PROSPECTS FOR DATA MINING APPLICATIONS IN ASSESSING INVESTMENT ATTRACTIVENESS OF THE REGION

E. M. SMIRNOVA
Ivanovo State University, Institute of Mathematics, Information Technologies and Natural Sciences
1st year Master in Applied Computer Science

Introduction

Existing **approaches to assessing investment attractiveness** do not give a complete picture of the study, confining itself to an expert opinion on the investment attractiveness of the region based on the *assessment of investments* made. In our opinion, **the most preferable approach** is to study the various areas of public life of the territory – *political*, *economic*, *environmental*, *industrial innovation* – for the purpose of assessing the attractiveness of investment. This involves *collecting data in all areas of public life in the region*, which are important for the *effectiveness of the investment project* in the region.

The **aim of the work** is to research **Data Mining technology** and use **Data Mining methods** to assess the region's investment attractiveness.

The **objectives of the study** are:

- to *consider* the concept and *define* data mining technology;
- to *identify* the types, methods and stages of Data Mining;
- to *list* tasks and *identify* areas of application of Data Mining technology;
- to *consider* the sources and features of the information for data mining purposes;
- to *explore* the modern uses of Data Mining technology in the field of analysis of the political, economic, environmental, industrial innovation and labor state of the territory.

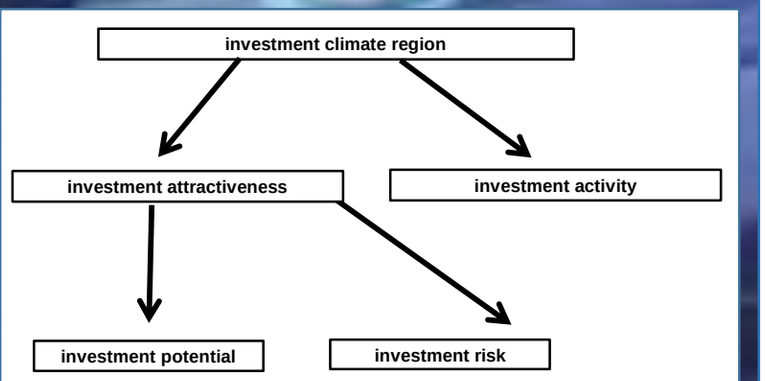


Figure 1. Components of the Investment Climate Region

Methods and Materials

Data mining methods and tools are used to study the state of the region from the point of view of investment attractiveness.

Data Mining methods and algorithms **include**: *artificial neural networks*, *decision trees*, *symbolic rules*, *methods of the nearest neighbor and k-nearest neighbor*, *method of support vectors*, *Bayesian networks*, *linear regression*, *correlation and regression analysis*, *hierarchical methods of cluster analysis*, non-hierarchical methods of cluster analysis, including *k-medium and k-median algorithms*; *methods for finding associative rules*, including the *Apriori algorithm*; *limited-overs*, *evolutionary programming and genetic algorithms*, a variety of *data visualization techniques* and many other methods.

MS Excel software, the *Deductor software package*, and SPSS software platform are used as an **instrumental base** for developing a model of investment attractiveness of the regions.

Results and Discussion

The **scientific novelty** of the study lies in the *systematization* of information and the *search for non-trivial relationships* in the analysis of the *region's investment attractiveness* through analytical programs designed to *classify*, *cluster* and *predict data*.

The **results** can be used to *assess the investment climate of the region*, *compare it with other regions* and *build models for further development of investment attractiveness* both of a single region, and the state as a whole (*Figure 1*). The indicators under study to establish the investment attractiveness of a particular territory include **statistics** on certain aspects of the socio-economic and political state of the territory (*Figure 2*).

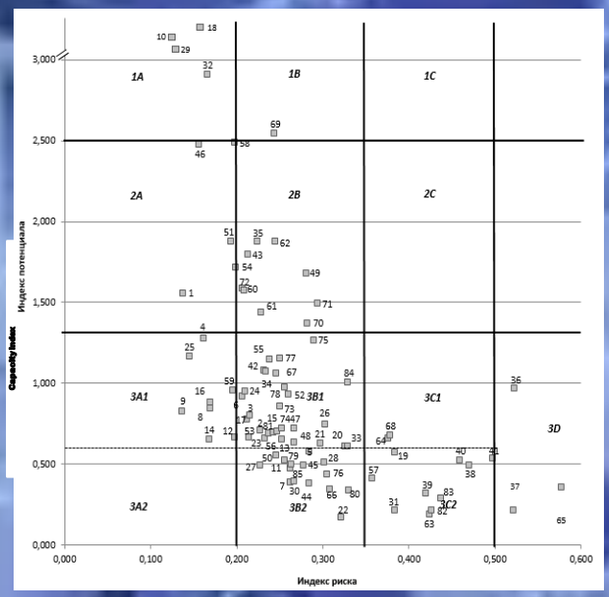
Data analysis includes the following **steps**:

Stage 1. Data series were collected for each of the 56 indicators used for all 85 regions of Russia covered by the study.

Stage 2. Using expert scales, aggregated estimates of all factors of investment attractiveness were calculated, as well as an integral index of investment attractiveness for each region.

Stage 3. On the basis of cluster analysis and allocation of index thresholds, regions were divided into three enlarged categories and nine investment attractiveness groups.

Figure 2. Investment rating of Russian regions 2020



Risk index

1A - maximum potential - minimum risk
1B - high potential - moderate risk
1C - high potential - high risk
2A - medium potential - minimum risk
2B - medium potential - moderate risk
2C - medium potential - high risk
3A1 - downgraded potential - minimum risk
3A2 - low potential - minimum risk
3B1 - reduced potential - moderate risk
3C1 - reduced potential - high risk
3B2 - low potential - moderate risk
3C2 - low potential - high risk

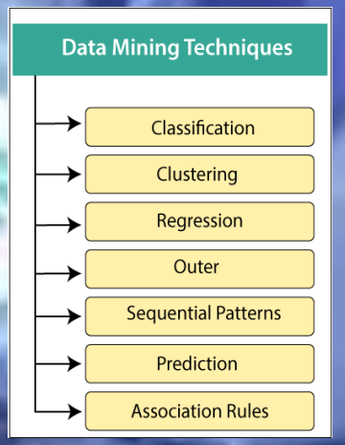


Figure 3. Data mining techniques

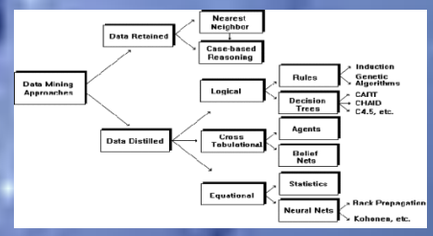


Figure 5. Data mining Approaches

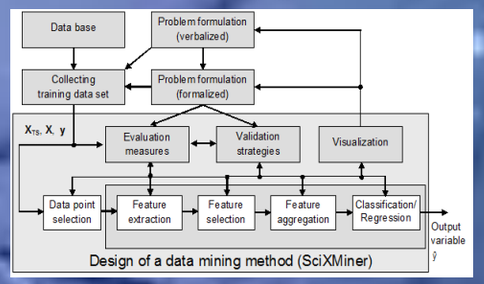


Figure 4. Data mining method

Conclusion

Further research is aimed at *developing a mathematical apparatus* to assess the investment attractiveness of the region using *formalized and informalized assessment criteria* and develop a *methodology for assessing the investment attractiveness of the region* using **cluster analysis tools**.