% Wald Confidence Limits		Chi-Square	Pr > ChiSq
lr - lrtm	-0.0158	0.00190	-0.0196	-0.0121	69.1614	<.0001

Table 4. Logistic Regression without (lr) and with (lrtm) text variables ROC Curves comparison Source: Output from Base SAS® version 9.4

To test it the following null and alternative hypothesis have been defined:

H_0: ROC curves of the models with and without text variables are not significantly different.

H_1: ROC curves of the models with and without text variables are significantly different.

Because p-value = 0.0001 is lower than $\alpha = 0.01$ then the null hypothesis can be rejected in favor of the alternative hypothesis. That means the logistic regression model with text variables is significantly better than the model with numerical inputs alone.

CONCLUSION

Including text variables into the input data set had a significant influence on improving the results of each of the three types of models. The most interesting insights and conclusions of the conducted text mining process are:

- Terms most correlated with defaulted loans are: child, partner, verification, treatment, son, please, phone, medicine, help, parent;
- Terms most correlated with loans repaid without problems: star, test, payday, loan, TV, vacation, December, system, sudden, fund;
- Generally, if the borrower in a loan description made reference to his family or help then it was more likely that it will default;
- The biggest difference in AUC value between models with and without text variables was observed for logistic regression.

Therefore, especially in the case of using logistic regression, this paper recommends adding text mining variables into input data, but at the same time keeping in mind the trade-off between complexity of model and more accurate results.

REFERENCES

Basel Committee on Banking Supervision 2005. "Working Paper No. 14 Studies on the Validation of Internal Rating Systems." *Bank for International Settlements.*

Berry, M. Kogan, J. 2010. *Text Mining – Applications and Theory.* John Wiley & Sons, Ltd

Bohme, R. Potzsch, S. 2010. *Privacy in Online Social Lending*. Technische Universitat Dresden International Computer Science Institutete,

Chakraborty G. Pagolu M. Garla S. 2013. *Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS.®* SAS Institute Inc., Cary, North Carolina, USA

Chen, D. Chaodong, H. 2012. "A Comparative Study of online P2P Lending in the USA and China." *Journal of Internet Banking and Commerce*, 17.2.

Engelmann, B. Rauhmeier, R. 2006. *The Basel II Risk Parameter, Estimation, Validation, Stress sting - with Applications to Loan Risk Management.* Springer - Verlag Berlin Heidelberg.

Galloway, I. 2009. “Peer-to-peer lending and community development finance.” *Community Investments,* 21.3 : 19-23.

Izzi, L. Oricchio, G. Vitale, L. 2012. *Basel III Credit Rating Systems. An applied Guide to Quantitative and Qualitative Models.* Palgrave Macmillan

SAS Institute Inc. (2012). *SAS® Text Miner 12.1 Reference Help.* Cary, NC: SAS Institute Inc.

Tkach, D. 1997. Text Mining Technology: Turning Information into Knowledge IBM White paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Piotr Małaszek
Warsaw University of Life Science
+48 697541077
malaszekpiotr@gmail.com