AJSIHD
GLOBAL RESEARCH NETWORK

AMERICAN Journal of Science on Integration and Human Development
Volume 01, Issue 08, 2023 ISSN (E): 2993-2750

Qualitative Issues of the Final Result of Corruption Risk Prediction

Tukhtasinov Nurillo Ismailjon ugli
Senior teacher at Tashkent State Law University

Abstract:

The work contains an analysis of various approaches to classification, and methods of preliminary data processing for calculating corruption risks by region, carrying out analysis.

Keywords: government procurement, measurement of corruption, risk markers, data classification.

An analysis of world practice indicates that the role of public procurement plays an important role in regulating the market, supporting economic social stability, and ensuring economic growth.

Government procurement in Uzbekistan is “a serious pain point in terms of corruption.” There are problems “starting with the design of facilities, construction, repairs, technical support for ministries, departments and budgetary institutions.” “There are a lot of simulation procedures, when approaches have already been agreed upon in advance. A lot of inflated prices, even despite the procedures being followed. That is, the results are also imitated, and inflated prices are often the result. [6]

There are various methods for measuring corruption. The use of big data processing technologies using risk markers forms objective indices of objective corruption during processing and, based on analysis with various indicators, allows one to judge the degree of influence of corruption on the overall corruption situation in the country[7].

Not long ago, indices measuring corruption acquired political and informational significance as a democratic tool of governance, provision of public goods, and economic growth. Understanding the importance of measurement, international organizations regularly monitor the extent of corruption in their countries and even link funding opportunities to various indicators of corruption. In particular, corruption is defined in the field of public procurement in the allocation and execution of public contracts in violation of previously clear rules and principles of open procurement procedures by using a closed network, denying access to all other bidders. Understanding clientelistic regimes as a subtype of limited access procedure determines the type and prevalence of corruption, creating a form of corruption that favors certain suppliers, creating a situation where one supplier is chosen. [8]

In the absence of reliable objective measures, there are currently three main sources of corruption indicators: 1) surveys of corruption perceptions and attitudes (which are the most widely used); 2) reviews of legal and regulatory frameworks; 3) detailed analysis and audit of individual cases. Unfortunately, each of them has serious shortcomings that leave us without any measure of corruption that is sufficiently reliable and valid to compare countries over time or examine them within countries. To fill some of the gap between the demand for corruption

indices and the state of currently available data, the current methodology for studying corruption indicators allows us to rely on the following factors:

1. Which rely exclusively on objective indicators that describe the behavior of factors.

2. Define procedures at the micro level as individual operations.

3. Allows consistent temporal comparisons within and between countries.

4. is based on a deep understanding of the process of extracting corrupt bribes.

In the context of public procurement, corruption refers to the allocation and execution of contracts through pre-arranged rules and system principles using a closed network to restrict suppliers' access to tender bids[8,9]. This work will address the problem of insufficient knowledge of approaches and methods for classification and preliminary analysis of risk marker data, assessment of corruption risks by region. The impact of corruption is caused by many factors, which does not leave its mark. Also, the risks of corruption have an impact on various components in the state.

In this work, we will consider the subjects of the Uzbek Republic with high and low ratings of socio-economic status. I would like to note that the economic development of regions is a system of measures aimed at implementing long-term tasks of the socio-economic development of the state, taking into account the rational contribution of the regions to solving these problems, determined by the real prerequisites and limitations of their development. The development of regions varies depending on the socio-economic and political orientation of the state at a particular stage of development. The state's regional development strategy is heterogeneous in relation to the regions that form it. This is due to significant differences between regions in the field of resource provision, economic structure, and the achieved level of development of various sectors of the economy.

In recent years, the independence of regions has increased, and they bear increasing responsibility for the results of regional economic development. The socio-economic state of regions is determined by both objective (macroeconomic conditions, the region’s position in the social division of labor, sectoral structure, geographical location, natural resources) and subjective factors, and primarily by methods of regional management.

In recent years, economic reforms have shown that regions that apply progressive methods of managing their development are less susceptible to crisis trends.

Data mining, a process that includes both the preliminary steps of data processing and the operations of visualizing the results and interpreting the conclusions. There are several stages that make up the complete data mining process:

1. Analysis of the subject area and statement of the problem, including definitions of the research topic, consideration of all ways of using data for analysis.

2. Data extraction and pre-processing Since we were provided with a ready-made set for research, our task is to determine the nature of this data, the relationship with what to understand and in what format this data should be extracted and aggregated.

3. Pre-processing of data, a process that includes converting data into a format convenient for us to work with them, this is usually the removal of contradictory or erroneous data, outliers and interpolation of missing values, and also includes the analysis of dependencies in input parameters that can unexpectedly influence the result of the model.

4. Construction of data mining models, at this stage, meaningful analysis of the obtained data, solving problems of information content of features, classification, typically used methods and algorithms of “supervised learning”. The number of methods used is large, so our task is to study various methods for working with data and select the appropriate analysis for our subject area, the complexity of the area under study and the dimension of the feature space.

5. Interpretation of the results obtained, including visualization of the information received in the form of visual information for the process of understanding. The choice of method for visualizing and presenting results largely depends on the initial task and the specifics of further application of the results obtained.

6. Using the results obtained, this stage is the most important and resource-consuming, depending on the complexity and volume of the data[12].

The quality of data plays a significant role in the study, so any data goes through this process for ease of working with it. Data contamination occurs for various reasons, mostly due to human error.

The main types of errors are inconsistency of information, omissions, anomalous values, etc. inconsistency of information and abnormality of meaning are closely related to the subject area. Such anomalies and outliers can be assessed by constructing standard deviation curves and preliminary cluster analysis, which can identify obvious outliers, if, of course, clustering on the data is possible.

At the data cleaning stage, it is also important to minimize records with missing values that cannot be filled based on associated features.

A feature description is a set of characteristics. Classification can be binary when we classify into two classes. These are the most common cases. However, there are problems with overlapping classes, when one object can belong to several classes at the same time, or unclear classification, when the probability of belonging to each of the classes is assessed. The data contains statistical data on the processes of public procurement and levels of development of human development regions in the constituent entities of the Uzbek Republic.

Innovative approach - presents several classifiers, the result of which is combined in some way to obtain the final result. A frequently used method is to combine the decisions of individual classifiers included in the ensemble and select. The main reason that arouses interest in building ensembles is that almost always a set of classifiers has a higher classification quality, which significantly exceeds the quality of individual classifiers and is robust (stable) in relation to the "noise" of the training sample.

A necessary condition for the high quality of an ensemble of classifiers is that the classifiers that make it up have an accuracy better than random and are significantly different (diversified), i.e. made mistakes in various precedents[14].

Groups of methods for constructing classifiers:

1. Manipulating examples of the training sample;

2. Manipulation of signs;

3. Injection of randomness into the inductive algorithm;

4. Manipulation of class labels;

5. Bayesian voting.

The first group of methods consists of either using a basic inductive algorithm on different subsamples of the original training set, or iteratively reweighting observations. The simplest method for forming subsamples was proposed by Breiman [15].

The method is based on the formation of a training sample for each ensemble classifier using bootstrap, i.e. a random sample (the same size as the original training sample) with a return from the original training sample and using the voting method to aggregate the decisions of individual classifiers.

The method is called bagging or bootstrap aggregating. A method that uses iterative reweighting of observations of the training sample – boosting – was proposed in [16]. The idea of boosting is

that the ensemble classifiers are built sequentially and at each iteration the observations of the training sample are corrected (reweighted) (at the first iteration the weights of all observations are equal).

The correction is carried out in such a way that the corresponding the classifier made fewer errors on those observations on which Classifiers built on previous iterations of the algorithm made mistakes. In addition, each classifier is assigned a certain weight based on the number of errors it makes.

- Random forest
- Gradient boosting;

The study of corruption risks in public procurement is characterized by a large volume of structured and redundant data. Processing such data is labor-intensive work that requires a modern approach at all stages of data analysis.

In this final qualifying work, the problem of applicability of classification algorithms to data was solved in order to capture the relationship between developed regions and poor ones, according to the rating of socio-economic status.

The main results of the work are:

- calculation of the relationship between corrupt applications to the total number of applications in the region, the ratio of corrupt contracts to the regional development rating was confirmed
- calculation of correlation to establish a connection between the ratio of corrupt applications in the region and the total number.

A comprehensive study was carried out on real data, classification algorithms were implemented to predict risks using three algorithms.

The quality of the classification was experimentally obtained; based on the characteristics, it is possible to predict the risks of corruption in the regions.

List of used literature:

1. Левин М.И. лекции по экономике коррупции / М.И. Левин, Е.А. Левина, Е.В. Покатович. – М.: Издательский Дом НИУ ВШЭ, 2011.-360 с.
2. Биктеев Р.Р. Влияние коррупции на благосостояние государства и общества // Молодежный научный форум: Общественные и экономические науки: Электр. сб. ст. по мат. IV междунар.студ. науч.-практ. конф. № 4(4).URL: https://nauchforum.ru/archive/MNF_social/ 4(4).pdf (дата обращения:20.04.2019)
3. Попова О.А. Государственные закупки как эффективный инструмент регулирования развития экономики региона// Экономика предприятий, регионов и отраслей. –2014. – С.61-63.
4. Котляров С. Б. Индикаторы измерения коррупции: понятие и совершенствование / С. Б. Котляров, Е. А. Чичеров // Актуальные проблемы права: материалы IV Междунар. науч. конф. – М.: Буки-Веди, 2015. – С. 6-9.
5. Фадеев А.В. Антикоррупционное законодательство России и зарубежных стран: криминологическое исследование [Текст]: автореферат дис....канд.юрид. наук: 12.00.08 / А.В. Фадеев. – М., 2012. – 20 с.