Spring/Summer Young Researchers' Colloquium

SYRCoSE Software Engineering

MAY 28-30, 2020 SAINT PETERSBURG, RUSSIA

Application of machine learning technology to analyse the probability of winning a tender for a project



May 28-30, 2020 Saint Petersburg, Russia



Nikita Kultin, docente, kultin_nb@spbstu.ru Danila Kultin, M. E., kultin.dn@edu.spb.stu.ru Speaker: Roman Bauer, B. E., bauer.rv@edu.spb.stu.ru

Introduction

Now



Excel-data

How it could be solved



Machine learning in tender competition analysis - Speaker: Roman Bauer

Eye-to-eye agreements Rough estimates Mood of stakeholders Other factors

Unformal factors





Resources' optimization







The purpose

Main purpose:

for tender projects development

Tasks:

- 1. Perform statistical analysis of data on the characteristics of projects

To justify the possibility of using machine learning technology to reduce costs

2. Based on the results of the analysis, select an adequate family of algorithms 3. Based on the quality criterion, select the best algorithm from the family 4. Using training and test data samples, determine the quality of the algorithm







Training data structure

Attributes of the project

Preliminary budget, type of work existing in the tender, preliminary project implementation period, place of implementation

Characteristics of the executing company

Project manager, sales manager, responsible manager

Characteristics of the company of tender owner

Client's office, the legal form of the client's company

Other attributes

Collaboration efficiency, application submission date



Statistical analysis

Correlation Features' distribution analysis analysis





Normalization

 $x - x_{min}$ x_{norm} $x_{max} - x_{min}$





Algorithm choosing

Logistic regression

 $\frac{1}{2}ww^{T} + C\sum_{j=1}^{n}\log\left(\exp\left(-y_{i}(x_{i}^{T}w+b)\right)+1\right)$

x - algorithm output y - true answer w - vector of weights **C** - total error penalty **b** - intercept **Included L2-regularization**

Support vector machine

$$\frac{1}{2}ww^{T} + C\sum_{i=1}^{n} \zeta_{i}, \text{ subject to}$$
$$y_{i}(w^{T}\phi(x_{i}) + b) \ge 1 - \zeta_{i}.$$

x - algorithm output y - true answers w - vector of weights C - total error penalty **b** - intercept ζ_i - error on each object



Quality assessment [1]

	Logistic regression method		Support vector method (polynomial loss function)			Number of vectors	
Metrics	Р	R	F	Р	R	F	
Value for class "1"	1.00	0.73	0.84	0.57	0.73	0.64	11
Value for class "0"	0.97	1.00	0.98	0.96	0.93	0.95	84
Average	0.97	0.97	0.97	0.92	0.91	0.91	95

Precision (P) – proportion of relevant instances among the retrieved instances Recall (R) – proportion of the total amount of relevant instances that were actually retrieved F-measure (F) – harmonic mean between precision and recall

Machine learning in tender competition analysis - Speaker: Roman Bauer

Table 1 Communication of algorithms and liter meeting





Quality assessent [2]

Logistic regression

Table 5. Cross validation results

Quality ratings	1	2	3	4	5	6	7
Logistic regression method	0.95	0.93	0.95	0.95	0.90	0.86	0.88
Support vector method	0.91	0.93	0.95	0.95	0.93	0.86	0.86



Testing on the real data

The algorithm correctly classified 40% of unpromising projects. It has not missed a single truly promising project.

Equivalent estimate of company's work quality without algorithm

Metrics	Р	R	F	Number of vecto
Value for class "1"	0.012	1.000	0.023	4
Value for class "0"	1.000	0	tends to 0	49

Estimate of company's work quality with algorithm

Metrics	Р	R	F	Number of vect
Value for class "1"	0.094	1.000	0.172	4
Value for class "0"	1.000	0.408	0.508	49



ors	





Conclusions

- 2. The classification algorithm can be constructed for use in any company current projects, reduce the cost of the pre-project stage.
- estimates, add probability estimates of belonging to the class.

1. Using machine learning technology to build a binary classification algorithm for projects is possible, but requires a significant amount of information about the characteristics of projects and the results of previous tenders.

engaged in project activities. Using the algorithm will allow you to rank

3. In the future, it is possible to refine the algorithm in the direction of tuning the hyperparameters of the algorithm to obtain more accurate and reasonable







Nikita Kultin, docente, kultin_nb@spbstu.ru Danila Kultin, M. E., kultin.dn@edu.spb.stu.ru Speaker: Roman Bauer, B. E., bauer.rv@edu.spb.stu.ru

May 28-30, 2020 Saint Petersburg, Russia

Thank you for attention!