2022

Bidders Recommender for Public Procurement Auctions

Using Machine Learning

IASMINA-OANA SILASCHI
RESEARCH PROJECT

Table of Contents

1. Classification

1.1. ACM:

J.1

1.2. MSC:

68T09

91-11

1.3. Key words:

machine learning

artificial intelligence

ranking

economic data

data processing

public procurement auctions

2. Abstract

This application will rank bidders considering the best fit for a public auction. Taking into consideration the persistent issues regarding Romania's integrity in the case of contracting companies for public procurement auctions, this application could validate if indeed the best decision has been made in auctions from the past, as well as be helpful regarding this matter in future tenders.

Even though this subject has already been researched, it has never been specifically tailored for tenders from Romania and using a dataset of companies from this country. Therefore, considering the specific legislation of this European country, it will have original results. Moreover, besides computing the profile of the ideal candidate for the contracting authority, my algorithm will also rank the bidding companies according to their compatibility.

3. Introduction

3.1. Public Procurement Auctions Bidding

3.2. Ranking using Machine Learning

4. Related Work

Article 1: (Casalegno, 2022)

Relevance:

This article is relevant because it focuses on how to implement Machine Learning models for Learning to Rank, explaining how this pertains in many applications. Also, the scoring model is thoroughly explained in an intuitive way.

Abstract:

In this post, by “ranking” we mean sorting documents by relevance to find contents of interest with respect to a query. Ranking models typically work by predicting a relevance score for each input. Once we have the relevance of each document, we can sort (i.e., rank) the documents according to those scores. The scoring model can be implemented using various approaches.

Structure:

- Ranking: What and Why?
- Ranking Evaluation Metrics
- Mean Average Precision (MAP)
- Discounted Cumulative Gain (DCG)
- Machine Learning Models for Learning to Rank
    - Pointwise Methods
    - Pairwise Methods
    - Listwise Methods
- Conclusions
- References

References:

The references were noted in the structure: <author>, <title>, <year>, each of them being links to the respective articles.

Article 2: (Chen, 2009)

Relevance:

I believe this article is relevant to my research project because it includes various mathematical functions and computations which are vital to the implementation of this subject.

Abstract:

Learning to rank has become an important research topic in machine learning. While most learning-to-rank methods learn the ranking functions by minimizing loss functions, it is the ranking measures that are used to evaluate the performance of the learned ranking functions. In this work, we reveal the relationship between ranking measures and loss functions in learning- to-rank methods, such as Ranking SVM, RankBoost, RankNet, and ListMLE. We show that the loss functions of these methods are upper bounds of the measure- based ranking errors. As a result, the minimization of these loss functions will lead to the maximization of the ranking measures.

Structure:

- Introduction
- Related work
  - Loss functions in learning to rank
  - Ranking measures
  - Previous bounds
- Main results
  - Essential loss: ranking as a sequence of classifications
  - Essential loss: upper bound of measure-based ranking errors
  - Essential loss: lower bound of loss functions
  - Summary
- Discussion
- Conclusion and future work
- References

References:

The references are very complex and different from one another, for example:

[1] R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. Addison Wesley, May 1999.

7] R. Herbrich, K. Obermayer, and T. Graepel. Large margin rank boundaries for ordinal re-gression. In Advances in Large Margin Classifiers, pages 115–132, Cambridge, MA, 1999.MIT

Article 3: (Liu)

Relevance:

This is a very comprehensive article which amasses plenty of useful insights for this subject, and it will serve as a great source to find information.

Abstract:

Learning to rank for Information Retrieval (IR) is a task to automatically construct a ranking model using training data, such that the model can sort new objects according to their degrees of relevance, preference, or importance. Many IR problems are by nature ranking problems, and many IR technologies can be potentially enhanced by using learning-to-rank techniques. The objective of this tutorial is to give an introduction to this research direction. Specifically, the existing learning-to-rank algorithms are reviewed and categorized into three approaches: the pointwise, pairwise, and listwise approaches.

Structure:

- 1 Introduction 226
  - Ranking in IR 228
  - Learning to Rank 235

References:

Most of the references in this article are as the following example, <author> <title> <journal> <volume> <page> <year>:

[111] S. E. Robertson, “Overview of the okapi projects,” Journal of Documentation, vol. 53, pp. 3–7, 1997.

Article 4: (Machine learning (ML) applications: ranking, 2022)

Relevance:

This article includes suggestive drawings to help better understand the topic, and it is short and concise, so I believe it is a great starting point to dive into the researched subject.

Abstract:

Ranking is a type of machine learning that sorts data in a relevant order. Companies use ranking to optimize search and recommendations.

Structure:

- What is a ranking model?
- How does ranking work?
- Why should I care?
- Use cases
- The fastest way to build a ranking model

References:

In this online article, the references are integrated in the article by having the referenced parts as links to the respective articles or sources.

Article 5: (S. Pandey, 2019)

Relevance:

The techniques that can be used are explained very well, as well as compared to one another, thus being a very interesting and helpful article.

Abstract:

Information retrieval is the topic area in which much research has been done and is still going on. The rapidly growing web pages make it very crucial to search up to date documents. In continuation of research works on learning to rank, this research focuses on implication of machine learning techniques for IR ranking.

Structure:

- Introduction
- Related Works
- Objectives of Current Research
- Proposed Methodology
- Experimental Results

References:

The references are numbered and contain the author, title and some further information that depends on the nature of the reference, and you can also view them in the context of the paper, as they each have this option. For instance:

Andrei. Broder, "A taxonomy of web search", ACM Sigir forum., vol. 36, no. 2, 2002.

Show in Context

5. Literature Review

6. Material and Methods

6.1. Dataset

Table 1: **Bids**

noticeNo
contractValue
title
currencyCode
publicationDate
countyCode
contractingAuthorityName
cpvCodeName
cpvCode
cpvCodeType
cNoticeNo
cNoticeEstimatedContractValue
cNoticePublicationDate
cNoticeTitle
organizationName
organizationId

Table 2: **Contracting Authorities**

contractingAuthorityId
name
CUI

Table 3: **Oganizations**

organizationId
name
CUI

6.2. Method

<u>Forecasting phase:</u>

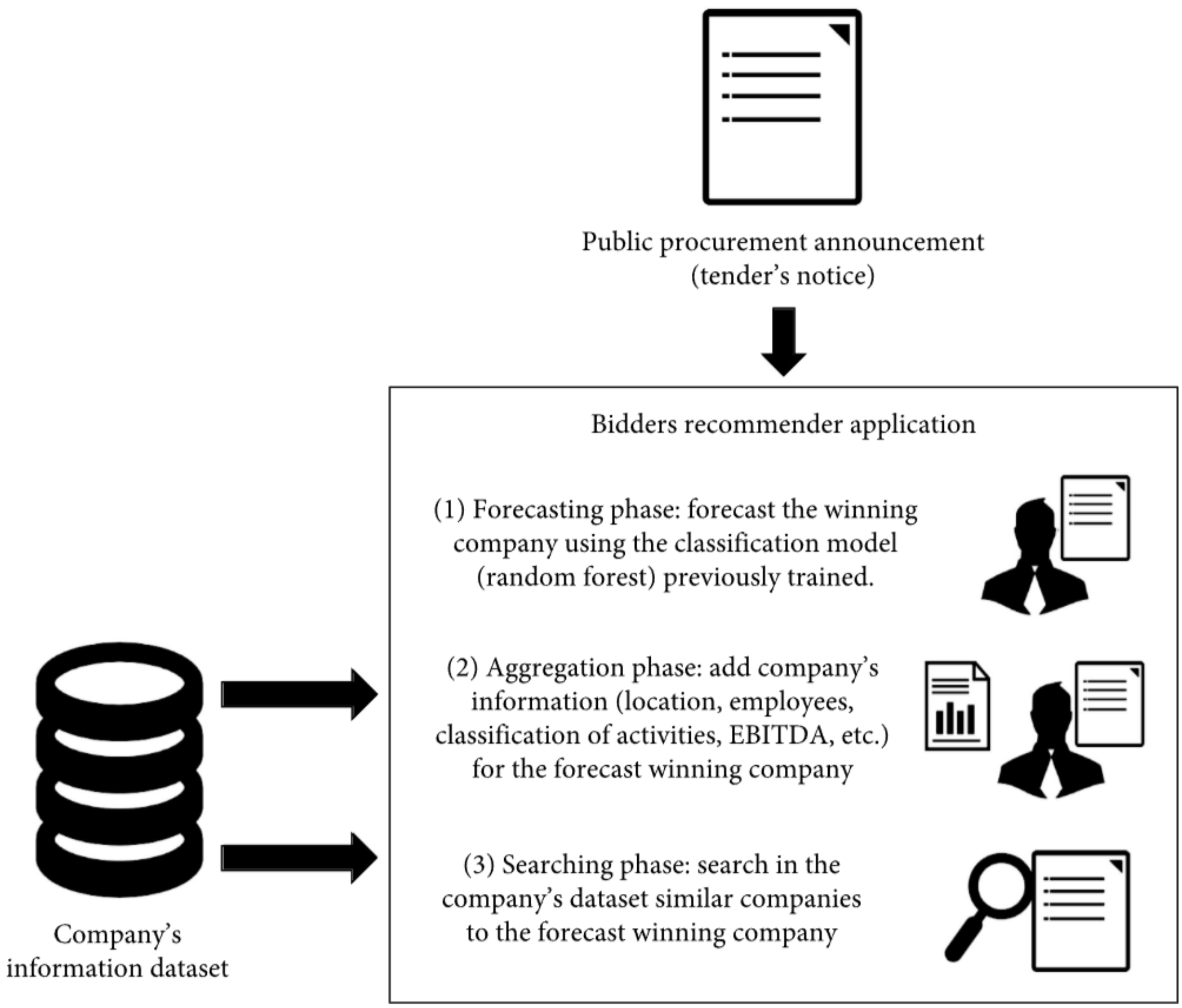
Forecast the winning company using a previously trained AI.

<u>Aggregation phase:</u>

Add the company's information for the forecast winning company.

<u>Searching phase:</u>

Search in the dataset similar companies to the forecasted winning company and rank them according to their compatibility.



6.3. Algorithm

Random forest (RF), introduced by Breiman in 2001, is an ensemble learning method for classification or regression that operates by constructing a multitude of decision trees at training times and outputting the class, which is the mode of the classes (classification) or mean prediction (regression) of the individual trees. It is a popular learning algorithm that offers excellent performance, no overfitting, a versatility of applicability to large-scale problems and in handling different types of data. Particularly, Random Forest has been applied with remarkable success in tender datasets. It provides its own internal generalization error estimate, called the out-of-bag (OOB) error. Simplified algorithm of RF for classification summarized:

(1) For $b = 1$ to B (number of trees):

(a) Draw a bootstrap sample $\mathbb{Z}^*$ of size N from the training data.

(b) Grow a random forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size $n_{\min}$ is reached.

(i) Select m variables at random from the p variables.

(ii) Pick the best variable/split point among the m.

(iii) Split the node into two daughter nodes.

(2) Output the ensemble of trees $\{T_b\}_1^B$.

To make a prediction at a new point x, let $\widehat{C}_b(x)$ be the class prediction of the $b-$ th random forest tree. Then, $\widehat{C}_{rf}^B(x)$ = majority vote $\{\widehat{C}_b(x)\}_1^B$.

7. Experiments

A real case study:
I will take a tenders list from Romania, a list of the organizations as well as a list of contracting authorities in order to perform experiments.

The sample consists of 47,974 contracting authorities, 289,472 companies and 42,474 bids.

7.1. Data Processing

Data preprocessing of the tender dataset is necessary due to the fact that information has not been verified automatically to correct human errors, such as incorrect formatting, wrong values, empty fields, and so on. Data preprocessing can be divided into the following 5 consecutive tasks: extraction, reduction, cleaning, transformation, and filtering.

At first, there were 289,472 companies. After data preprocessing, there were 230,770 tenders. An algorithm to check the integrity of the VAT was used to filter the companies.

7.2. Statistical Analysis

Firstly, the most relevant information of the tender dataset will be explained, quantitatively. Secondly, the company dataset will also be explained, as well as the contracting authorities dataset, and, finally, the correlations between all three datasets will be analyzed.

The tender dataset the quantitative description of the tender dataset: total numbers, means, medians, maximum, percentages, etc.

8. Conclusion

9. Bibliography

1. A rank-and-compare algorithm to detect abnormally low bids in procurement auctions. (2012). In L. D. Pier Luigi Conti, *Electronic Commerce Research and Applications* (pp. 192-203).
2. Application of artificial intelligence in control systems of economic activity. (2019). In O. Melnychenko.
3. Casalegno, F. (2022). Learning to Rank: A Complete Guide to Ranking using Machine Learning. *Towards Data Science*.
4. Chen, W. &.-Y. (2009). Ranking Measures and Loss Functions in Learning to Rank. *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7-10 December 2009*, (pp. 315-323). Vancouver, British Columbia, Canada.
5. Dragan Pamučar, D. B. (2022). Application of neuro-fuzzy system for predicting the success of a company in public procurement. In *Decision making: Applications in Management and Engineering* (pp. 135-153).
6. Laura Abrardi, C. C. (2021, July 26). *Artificial intelligence, firms and consumer behavior: A survey.* Retrieved from Wiley Online Library: https://onlinelibrary.wiley.com/doi/full/10.1111/joes.12455
7. Leon Yang Chu, H. N. (2020). Position Ranking and Auctions for Online Marketplaces.
8. Lisa Chever, S. S.-B. (2016). The law of small numbers: investigating the benefits of restricted auctions for public procurement.
9. Liu, T.-Y. (n.d.). *Learning to Rank for Information Retrieval.* Retrieved from http://didawiki.di.unipi.it/lib/exe/fetch.php/magistraleinformatica/ir/ir13/1_-_learning_to_rank.pdf
10. *Machine learning (ML) applications: ranking.* (2022, Mar 17). Retrieved from DEV: https://dev.to/mage_ai/machine-learning-ml-applications-ranking-238d
11. Manuel J. García Rodríguez, V. R. (2020, November 25). *Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain.* Retrieved from Hindawi: https://www.hindawi.com/journals/complexity/2020/8858258/
12. Manuel J.García Rodrígueza, V. R.-M.-P. (2022). Collusion detection in public procurement auctions with machine learning algorithms. In *Automation in Construction.*
13. Milutin ŽIVKOVIĆ, ,. M. (n.d.). MULTICRITERIA APPROACH TO PUBLIC PROCUREMENT FOR. In *Fiability & Durability* (pp. 181-187).
14. Montequín, V. R. (2019). Public Procurement Announcements in Spain: Regulations, Data Analysis, and Award Price Estimator Using Machine Learning.
15. S. Pandey, I. M. (2019). Information Retrieval Ranking Using Machine Learning Technique. *Amity International Conference on Artificial Intelligence (AICAI)*, (pp. 86-92).