Burgas State University "Prof. Dr. Assen Zlatarov"

ABSTRACT

ON THE DISSERTATION FOR AWARDING THE EDUCATIONAL AND SCIENTIFIC DEGREE "DOCTOR"

BY **Petar Rosenov Petrov**

TOPIC: **Modeling of Processes for Big Data Analysis through Generalized Nets**

Scientific field:	5 Engineering sciences
Professional field:	5.3 Communication and computer technology

SCIENTIFIC ADVISORS:

1. Acad. Prof. Dr. Krasimir T. Atanasov
2. Assoc. Prof. Dr. Vesselina K. Bureva

BURGAS, 2025

The dissertation was discussed and approved for defense at an extended meeting of the Department of Computer Systems and Technologies held on June 26, 2025, at Burgas State University "Prof. Dr. Assen Zlatarov" - Burgas.

The dissertation consists of 148 pages, including 67 figures and 4 tables. A total of 138 references are cited. The results have been published in 10 articles.

The defense of the dissertation will take place on .... 2025 at .... in room .... at Burgas State University "Prof. Dr. Asen Zlatarov" - Burgas before a scientific jury composed of:

1. Prof. Sotir Nikolov Sotirov, PhD
2. Assoc. Prof. Deyan Georgiev Mavrov, PhD
3. Prof. Maria Mihaylova Nisheva-Pavlova
4. Prof. Daniela Ananieva Orozova, PhD
5. Corresponding Member Prof. Stefan Todorov Hadjitodorov, DSc

Reserve members: 1. Assoc. Prof. Todor Petkov Petkov
2. Assoc. Prof. Nora Angelova Angelova

The defense materials have been made available to interested parties in office …………………………………...

Author: Petar Rosenov Petrov
Title: Modeling of Processes for Big Data Analysis through Generalized Nets

Acknowledgements

I would like to express my deep gratitude to the supervisors of my dissertation, Assoc. Prof. Veselina Bureva, PhD, and Acad. Krasimir Atanasov, DSc, for their knowledge, assistance, opportunities, and inspiration.

I would also like to thank Assoc. Prof. Velichka Traneva, PhD, Assoc. Prof. Stoyan Tranev, who contributed with opportunities for additional research and practical software implementations in the field of indexed matrices and intuitionistic fuzzy sets through our joint participation in NIH-440/2020 "Indexed matrices as a tool for data mining."

I would also like to thank the head of the Intelligent Systems Laboratory at Prof. Dr. Assen Zlatarov University, Prof. Sotir Sotirov, for his overall support during my studies, first as a master's student and later as a doctoral student.

I would also like to thank all my colleagues from the Department of Computer Systems and Technologies at Prof. Dr. Asen Zlatarov University for their support and cooperation.

Part of the research in the dissertation and publications were supported by Project BG05M20P001-1.002-0011 "Establishment and Development of a Competence Center for Mechatronics and Clean Technologies MIRACle (Mechatronics, Innovation, Robotics, Automation, Clean technologies)".

Introduction

In modern everyday life, people are surrounded not just by an "ocean of data," but by an entire "universe of data"! Data is everywhere around us and accompanies us in all areas of human activity. Today, data is key to the development not only of artificial intelligence, but also of all areas – for example, healthcare, education, economics, transportation, agriculture, construction, and many others. It would not be an exaggeration to say that data is the building block of a smart and sustainable future! Data has always been a valuable asset, especially when it comes to making informed and reasoned decisions, creating strategies or plans. In the past, people had data at their disposal, but not in the unprecedented quantities we have today, nor with the ease of storage and processing that we enjoy now.

The growth in the amount of data, the possibilities for its storage and processing, as well as the awareness of its meaning and usefulness, led to the formation of data science and the concept of big data. However, big data also brings major challenges related to its optimal storage and processing, as well as the extraction of knowledge from it. Traditional models often fail to reflect the complexity, uncertainty, and dynamics inherent in real-world processes. This requires the application of more adaptive and powerful theoretical frameworks that allow not only formalization but also effective processing of unstructured, uncertain, incomplete, or contradictory information. This dissertation examines the Generalized Nets proposed by K. Atanasov, which are a means of modeling processes that run parallel in time. They include, as special cases, Petri nets and all their modifications. In this sense, they are a natural basis for describing complex systems and algorithms. In combination with intuitionistic fuzzy sets (IFS) and intuitionistic fuzzy evaluations (IFE), they represent a good theoretical basis that is both mathematically correct and intuitive for human logic. Indexed matrices are also considered, as well as some extensions of intuitionistic fuzzy sets.

This paper aims to explore the possibilities for modeling big data analysis processes through generalized networks, using intuitionistic fuzzy sets and indexed matrices. By integrating these theoretical approaches with applied algorithms for cluster analysis, classification, and optimization, a multi-layered framework for intelligent data analysis under conditions of uncertainty is proposed.

Approbation of the results

The results were validated through presentations at several international conferences and in articles published in scientific journals.

Contents of the dissertation

The dissertation is 148 pages long and consists of an introduction, three chapters, contributions, a list of publications related to the dissertation, and a bibliography. The dissertation includes 67 figures and 4 tables, and the bibliography contains 138 titles.

Chapter 1. Theoretical concepts of big data, generalized nets, intuitionistic fuzzy sets and indexed matrices

The first chapter provides the basic definitions of generalized networks, intuitionistic fuzzy sets, and indexed matrices. It also provides an overview of the field of knowledge extraction from data and big data.

1.1. Data mining, Data Science and Big data

This section presents basic concepts and techniques related to knowledge extraction from data and big data.

1.2. Generalized nets

Here are some basic definitions that are necessary for the rest of the presentation and brief historical notes on the theory of generalized networks.

1.3. Intuitionistic fuzzy sets

Here we give basic definitions that are necessary for the exposition below and brief historical notes on the theory of intuitionistic fuzzy sets. Intuitionistic fuzzy pairs, operations on intuitionistic fuzzy sets, and a geometric interpretation of intuitionistic fuzzy sets are presented.

1.4. Index matrices and some extensions of the intuitionistic fuzzy sets

Here are the basic definitions of indexed matrices. Extensions of intuitionistic fuzzy sets, such as circular intuitionistic fuzzy sets and elliptical intuitionistic fuzzy sets, are also presented. Applications of the extensions in circular intuitionistic fuzzy matrices and elliptical intuitionistic fuzzy matrices are discussed.

Aim and tasks of the dissertation

The aim of the dissertation is to study various processes related to big data analysis (Big Data, Data Mining, Data Science) by modeling them using generalized networks and their software implementation, as well as developing intuitionistic fuzzy extensions of optimization tasks and their implementations. To achieve this goal, the following tasks have been set:

1. Analyze the areas of big data analysis, generalized networks, intuitionistic fuzzy sets, and indexed matrices;
2. Development of the generalized nets models of DBSCAN and BIRCH clustering algorithms using intuitionistic fuzzy evaluations;
3. Development of an application with intuitive fuzzy evaluations in various fields such as waste sorting, database analysis, and scientific term recognition;
4. Development of an algorithm with intuitionistic fuzzy evaluations for data analysis and recommendations for student academic performance, and application of intuitionistic fuzzy evaluations for data analysis from state matriculation exams in the secondary education system in Bulgaria.;

5. Development of software implementations of algorithms for solving the intuitionistic fuzzy knapsack problem based on IFP, circular and elliptical intuitionistic fuzzy sets;

The text marked with [n*] indicates the author's articles included in the dissertation.

Chapter 2. Applications of generalized nets and intuitionistic fuzzy evaluations in the area of big data analysis and artificial intelligence

2.1. Generalized nets in big data analysis

This section examines the role of big data analysis and knowledge extraction as key areas in contemporary science and industry. It presents the possibility of using generalized networks and intuitionistic fuzzy evaluations in the modeling of dynamic, parallel, and uncertain processes characteristic of big data analysis.

2.1.1. Generalized network model of the algorithm Density-based spatial clustering of applications with noise (DBSCAN) and its applications on a set of data related to diabetes

In this dissertation a generalized nets model of DBSCAN with intuitionistic fuzzy evaluations is presented (Fig. 1). The GNDraw software is used for GN model construction [54]. The GN model contains the following set of transitions:

$$A = \{Z_1, Z_2, Z_3, Z_4, Z_5\},$$

where the transitions describe the processes:

- Z_1 – Data sources;
- Z_2 – Preprocessing the received dataset;
- Z_3 – Scanning data for selection initial core point/s;
- Z_4 – Forming clusters with minimum points;
- Z_5 – Intuitionistic fuzzy evaluation of the received clusters.

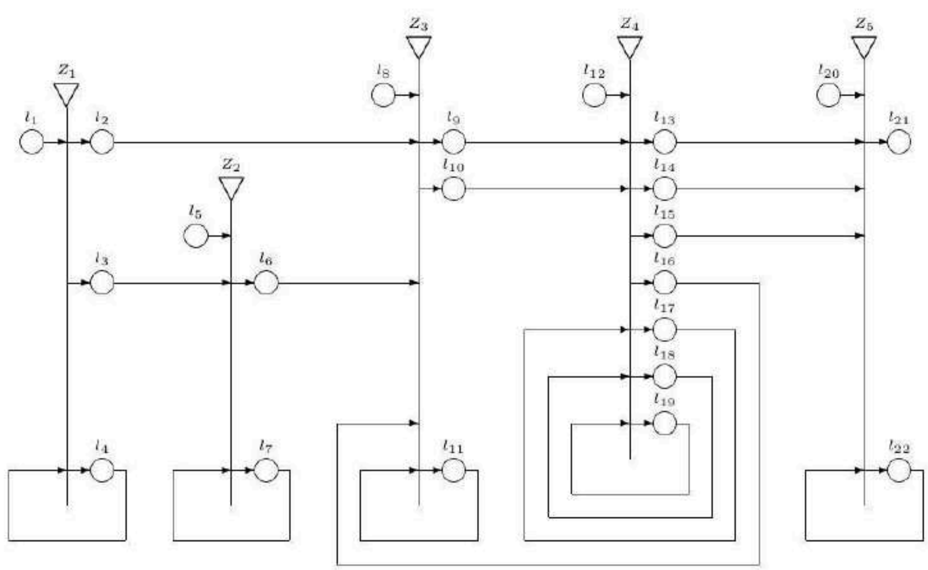

Fig. 1. *Generalized net model of the process of cluster analysis using DBSCAN algorithm with intuitionistic fuzzy evaluations*

2.1.2. Generalized Net Model of Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) with Intuitionistic Fuzzy Evaluations

In this dissertation the GN model of BIRCH with intuitionistic fuzzy evaluations is presented (Fig. 2). The GNDraw software is used for GN model construction [54]. The GN model contains the following set of transitions:

$$A = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7\},$$

where the transitions describe the processes:

- Z_1 – Data sources;
- Z_2 – Preprocessing the selected data;
- Z_3 – Scanning data into memory;
- Z_4 – Condense data;
- Z_5 – Global clustering;
- Z_6 – Refining clusters;
- Z_7 – Intuitionistic fuzzy evaluations.

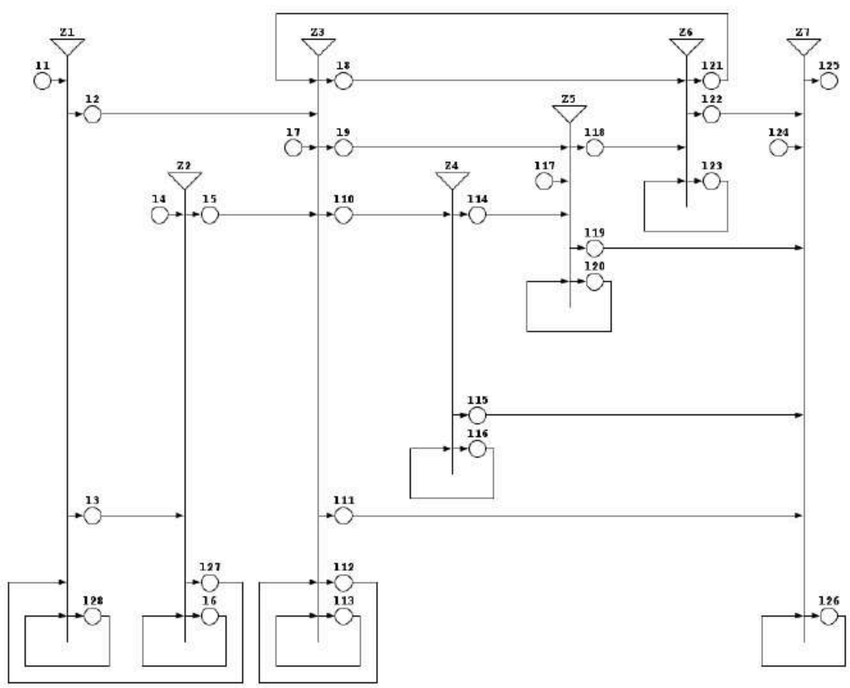

Fig. 2. *Generalized net model of the process of cluster analysis using BIRCH algorithm using intuitionistic fuzzy evaluations*

2.2. Intuitionistic fuzzy evaluations in big data analysis, artificial intelligent and data-driven decision making

This section presents intuitionistic fuzzy evaluations as an effective tool for dealing with uncertainty, incompleteness, and inconsistency of information. The focus is on their application in big data analysis, artificial intelligence, and decision-making, where IFEs provide flexibility and more accurate modeling of processes in conditions of uncertainty.

2.2.1. Application of Intuitionistic Fuzzy Evaluations of Garbage Sorting using a Robotic Arm and possibilities for data mining in the process of garbage collecting and recycling

Modern technologies and intelligent systems certainly have a place in the study of waste sorting and recycling. This section presents the implementation of a waste sorting algorithm using a robotic arm and intuitive fuzzy evaluations. The implementation was carried out using two DOBOT Magician robots and a mini conveyor belt. The idea is to evaluate the waste sorting procedure using intuitionistic fuzzy evaluations. The first DOBOT Magician robot picks up a piece of waste using a suction pump and places it on the mini conveyor belt. The color of the object is then scanned using a color sensor on the second DOBOT Magician robot. The results obtained after placing the object on the conveyor belt are evaluated in the DOBOT Studio console. The sorting area is represented by an intuitionistic fuzzy interpretation of the red-green-blue (RGB) color system. A conversion process is performed using color coordinates and conversion formulas to obtain intuitionistic fuzzy evaluations. Research work on waste sorting is discussed in [37, 48]. Examples of intuitionistic fuzzy evaluations in various fields of science have already been published [64, 71, 100].

The intuitionistic fuzzy interpretation of "Red-Green-Blue (RGB) color system" is used for garbage sorting intuitionistic fuzzy evaluations calculation. The IF-geometrical interpretation is presented by an equilateral triangle with vertices assigned to the three basic colors "Red", "Green" and "Blue" with coordinates respectively $\langle 0, 0\rangle$, $\langle \frac{2\sqrt{3}}{3}, 0\rangle$ and $\langle \frac{sqrt3}{3}, 1\rangle$. According to [8], the equilateral triangle interpreted as the "Red-Green-Blue (RGB) color system is presented in Fig. 3.

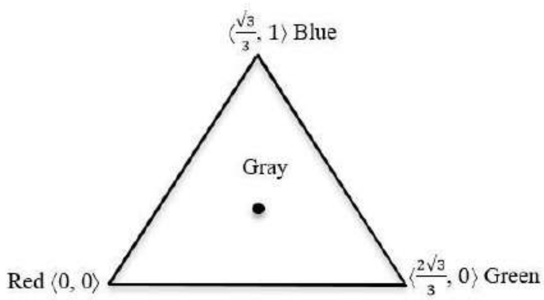

Fig. 3. *The equilateral triangle interpreted the "Red-Green-Blue (RGB) colour system*

The color data from the DOBOT Magician robot is represented by an equilateral triangle interpreting the red-green-blue (RGB) color system. Then, the recognition results are transformed into a rectangular intuitionistic fuzzy triangle using the transformation function F with the following form:

$$\langle u, v \rangle = F(x, y) = \langle \frac{x\sqrt{3}-y}{2}, y \rangle .$$

Thus, the three vertices of the equilateral triangle are transformed into the right-angled triangle in Fig. 3. They have the following form:

$$F(0, 0) = \langle 0, 0 \rangle,$$

$$F(\frac{2\sqrt{3}}{3}, 0) = \langle 1, 0 \rangle,$$

$$F(\frac{\sqrt{3}}{3}, 1) = \langle 0, 1 \rangle.$$

As a prototype for the concept, an installation consisting of two DOBOT Magician robots and a mini conveyor belt has been created.

2.2.2. Intuitionistic fuzzy evaluations for user request frequency

Intuitionistic fuzzy evaluations can help optimize databases, especially in the context of big data. Let us have a database in which an IFE is assigned to each fact. These IFEs will indicate the most frequently used facts and the least frequently used ones. Let's say that initially the facts have an IFE of $\langle 0, 1 \rangle$. If the relevant fact is used, its IRO increases to $\langle c, 1 - c \rangle$, where c is a fixed constant in the interval [0, 1].

Let in some time-moment the fact F have the IFE $\langle a, b \rangle$. Then, after its next use, the evaluation will be $\langle a + c - ac, b - bc \rangle$.

If the fact is not used for a long period of time, then its IFE will obtain the value $\langle a - c + ac, b + c - ac \rangle$.

Nowadays, large volumes of data in various formats are collected from multiple data sources. The performance of data queries on large data sets is improved to enable good parallel or distributed query execution. The diversity of information requires different types of storage for the purpose of storing and processing the information.

2.2.3. Application of the support vector machine for detecting scientific terms and determining keywords in a system for offering educational resources and scientific articles

Support Vector Machine is a supervised machine learning algorithm which is widely used for classification and regression. First proposed by Boser et al., 1992, Guyon et al., 1993, Cortes and Vapnik, 1995, Vapnik et al, 1997 [24, 49, 28], SVM is particularly effective for large-scale data sets, making it applicable in many areas.

In order to create a classifier that recognizes whether a given word is a scientific term, it is first necessary to formalize the textual information by representing the words in numerical form suitable for input into an SVM model. This is accomplished by vectorizing the words.

Determining whether a word or phrase is a scientific term can be useful in algorithms based on intuitionistic fuzzy evaluations, which make decisions or give recommendations on potential additional keywords, as well as in algorithms that recommend similar/related scientific articles or educational resources.

Let's look at suggesting similar scientific articles or educational resources based on keywords.

The IFE of keywords shows the relevance of the potential training resource to the keywords in the submitted request. The higher the relevance, the greater the correspondence of the resource based on the keywords. Resources with a relevance score of 0 are rejected at this stage. Let us consider the following examples:

a - matching keywords count;

b - non-matching keywords count;

c - total keywords count;

c - a - b - partially matching keywords count;

By this, the IFE of the i-th learning resource, in respect of the keywords, is calculated as follows:

$$k_i = \langle \mu_k, \nu_k \rangle,$$

where

$$\mu_k = \frac{a}{c}; \quad \nu_k = \frac{b}{c};$$

The obtained IFEs can be sorted in ascending order by their degree of membership and select the first few items for recommendation.

2.2.4. Application of intuitionistic fuzzy evaluations of student academic performance in school using intuitionistic fuzzy evaluations

In this dissertation, a proposed methodology applies the principles of Intuitionistic Fuzzy Sets (IFS) to assess academic performance on three levels: individual students, class groups, and schools. The approach is based on the construction of Intuitionistic Fuzzy Evaluations (IFE) for each entity, derived from academic performance data, typically obtained from electronic gradebooks.

Each student has grades for multiple subjects. For each subject we calculate the intuitionistic fuzzy evaluation for each subject and then aggregate the intuitionistic fuzzy evaluations to obtain the student's overall intuitionistic fuzzy evaluation. Let us define the following:

- Let S_i be the *i*-th subject studied by the student;

- Let $G_{S_i} = \{g_1, g_2, ..., g_k\}$ be the subset of k grades, obtained by the student in subject S_i, where $g_i \in \{2,3,4,5,6\}$;
- Let $g_{\max} = 6$ be the maximum possible grade;
- Define the average grade in subject S_i as follows: $\overline{g}_{S_i} = \frac{1}{k}\sum_{i=1}^{k} g_i$.

The normalized degree of success (membership) is given by:

$$\mu_{S_i} = \frac{\overline{g}_{S_i}}{g_{\max}}$$

The degree of difficulty (non-membership) is defined using a buffer parameter $\delta \in [0, 0.2]$, typically δ=0.1 as follows:

$$\nu_{S_i} = \max(0, 1 - \mu_S - \delta)$$

The hesitation degree (uncertainty) is calculated as:

$$\pi_{S_i} = 1 - \mu_{S_i} - \nu_{S_i}$$

The final intuitionistic fuzzy evaluation for subject S_i is:

$$IFE_{S_i} = \langle \mu_{S_i}, \nu_{S_i} \rangle$$

Let a student A_i be enrolled in n subjects $S_1, S_2, ..., S_n$, each with an associated intuitionistic fuzzy evaluation:

$$IFE_{S_j}^{(A_i)} = \langle \mu_{S_j}^{(A_i)}, \nu_{S_j}^{(A_i)} \rangle,$$

for $j = 1,2,...,n$.

The overall intuitionistic fuzzy evaluation for student A_i denoted as:

$$IFE_{A_i} = \langle \mu_{A_i}, \nu_{A_i} \rangle$$

The degree is computed by averaging the respective components over all subjects. Thus, the membership degree is:

$$\mu_{A_i} = \frac{1}{n}\sum_{j=1}^{n}\mu_{s_i}^{(A_i)}$$

The non-membership degree is:

$$\nu_{A_i} = \frac{1}{n}\sum_{j=1}^{n}\nu_{s_i}^{(A_i)}$$

The uncertainty degree is:

$$\pi_{A_i} = \frac{1}{n}\sum_{j=1}^{n}\pi_{s_i}^{(A_i)}$$

This evaluation provides a comprehensive intuitionistic fuzzy profile of student A_i based on performance across all enrolled subjects.

Let class C_j consists of m students $A_1^{(C_j)}, A_2^{(C_j)}, \ldots, A_m^{(C_j)}$, each with an individual intuitionistic fuzzy evaluation:

$$IFE_{A_i}^{(C_j)} = \left\langle \mu_{A_i}^{(C_j)}, \nu_{A_i}^{(C_j)} \right\rangle,$$

for $i = 1,2,\ldots,m$.

The overall intuitionistic fuzzy evaluation for class C_j is defined as:

$$IFE_{C_j} = \left\langle \mu_{C_j}, \nu_{C_j} \right\rangle$$

with components calculated as:

- membership degree:

$$\mu_{C_j} = \frac{1}{m}\sum_{i=1}^{m}\mu_{A_i}^{(C_j)}$$

- non-membership degree:

$$\nu_{C_j} = \frac{1}{m}\sum_{i=1}^{m}\nu_{A_i}^{(C_j)}$$

- uncertainty degree:

$$\pi_{C_j} = \frac{1}{m}\sum_{i=1}^{m}\pi_{A_i}^{(C_j)}$$

This evaluation provides a class-level intuitionistic fuzzy profile, reflecting the general academic status, difficulty levels, and performance uncertainty within class C_j.

Let school S_r consists of l classes $C_1^{(S_r)}, C_2^{(S_r)}, ..., C_l^{(S_r)}$, each with a class-level intuitionistic fuzzy evaluation:

$$IFE_{C_j}^{(S_r)} = \left\langle \mu_{C_j}^{(S_r)}, \nu_{C_j}^{(S_r)} \right\rangle$$

for $j = 1,2,...,l$.

The overall intuitionistic fuzzy evaluation for school S_r is defined as:

$$IFE_{S_r} = \left\langle \mu_{S_r}, \nu_{S_r} \right\rangle$$

with components calculated as:

- membership degree:

$$\mu_{S_r} = \frac{1}{l}\sum_{j=1}^{l} \mu_{C_j}^{(S_r)}$$

- non-membership degree:

$$\nu_{S_r} = \frac{1}{l}\sum_{j=1}^{l} \nu_{C_j}^{(S_r)}$$

- uncertainty degree:

$$\pi_{S_r} = \frac{1}{l}\sum_{j=1}^{l} \pi_{C_j}^{(S_r)}$$

This evaluation provides a high-level overview of the academic performance, learning difficulties, and uncertainty patterns across the entire school S_r, based on class-level aggregation.

The presented process of forming intuitionistic fuzzy evaluations and the analysis afterwards can be modelled using Generalized nets as suggested by the research work of Sotirova et al [99, 98, 100, 97].

2.2.5. Application of Intuitionistic Fuzzy Evaluations for Data Analysis from State Matriculation Exams in the Bulgarian Secondary Education System

This dissertation examines a methodology based on the construction of intuitionistic fuzzy evaluations, based on average results from a given State matriculation exam (SME) for a school, town, and region. Data on State matriculation exams results for each academic year are available on the Bulgarian open data portal [76].

In the first step of the methodology, the data is processed and the IFE is calculated for each SME conducted in a given school. Since some schools have a larger number of students who took the exam than others, this is reflected in the degree of uncertainty—the fewer students who took the exam, the higher the degree of uncertainty. The IFE for a result in a given subject is calculated by first normalizing the results, taking into account that the results are in the Bulgarian six-point system (from 2 to 6):

$$\mathrm{norm} = \frac{\gamma - \gamma_{\min}}{\gamma_{\max} - \gamma_{\min}},$$

where γ is the result of SME for a subject for the school, $\gamma_{\min} = 2$ and $\gamma_{\max} = 6$, in the scope of the Bulgarian school education system. We assign: $\mu_{base} = norm$.

The degree of uncertainty is calculated with respect to the count of students who held the exam – n and sensitivity coefficient $\alpha = 0.5$, as follows:

$$\pi_{\mathrm{base}} = \min\left(\frac{\alpha}{\sqrt{n}}, 1\right).$$

The degree of non-membership is calculated as follows: $\nu_{base} = 1 - \mu_{base} - \pi_{base}$

To obtain the final IFE we normalize the values:

$$T = \mu_{base} + \nu_{base} + \pi_{base}$$

$$\mu = \frac{\mu_{base}}{T}, \nu = \frac{\nu_{base}}{T}, \pi = 1 - \mu - \nu$$

The IFE for SME for the i-th subject in a school is denoted as: $\mathrm{IF}_{s_i} = \langle \mu_i, \nu_i \rangle$ and $\pi_i = 1 - \mu_i - \nu_i$.

After the successful calculation of the IFEs for each subject for the school, the aggregated IFE for the school should be calculated. Let's denote the subjects count for which the school held SMEs with *m*.

$$\mathrm{IF}_{school} = \left\langle \frac{1}{m}\sum_{i=1}^{m}\mu_i,\ \frac{1}{m}\sum_{i=1}^{m}\nu_i,\ \frac{1}{m}\sum_{i=1}^{m}\pi_i \right\rangle$$

In similar manner we can obtain IFE for a locality based on the IFEs of the schools in it. The school count in the locality is denoted by *k*.

$$\mathrm{IF}_{locality} = \left\langle \frac{1}{k}\sum_{j=1}^{k}\mu_j,\ \frac{1}{k}\sum_{j=1}^{k}\nu_j,\ \frac{1}{k}\sum_{j=1}^{k}\pi_j \right\rangle$$

Using localities IFEs we can aggregate and obtain IFE for a municipality. The number of schools in a municipality is *l*.

$$\mathrm{IF}_{municipality} = \left\langle \frac{1}{l}\sum_{j=1}^{l}\mu_j,\ \frac{1}{l}\sum_{j=1}^{l}\nu_j,\ \frac{1}{l}\sum_{j=1}^{l}\pi_j \right\rangle$$

Using municipalities IFEs we can aggregate and obtain IFE for the region. The number of schools in a region is *q*.

$$\mathrm{IF}_{region} = \left\langle \frac{1}{q}\sum_{j=1}^{q}\mu_j,\ \frac{1}{q}\sum_{j=1}^{q}\nu_j,\ \frac{1}{q}\sum_{j=1}^{q}\pi_j \right\rangle$$

With the obtained results we can apply the DBSCAN algorithm in which we can choose which data to use for its application:

- school results for given subject;
- school results for a whole school in comparison with the other schools;
- results for localities;
- results for regions.

The presented process of forming IFEs and the analysis afterwards and the respective decision-making can be modelled using generalized nets by the example of the research work of Sotirova et al [99, 98, 100, 97].

Chapter 3. Integrated implementations of models using intuitionistic fuzzy logic and intelligent data analysis algorithms

This chapter presents software implementations of algorithms based on intuitionistic fuzzy sets, indexed matrices, and machine learning methods. The goal is to build intelligent tools for analysis and decision-making in various contexts, including optimization, classification, and educational analysis. The application of C++, Python, and JavaScript in the implementation of the algorithms is described, and the importance of combining mathematical rigor with practical applicability in the analysis of large and complex data is emphasized.

3.1 Implementation of algorithm for solving intuitionistic fuzzy knapsack problem using indexed matrices for intelligent data analysis

In this dissertation an algorithm for solving the intuitionistic fuzzy knapsack problem by indexed matrices is presented.

Step 1. We construct the following indexed matrix (IM) according to the problem (9):

$$A[K, L, \{a_{k_i,p(w)}\}] = \begin{array}{c|cc} & p & w \\ \hline k_1 & \langle \mu_{k_1,p}, \nu_{k_1,p} \rangle & \langle \mu_{k_1,w}, \nu_{k_1,w} \rangle \\ \vdots & \vdots & \vdots \\ k_m & \langle \mu_{k_m,p}, \nu_{k_m,p} \rangle & \langle \mu_{k_m,w}, \nu_{k_m,w} \rangle \end{array}, \tag{9}$$

where $K = \{k_1, \ldots, k_m\}$ for $i = 1, \ldots, m$ and $L = \{p, w\}$ with intuitionistic fuzzy (IF) elements. $\{a_{k_i,p}, a_{k_i,w}\}$ are respectively IF profit and weight of k_i-th object.
The data values for profit and weight of k_i-th object are transformed into IFPs as demonstrated in [112]. Let us have the set of intervals $[i_1, i_I]$ for $1 \le i \le m$ and let:

$$A_{\min,i} = \min_{i_1 \le j \le i_I} x_{i,j} < \max_{i_1 \le j \le i_I} x_{i,j} = A_{\max,i}.$$

The conditions $0 \le \mu_{i,j}, \nu_{i,j} \le 1$ and $0 \le \mu_{i,j} + \nu_{i,j} \le 1$ are satisfied. We can use the expert approach described in detail in [8] to convert the data into IFPs.
The IFP $a_{k_i,p}$ (or $a_{k_i,w}$) presents

- degree of perception – the positive evaluation of an expert for the profit (or weight) of k_i-th object divided by the (maximum-minimum) evaluation of profit (or weight);
- degree of non-perception – the negative evaluation of the of the expert for the k_i-th object divided by the (maximum-minimum) evaluation.

The hesitation degree $\pi_{k_i,p} = 1 - \mu_{k_i,p} - \nu_{k_i,p}$, ($\pi_{k_i,w} = 1 - \mu_{k_i,w} - \nu_{k_i,w}$) corresponds to the uncertain evaluation of the expert for the profit (weight) of the k_i-th object.
At this step, a check is made on the input data for the weight of the objects for not exceeding the capacity of the knapsack C as follows:

for $i = 1$ to m:
if $a_{k_i,w} > C$ then $A_{(k_i,\perp)}$

Let we denote by $|K| = m$ the number of elements of the set K, then $|L| = 2$. We also define:

$$X[K, L] = \begin{array}{c|cc} & p & w \\ \hline k_1 & x_{k_1,p} & x_{k_1,w} \\ \vdots & \vdots & \vdots \\ k_m & x_{k_m,p} & x_{k_m,w} \end{array}, \quad (10)$$

for $1 \le i \le m$:

$$\{x_{k_i,w}, x_{k_i,p}\} \in \begin{cases} \langle 1,0 \rangle, & \text{if the product } k_i \text{ is selected} \\ \langle 0,1 \rangle, & \text{otherwise} \end{cases}$$

Let in the beginning of the algorithm, all elements of IM X X are equal to $\langle 0, 1 \rangle$. Construct IM S^0 for the initial state:

$$S^0[u_0, L] = \begin{array}{c|cc} & p & w \\ \hline u_0 & \langle 0, 0 \rangle & \langle 0, 0 \rangle \end{array}.$$

Step 2. For $i = 1$ to m, do:

- Create IMs $R_i[k_i, L] = pr_{k_i,L} A$;

- Calculate $SH^{i-1}1 = \left[\frac{u_i}{ui-1}; \perp\right] S^{i-1}$

For $h = 1$ to $i+1$, do:

- $SH^{i-1}1 = SH^{i-1}1 \oplus_{(\max,\min)} \left[\frac{u_h}{k_i}; \perp\right] R_i$

Afterwards:

- $S^i[U^i, L] = S^{i-1} \oplus_{(\max,\min)} SH_1^{i-1};$

For $h = 1$ to $i+1$, do:

- Check the conditions for the capacity of the knapsack:
  If $s^i h, w > C$, then $S^i(h, \perp)$
- Perform operation "Purge": U^i: $S^i = Purge_{U^i} S^i$

Step 3. At this step, the index of the largest IF profit is found by: $Index_{(\max_R),p}(A) = \langle u_g, p \rangle$.

For $i = m$ to 1, do:

- Find α-nearest elements of $s^i_{u_g,p}$ (or $s^i_{u_g,w}$) of S^i, where $\alpha = 0.5$ and choose the closest element from them: $s^i_{ug*,p}$ (or $s^i_{ug*,w}$).
- If $s^i_{ug*,p} \in S^i$ and $s^i_{ug*,p} \notin S^{i-1}$, then:
  - $x_{k_i,p} = \langle 1, 0 \rangle$ and $x_{k_i,w} = \langle 1, 0 \rangle$
  - Update: $s^i_{u_g,p} = s^i_{u_g*,p} - a_{k_i,p}; \quad s^i_{u_g,w} = s^i_{u_g*,w} - a_{k_i,w}$.

Step 4. The optimal profit is: $AGIO_{\oplus_{(\#_q)}} \left(pr_{K,p}A \otimes_{(\min,\max)} pr_{K,p}X\right)$.

If $q = 1$, then we obtain a pessimistic scenario and the operation finds the pessimistic value of the addition of all elements. If $q = 2$, we obtain an average scenario. If $q = 3$, then we obtain an optimistic value of the addition of all elements. The decision maker chooses which forecast to prefer.

The optimal weight is:

$$AGIO_{\oplus_{(\#_q)}} \left(pr_{K,w}A \otimes_{(\min,\max)} pr_{K,w}X\right).$$

In this dissertation a software implementation in C++ for the above algorithm is created and applied for an example dataset of synthetic data.

3.2 Implementation of algorithm for solving circular intuitionistic fuzzy knapsack problem using indexed matrices for intelligent data analysis

In this dissertation an algorithm for solving the circular intuitionistic fuzzy knapsack problem by indexed matrices is presented.

Step 1. Construction of 3D evaluation intuitionistic fuzzy indexed matrix (IFIM)

We form 3D evaluation IFIM: $EV[K, C, E, \{ev_{k_i,c_j,d_s}\}]$, where:

- $K = \{k_1, k_2, \ldots, k_m\}$ – set of requests;
- $C = \{c_1, c_2\}$ – criteria: urgency (c_1) and expected duration (c_2);
- $E = \{d_1, \ldots, d_s, \ldots, d_D\}$ – experts;
- $\{ev_{k_i,c_j,d_s}\} = \langle \mu_{k_i,c_j,d_s}, \nu_{k_i,c_j,d_s} \rangle$ is the estimate of expert d_s for request k_i by criteria c_j.

The expert is not sure about its evaluation according to a given criterion due to changes in some uncontrollable factors and his evaluations are in the form of intuitionistic fuzzy pairs (IFPs). The score (rating) of expert d_s is represented by IFP: $re_s = \langle \delta_s, \epsilon_s \rangle$, where δ_s is the degree of competency and ϵ_s – the degree of incompetency. Then we create the following matrix:

$$EV^*[K, C, E, \{ev^*{}_{k_i,c_g,d_s}\}] = re_1 pr_{K,C,d_1} EV \oplus_{(\circ_1,\circ_2)} \ldots \oplus_{(\circ_1,\circ_2)} re_D pr_{K,C,d_D} EV.$$

$$EV := EV^*(ev_{k_i,l_j,d_s} = ev^*_{k_i,l_j,d_s},\ \forall k_i \in K, \forall l_j \in L, \forall d_s \in E).$$

α_E-th aggregation operation is applied to find the averaged IF value of the k_i-th request by the c_i-th criteria in a given moment $h_f \notin E$:

$$PI[K, C, h_f, \{pi_{k_i,c_g,h_f}\}] = \alpha_{E,\#_2}(EV^*, h_f) = \begin{array}{c|cc} h_f & c_1 & c_2 \\ \hline k_1 & \underset{s=1}{\overset{D}{\#_2}} \langle \mu_{k_1,c_1,d_s}, \nu_{k_1,c_1,d_s} \rangle & \underset{s=1}{\overset{D}{\#_2}} \langle \mu_{k_1,c_2,d_s}, \nu_{k_1,c_2,d_s} \rangle \\ \vdots & \vdots & \vdots \\ k_m & \underset{s=1}{\overset{D}{\#_2}} \langle \mu_{k_m,c_1,d_s}, \nu_{k_m,c_1,d_s} \rangle & \underset{s=1}{\overset{D}{\#_2}} \langle \mu_{k_m,c_2,d_s}, \nu_{k_m,c_2,d_s} \rangle \end{array},$$

where $h_f \notin E$.

Now we construct Circular-IFIM (C-IFIM) A, which consists of the evaluations in a present moment h_f for each request k_i by the two criteria:

$$A[K, C\{a_{k_i,c_g}\}] = \begin{array}{c|cc} & c_1 & c_2 \\ \hline k_1 & \langle \mu^a_{k_1,c_1}, \nu^a_{k_1,c_1}; r^a_{k_1,c_1} \rangle & \langle \mu^a_{k_1,c_2}, \nu^a_{k_1,c_2}; r^a_{k_1,c_2} \rangle \\ \vdots & \vdots & \vdots \\ k_m & \langle \mu^a_{k_m,c_1}, \nu^a_{k_m,c_1}; r^a_{k_m,c_1} \rangle & \langle \mu^a_{k_m,c_2}, \nu^a_{k_m,c_2}; r^a_{k_m,c_2} \rangle \end{array},$$

where $K = \{k_1, \ldots, k_i, \ldots, k_m\}, i = 1, \ldots, m; C = \{c_1, c_2\}, a_{k_i,c_g}$ for $i = 1, \ldots, m; g = 1, 2$ are obtained as C-IFPs by transforming the IFPs pi_{k_i,c_j,d_s} in C-IFPs:

- $\mu^a_{k_i,c_g} = \mu^{pi}_{k_i,c_g,h_f}$;
- $\nu^a_{k_i,c_g} = \nu^{pi}_{k_i,c_g,h_f}$;

- Radius $r^a_{k_i,c_g}$ is calculated as follows: $r^a_{k_i,c_g} = \max_{1\le s\le D}\left|\sqrt{\left(\mu^{ev}_{k_i,c_g,d_s} - \mu^{pi}_{k_i,c_g,h_f}\right)^2 + \left(\nu^{ev}_{k_i,c_g,d_s} - \nu^{pi}_{k_i,c_g,h_f}\right)^2}\right|$.

A similar approach for construction C-IFPs is developed in [16, 61].

Then, a check is made on the input data for the duration of the requests for not exceeding the given time of the ambulance team T. For $i = 1$ to m: If $a_{k_i,c_2} > T$, then $A_{(k_i,\perp)}$.

Let us denote by $|K| = m$ the number of the elements of the set K, then $|C| = 2$. We also define C-IFIM $X[K, C]$ with elements x_{k_i,c_g} (for $1 \le i \le m, \ \le g \le 2$) and:

$$\{x_{k_i,c_g}\} \in \begin{cases} \langle 1, 0; \sqrt{2}\rangle \\ \langle 0, 1; \sqrt{2}\rangle \end{cases}$$

Let in the beginning of the algorithm, all elements of IM X are equal to $\langle 0, 1; 0\rangle$. Construct IM $S^0[u_0, L] = \begin{array}{c|cc} & c_1 & c_2 \\ \hline u_0 & s^0_{u_0,c_1} & s^0_{u_o,c_2} \end{array} = \begin{array}{c|cc} & c_1 & c_2 \\ \hline u_0 & \langle 0, 1; \sqrt{2}\rangle & \langle 0, 1; \sqrt{2}\rangle \end{array}$.

Step 2.

For $i = 1$ to m:

Create IM $R_i[k_i, C] = pr_{k_i,C}A$;

$SH_1^{i-1} = \left[\frac{u_i}{u_{i-1}}; \perp\right] S^{i-1}$

For $h = 1$ to $i + 1$:

$SH_1^{i-1} = SH_1^{i-1} \oplus_{(\circ_1,\circ_2,*)} \left[\frac{u_h}{k_i}; \perp\right] R_i$

$S^i[U^i, L] = S^{i-1} \oplus_{((\circ_1,\circ_2,*))} SH_1^{i-1};$

For $h = 1$ to $i + 1$:

If $s^i_{h,w} > T$, then $S^i_{(h,\perp)}$.

Apply "Purge" operation $S^i = Purge_{U^i}S^i$.

Step 3.

At this step, the index of the most emergency request is found by:

$Index_{(\max_{R^{circ}}),c_1}(A) = \langle u_g, c_1\rangle$

For $i = m$ to 1:

Find the α-nearest elements of $s^i_{u_g,c_1}$ (or $s^i_{u_g,c_2}$) ($\alpha = 0.05$) of S^i and choose the closest element from them - $s^i_{u_{g*},c_1}$ (or $s^i_{u_{g*},c_2}$).

If $\{s^i_{u_g*,c_1}$ (or $s^i_{u_{g*},c_2}$)$\} \in S^i$ and $\{s^i_{u_g*,c_1}$ (or $s^i_{u_{g*},c_2}$)$\} \notin S^{i-1}$ then:

$x_{k_i,p} = \langle 1, 0; \sqrt{2}\rangle \ \ x_{k_i,w} = \langle 1, 0; \sqrt{2}\rangle$;

$$s^i_{u_g,c_1} = s^i_{u_g*,c_1} -_* a_{k_i,c_1}$$
$$s^i_{u_g,c_2} = s^i_{u_g*,c_2} -_* a_{k_i,c_2}$$

Step 4. The optimal benefit and time of satisfying the emergency requests are:

$$AGIO_{\oplus(\#_q,*)}\left(pr_{K,c_1}A \otimes_{(\circ_1,\circ_2,*)} pr_{K,c_1}X\right); AGIO_{\oplus(\#_q,*)}\left(pr_{K,c_2}A \otimes_{(\circ_1,\circ_2,*)} pr_{K,c_2}X\right).$$

If $q = 1$, $q = 2$ or $q = 3$, then we obtain respectively the pessimistic, the averaged or the optimistic value of the optimal benefit. When in all operations of the algorithm the pair of operators is used $\langle \circ_1, \circ_2 \rangle = \langle \max, \min \rangle$, this means that the optimistic scenario is applied. If $\langle \min, \max \rangle$ is used, then the pessimistic scenario is used. In the case of greater vagueness, the operation $* = \max$ is used, otherwise – $* = \min$.

In this dissertation a software implementation using C++ of the proposed algorithm is created and applied over an example data set.

3.3 Implementation of algorithm for solving elliptic intuitionistic fuzzy knapsack problem using indexed matrices for intelligent data analysis

In this dissertation an algorithm for solving the circular intuitionistic fuzzy knapsack problem by indexed matrices is presented.

Step 1. Construction of 3D evaluation intuitionistic fuzzy indexed matrix (IFIM)

We form 3D evaluation IFIM: $EV[K, C, E, \{ev_{k_i,c_j,d_s}\}]$, where:

- $K = \{k_1, k_2, \dots, k_m\}$ – is a set of assets;
- $C = \{c_1, c_2\}$ – is a set of criteria (for example yields and price);
- $E = \{d_1, \dots, d_s, \dots, d_D\}$ – experts;
- $\{ev_{k_i,c_j,d_s}\} = \langle \mu_{k_i,c_j,d_s}, \nu_{k_i,c_j,d_s} \rangle$ is estimate of expert d_s for asset k_i by criterion c_j.

Due to changes in some uncontrolled elements, the expert is unsure about the evaluation according to a particular criterion, and his evaluations take the shape of IFPs. The score of expert d_s is presented by an IFP: $re_s = \langle \delta_s, \epsilon_s \rangle$, where δ_s is degree of competency, ϵ_s – degree of incompetency. Next, we calculate:

$$EV^*[K, C, E, \{ev^*{}_{k_i,c_g,d_s}\}] = re_1 pr_{K,C,d_1} EV \oplus_{(\circ_1,\circ_2)} \dots \oplus_{(\circ_1,\circ_2)} re_D pr_{K,C,d_D} EV.$$

$$EV := EV^*(ev_{k_i,l_j,d_s} = ev^*_{k_i,l_j,d_s},\ \forall k_i \in K, \forall l_j \in L, \forall d_s \in E).$$

The degrees of membership and non-membership of the elliptic intuitionistic fuzzy pairs (E-IFP) are determined using the three aggregating operations:

- $\alpha_{K,\#_1,*}$ – pessimistic scenario,
- $\alpha_{K,\#_3,*}$ – optimistic scenario,
- $\alpha_{K,\#_2,*}$ – averaged scenario.

$$PI_{min}[K, C, h_m, \{pi_{min\,k_i,c_g,h_f}\}] = \alpha_{E,\#_1}(EV^*, h_m)$$

$$
= \begin{array}{c|cc}
h_m & c_1 & c_2 \\
\hline
k_1 & \underset{D}{\overset{s=1}{\#_1}} \langle \mu_{k_1,c_1,d_s}, \nu_{k_1,c_1,d_s} \rangle & \underset{D}{\overset{s=1}{\#_1}} \langle \mu_{k_1,c_2,d_s}, \nu_{k_1,c_2,d_s} \rangle \\
\vdots & \vdots & \vdots \\
k_m & \underset{D}{\overset{s=1}{\#_1}} \langle \mu_{k_m,c_1,d_s}, \nu_{k_m,c_1,d_s} \rangle & \underset{D}{\overset{s=1}{\#_1}} \langle \mu_{k_m,c_2,d_s}, \nu_{k_m,c_2,d_s} \rangle
\end{array}
$$

$PI_{max}[K,C,h_m,\{pi_{max\,k_i,c_g,h_f}\}]=\alpha_{E,\#_3}(EV^*,h_m)$

$$
= \begin{array}{c|cc}
h_m & c_1 & c_2 \\
\hline
k_1 & \underset{s=1}{\overset{D}{\#_3}} \langle \mu_{k_1,c_1,d_s}, \nu_{k_1,c_1,d_s} \rangle & \underset{s=1}{\overset{D}{\#_3}} \langle \mu_{k_1,c_2,d_s}, \nu_{k_1,c_2,d_s} \rangle \\
\vdots & \vdots & \vdots \\
k_m & \underset{s=1}{\overset{D}{\#_3}} \langle \mu_{k_m,c_1,d_s}, \nu_{k_m,c_1,d_s} \rangle & \underset{s=1}{\overset{D}{\#_3}} \langle \mu_{k_m,c_2,d_s}, \nu_{k_m,c_2,d_s} \rangle
\end{array}
$$

Then construct $PI* = PI_{min} \oplus_{(\circ_1,\circ_2,*)} PI_{max}$ and $PI[K,C,h_f,\{pi_{k_i,c_g,h_f}\}] = \alpha_{E,\#_2}(PI^*,h_f)$, whose elements are the coordinates of the centers of the E-IFPs evaluating the asset.

We now form elliptic intuitionistic fuzzy index matrix $A[K,C\{a_{k_i,c_g}\}]$, which represents current evaluations of the assets utilizing the approach from [59] by criteria for return and price:

$$
\begin{array}{c|cc}
 & c_1 & c_2 \\
\hline
k_1 & \langle \mu^a_{k_1,c_1}, \nu^a_{k_1,c_1}; rf^a_{k_1,c_1}, rs^a_{k_1,c_1} \rangle & \langle \mu^a_{k_1,c_2}, \nu^a_{k_1,c_2}; rf^a_{k_1,c_2}, rs^a_{k_1,c_2} \rangle \\
\vdots & \vdots & \vdots \\
k_m & \langle \mu^a_{k_m,c_1}, \nu^a_{k_m,c_1}; rf^a_{k_m,c_1}, rs^a_{k_m,c_1} \rangle & \langle \mu^a_{k_m,c_2}, \nu^a_{k_m,c_2}; rf^a_{k_m,c_2}, rs^a_{k_m,c_2} \rangle
\end{array}
$$

, where $K=\{k_1,\ldots,k_i,\ldots,k_m\}, i=1,\ldots,m; C=\{c_1,c_2\}$, a_{k_i,c_g} (for $i=1,\ldots,m; g=1,2$) are created as E-IFPs by converting the IFP pi_{k_i,c_j,d_s} using the following steps:

For $g=1$ to 2, $i=1$ to m:

$$\mu^a_{k_i,c_g} = \mu^{pi}_{k_i,c_g,h_f}; \nu^a_{k_i,c_g} = \nu^{pi}_{k_i,c_g,h_f}, rf^a_{k_i,c_g}$$

$$= \sqrt{\min_{1\le s\le D}\left(\mu^{ev}_{k_i,c_g,d_s}\right)^2 + \left\{\frac{\max_{1\le s\le D}\left(\mu^{ev}_{k_i,c_g,d_s}\right)^2 - \min_{1\le s\le D}\left(\mu^{ev}_{k_i,c_g,d_s}\right)^2}{\max_{1\le s\le D}\left(\nu^{ev}_{k_i,c_g,d_s}\right)^2 - \min_{1\le s\le D}\left(\nu^{ev}_{k_i,c_g,d_s}\right)^2}\right\}^2 \cdot \min_{1\le s\le D}\left(\nu^{ev}_{k_i,c_g,d_s}\right)^2}$$

$$rs^a_{k_i,c_g}$$

$$= \sqrt{\min_{1\le s\le D}\left(\mu^{ev}_{k_i,c_g,d_s}\right)^2 \cdot \left\{\frac{\max_{1\le s\le D}\left(\nu^{ev}_{k_i,c_g,d_s}\right)^2 - \min_{1\le s\le D}\left(\nu^{ev}_{k_i,c_g,d_s}\right)^2}{\max_{1\le s\le D}\left(\mu^{ev}_{k_i,c_g,d_s}\right)^2 - \min_{1\le s\le D}\left(\mu^{ev}_{k_i,c_g,d_s}\right)^2}\right\}^2 + \min_{1\le s\le D}\left(\nu^{ev}_{k_i,c_g,d_s}\right)^2}$$

The input data for the portfolio's budget is then checked to ensure that it does not exceed the investor's specified budget Bu. If the price of a given asset k_i exceeds the budget Bu, then the corresponding row of IM A is reduced by it.

Let us denote by $|K| = m$ the number of the elements of the set K, then $|C| = 2$. As well, we define elliptic intuitionistic fuzzy index matrix $X[K, C]$, containing the elements x_{k_i,c_g} (for $1 \leq i \leq m, \ \leq g \leq 2$) and: $\{x_{k_i,c_g}\} \in \begin{cases} \langle 1, 0; 0, 0\rangle, & \text{if the asset } k_i \text{ is selected} \\ \langle 0, 1; 0, 0\rangle & \text{otherwise} \end{cases}$.

Let us assume that at the beginning of the algorithm, all components of IM X are identical to $\langle 0, 1; 0, 0\rangle$.

Create IM $S^0[u_0, L] = \begin{array}{c|cc} & c_1 & c_2 \\ \hline u_0 & s^0_{u_0,c_1} & s^0_{u_o,c_2} \end{array} = \begin{array}{c|cc} & c_1 & c_2 \\ \hline u_0 & \langle 0, 1; 0, 0\rangle & \langle 0, 1; 0, 0\rangle \end{array}$.

Step 2.

For $i = 1$ to m do:

Create IM $R_i[k_i, C] = pr_{k_i,C}A; SH_1^{i-1} = \left[\frac{u_i}{u_{i-1}}; \perp\right] S^{i-1}$

For $h = 1$ to $i + 1$:

$\{SH_1^{i-1} = SH_1^{i-1} \oplus_{(\circ_1,\circ_2,*)} \left[\frac{u_h}{k_i}; \perp\right] R_i\}$

$S^i[U^i, L] = S^{i-1} \oplus_{((\circ_1,\circ_2,*))} SH_1^{i-1};$

For $h = 1$ to $i + 1$:

If $s^i_{h,w} > Bu$, then $S^i_{(h,\perp)}$.

The "Purge" operation is applied: $S^i = Purge_{U^i} S^i$.

Step 3.

This step finds the index of the highest stock return by:

$$Index_{(\max_{R^{elliptic}}),c_1}(A) = \langle u_g, c_1\rangle$$

For $i = m$ to 1:

Find the α-nearest elements of $s^i_{u_g,c_1}$ (or $s^i_{u_g,c_2}$) ($\alpha = 0.5$) of S^i and choose the nearest of them – $s^i_{u_{g*},c_1}$ (or $s^i_{u_{g*},c_2}$).

If $\{s^i_{u_g*,c_1}$ (or $s^i_{u_{g*},c_2}$)$\} \in S^i$ and $\{ s^i_{u_g*,c_1}$ (or $s^i_{u_{g*},c_2}$)$\} \notin S^{i-1}$, then

$\{x_{k_i,p} = \langle 1, 0; 0, 0\rangle$ and $x_{k_i,w} = \langle 1, 0; 0, 0\rangle$; $s^i_{u_g,c_1} = s^i_{u_g*,c_1} -_* a_{k_i,c_1}$; $s^i_{u_g,c_2} = s^i_{u_g*,c_2} -_* a_{k_i,c_2}\}$

Step 4.

We find the optimal return and price of the investment portfolio:

$$AGIO_{\oplus_{(\#_q,*)}}\left(pr_{K,c_1}A \otimes_{(\circ_1,\circ_2,*)} pr_{K,c_1}X\right);$$

$$AGIO_{\oplus_{(\#_q,*)}}\left(pr_{K,c_2}A\otimes_{(\circ_1,\circ_2,*)}pr_{K,c_2}X\right).$$

If $q=1, q=2$ or $q=3$, we obtain respectively pessimistic, averaged or optimistic value of the benefit.

In this dissertation the proposed algorithm is implemented in C++ and is applied over an example data set.

3.4 Implementation of a software solution for optimizing shipment delivery based on data analysis and application of the elliptic intuitionistic fuzzy decision-making backpack problem

This section discusses an algorithm similar to the one described in the previous section. It also includes an analysis of sensitivity to changes in the data. A C++ software implementation of the proposed algorithm has been developed and applied to a sample data set, including modified data, in order to analyze the sensitivity of the algorithm.

3.5 Implementation of clustering algorithms using DBSCAN and BIRCH

This section presents the applied implementation and investigation of two popular clustering algorithms – DBSCAN and BIRCH, with the aim of identifying patterns and structures in data through clustering. The implementations are developed in Python, using the scikit-learn, matplotlib, numpy, and pandas libraries. The algorithms are applied to various datasets, including real medical data and synthetically generated sets, in order to demonstrate their effectiveness, visualization, and evaluation through intuitionistic fuzzy indicators.

3.6 Creation of a classification model for determining membership in the "scientific term" category using the Support Vector Machine (SVM) method

This section presents the implementation of a classification model based on the Support Vector Machine (SVM) method, with the aim of automatically determining whether a given word or phrase belongs to the category of "scientific term." Essentially, this is a binary classification task in which each input element (term or phrase) is predicted to be a scientific term or not, based on extracted text characteristics.

The input data used to build and train the SVM classifier consists of text units (words or short phrases) extracted from keywords in articles published in a scientific journal, as well as from other academic and general language texts. Each element in the corpus is manually annotated with a binary label: 1 if it is a scientific term and 0 otherwise.

For the purposes of implementation, keywords from the field of number theory and intuitionistic fuzzy logic have been selected, while the remaining words are common general language words and phrases.

The classification model for scientific terms is built in stages using Python and the scikit-learn library. Each step corresponds to a specific component of machine learning—from data processing to model training and storage.

After creating and training an SVM model for classifying scientific terms, a web-based application programming interface (API) was developed that allows for the automated extraction of key scientific concepts from text. The implementation is built using the Flask web framework, and communication with the service is done via HTTP POST requests with JSON content.

3.7 Implementation of a software system analyzing students' academic performance

In this dissertation, a Python-based software system was developed to support the process of educational analysis and early diagnosis of potential difficulties in students. The main approach embedded in the system is based on intuitionistic fuzzy evaluations (IFE) – a mathematical framework that allows for more flexible and objective modeling of subjective data, taking into account simultaneously the degree of membership, degree of non-membership, and degree of uncertainty in the assessments.

The system was created to process data from real electronic gradebooks, including individual assessments, absences, participation in activities, and other indicators. Each student is represented by multiple IFEs based on various criteria, and the results are used to form aggregated intuitionistic fuzzy evaluations. This makes it possible to classify students by levels of educational risk and to identify groups in need of additional support.

The system was developed using Python, Flask, and the pandas library. The system accepts data files in CSV format (Fig. 41). After the file is submitted, an analysis of the students is performed and the IFE is calculated for each of them.

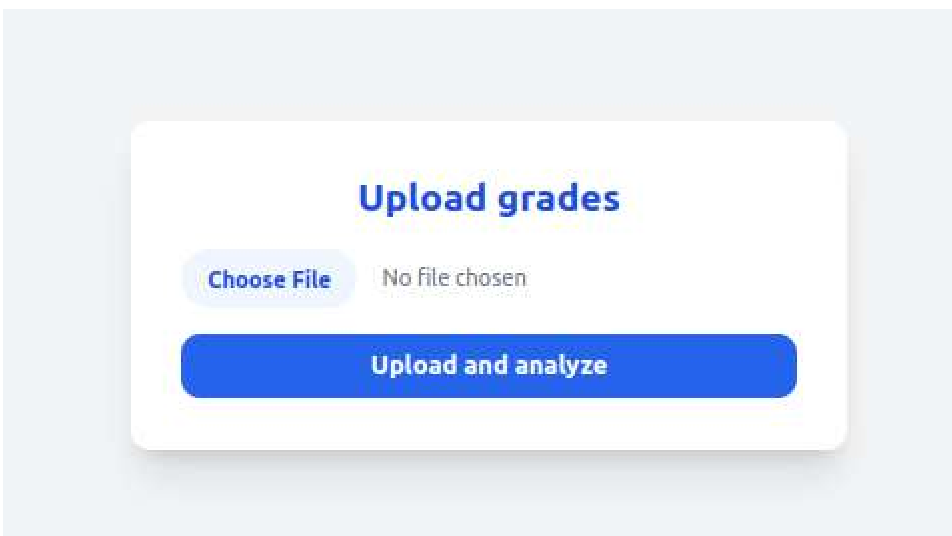

Fig. 4. *Interface for grades importing*

JavaScript and the Chart.js library are used to visualize the graph with the students' results (Fig. 5). In addition to the results, circles with colors corresponding to each student are also visualized, with the color of the group to which the student belongs (Fig. 6). Clicking on the student's circle displays a table with the individual IFEs for each student (Fig. 7).

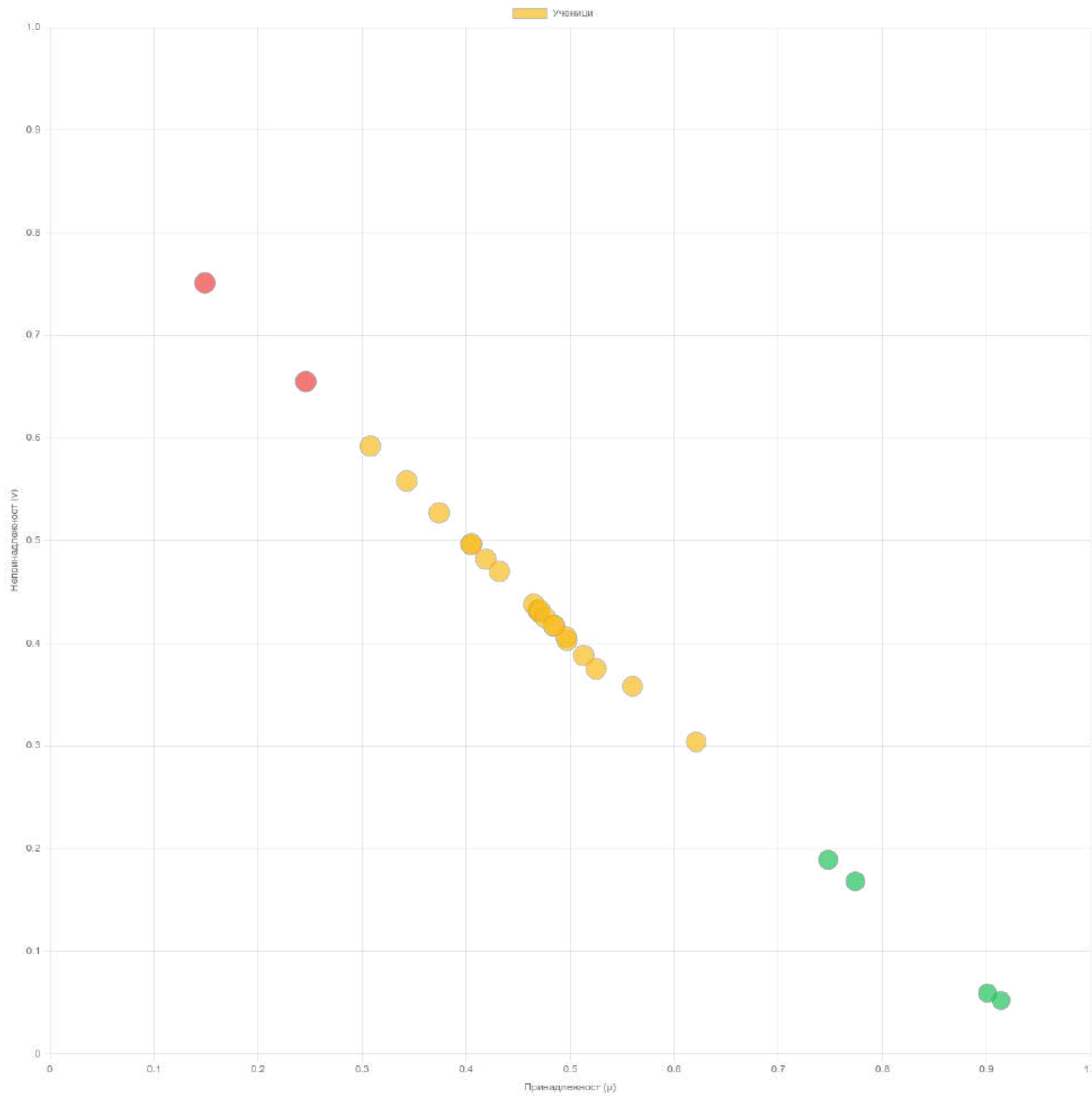

Fig. 5. *Graph with students IFEs*

Fig. 6. *Student results*

Details for student №3

μ: **0.621**, ν: **0.304**, π: **0.076**

Subject	μ	ν	π
English	0.42	0.48	0.1
Biology	0.44	0.46	0.1
Bulgarian	0.5	0.4	0.1
Geography	0.75	0.2	0.05
History	0.94	0.04	0.03
Mathematics	0.5	0.4	0.1
Physics	0.44	0.46	0.1
Sports	0.92	0.05	0.03
Phylosophy	0.88	0.07	0.05
Chemistry	0.42	0.48	0.1

***Fig.** 7. Results by subject for given student*

The development of software implementation of the IFE analysis system is of significant importance, as it transforms the idea of analyzing students' academic performance through intuitive fuzzy evaluations into a practical tool with real applicability in education. Thanks to the automation of calculations, the visualization of results, and the possibility of real-time analysis, the system allows teachers, psychologists, and school administrators to make more informed and reasoned pedagogical decisions. In addition, it creates a basis for long-term data accumulation and comparison, which supports both the individual progress of students and strategic planning at the school level. In this sense, the software implementation not only facilitates analysis - it provides a sustainable foundation for prevention, support, and development in the educational environment.

3.8 Implementation of a software system analyzing the results of the State Maturity Exams (SME) in the secondary education system of the Republic of Bulgaria

The educational results of the State Matriculation Exams (SME) have always been an important indicator for everyone involved in education—students, parents, teachers, principals, and the Ministry of Education and Science. In light of the modernization of the education system, learning for the industry of the future, and the increasingly crucial role of artificial intelligence in the labor market of the future, the topic of student results is becoming increasingly important. In the dissertation, a working framework is proposed and software for data mining is implemented, based on intuitionistic fuzzy evaluations [14, 15] and clustering algorithms [32, 56] for the purpose of analyzing state matriculation exams. The software implementation is based on the Python programming language.

The web-based platform operates in six steps. The first step involves providing data on the results of the DZI to schools (Fig. 8). The data is in JSON format and is loaded and processed in the platform. The DZI data is published on the Bulgarian open data portal [76].

Fig. 8. *Interface for upload of JSON file with results*

After calculating the IFE for each school, the data is submitted to the DBSCAN clustering algorithm. In order to adequately reflect the situation in the education system, it is desirable to select appropriate parameters for the DBSCAN algorithm so that approximately 6-7 clusters are obtained. It is also good for the clusters to depend on the degree of membership, which is why additional segmentation of the clusters by degree of membership of the obtained aggregated IFEs for schools is used.

Within the scope of this dissertation, different values for *eps* and *min_samples* were tested using the dataset for the 2023/2024 academic year. With values of *eps* = 0.5 and *min_samples* = 5, four clusters are obtained and no noise (outliers) is observed (Fig. 9). In this case, the clusters are obtained entirely from segmentation, but without it, a single cluster would be obtained (Fig. 10). This shows that these values are not entirely suitable for the algorithm. However, for a more basic analysis, the result has practical use.

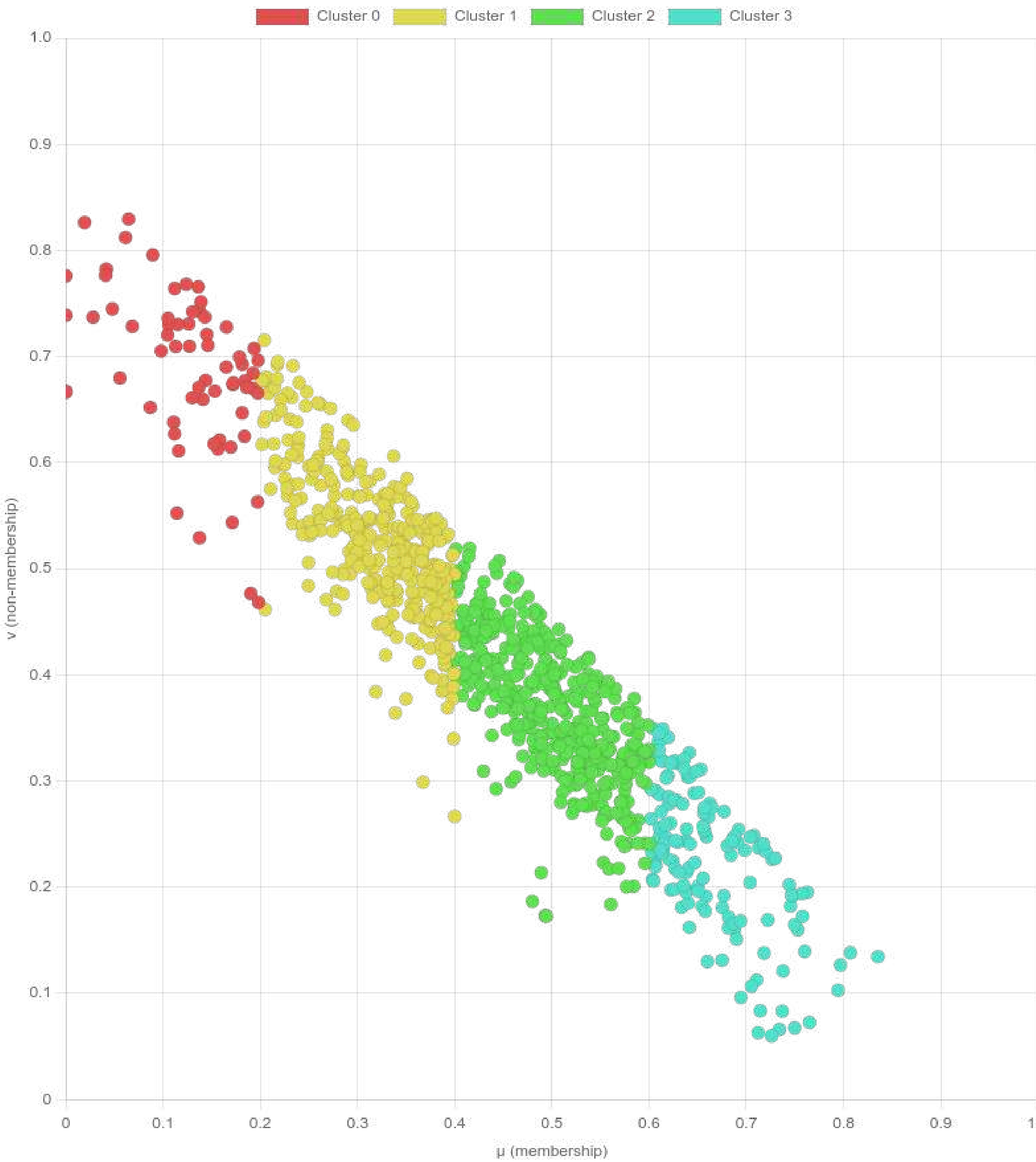

Fig. 9. *Result from DBSCAN at values eps = 0.5 and min_samples = 5 and segmentation by degree of membership*

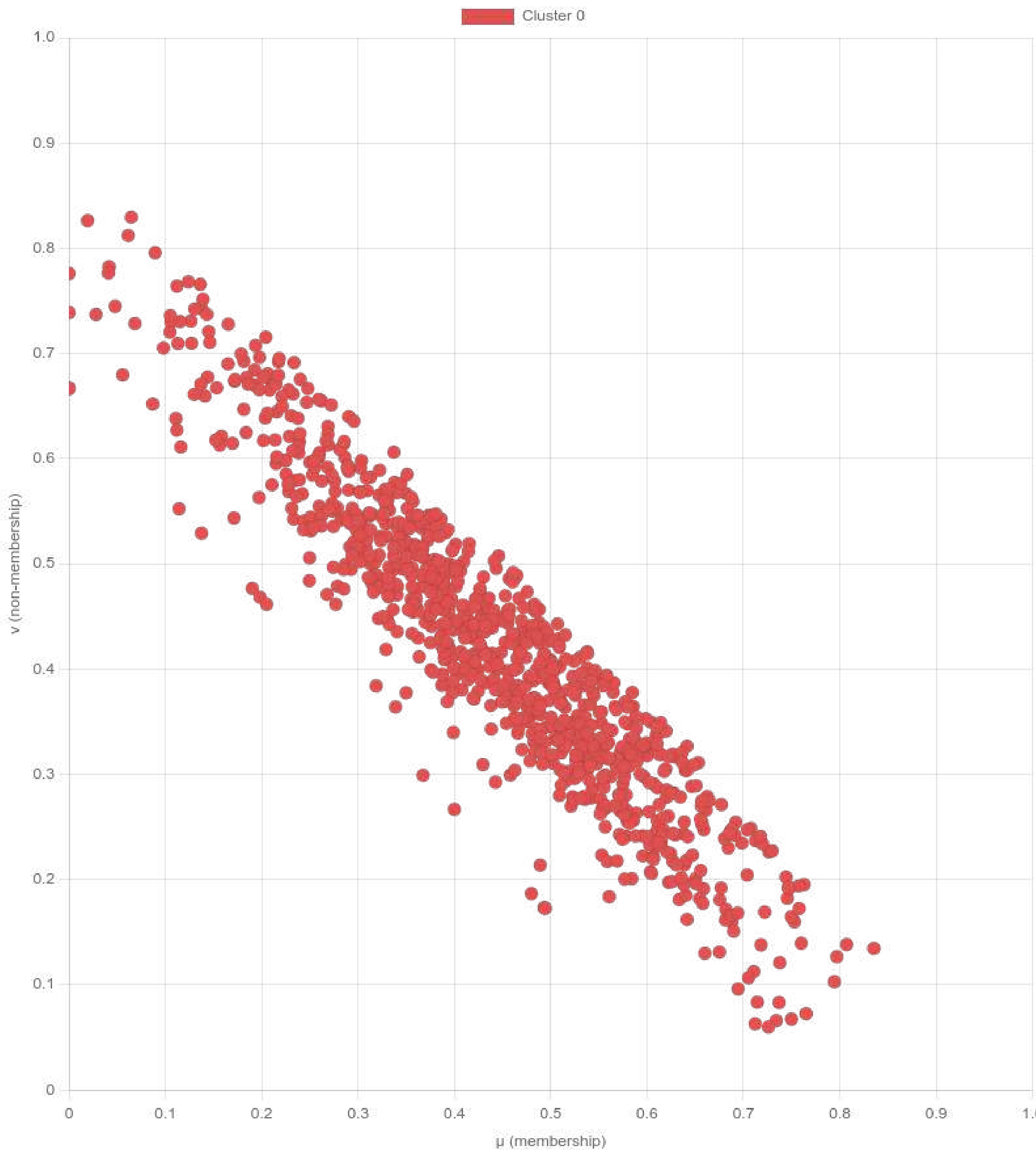

Fig. 10. *Result from DBSCAN at values eps = 0.5 and min_samples = 5 without segmentation by degree of membership*

The next value we will try is eps = 0.01 and min_samples = 5. Without segmentation, we get 34 clusters and quite a few schools that remain as noise (Fig. 11). With segmentation, we get

36 clusters and some schools that remain as noise (Fig. 12).

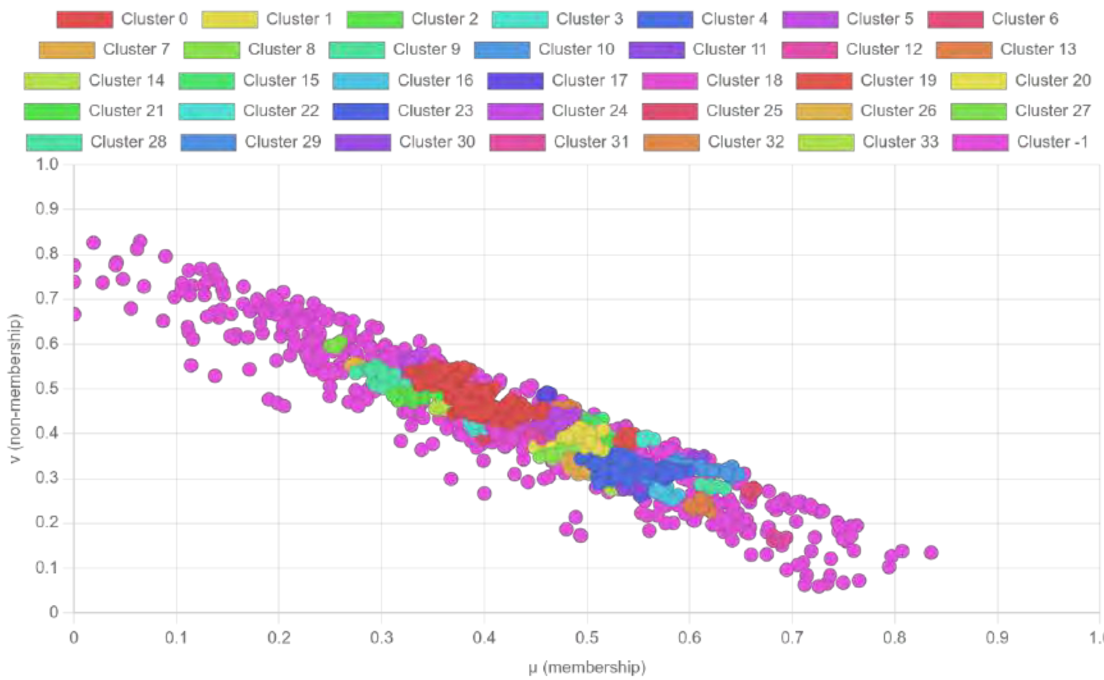

Fig. 11. *Result from DBSCAN at values eps = 0.01 and min_samples = 5 without segmentation by degree of membership*

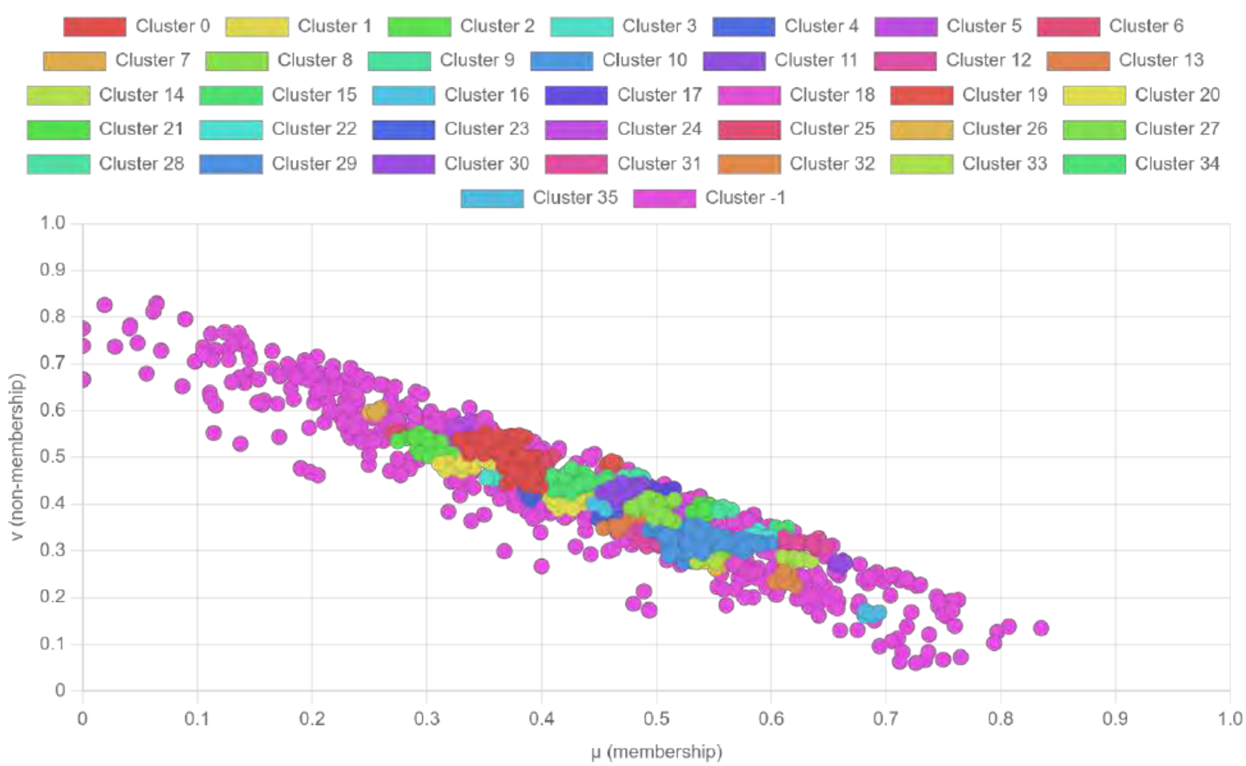

Fig. 12. *Result from DBSCAN at values eps = 0.01 and min_samples = 5 with segmentation by degree of membership*

The presence of too many clusters in this case is rather inconvenient for analysis, but nevertheless could be a potential indicator for complex analysis and a connection between certain schools for which similar measures would be needed. However, in this case, quite a few schools remain outside the clusters as noise, leading to a missed opportunity for analysis and knowledge extraction about them.

With values of eps = 0.02 and min_samples = 5, we obtain significantly fewer clusters and noise. Without segmentation, 5 clusters and noise are formed (Fig. 13), and with segmentation – 8 clusters and noise (Fig. 14). This option is considered convenient for subsequent analysis and conclusions.

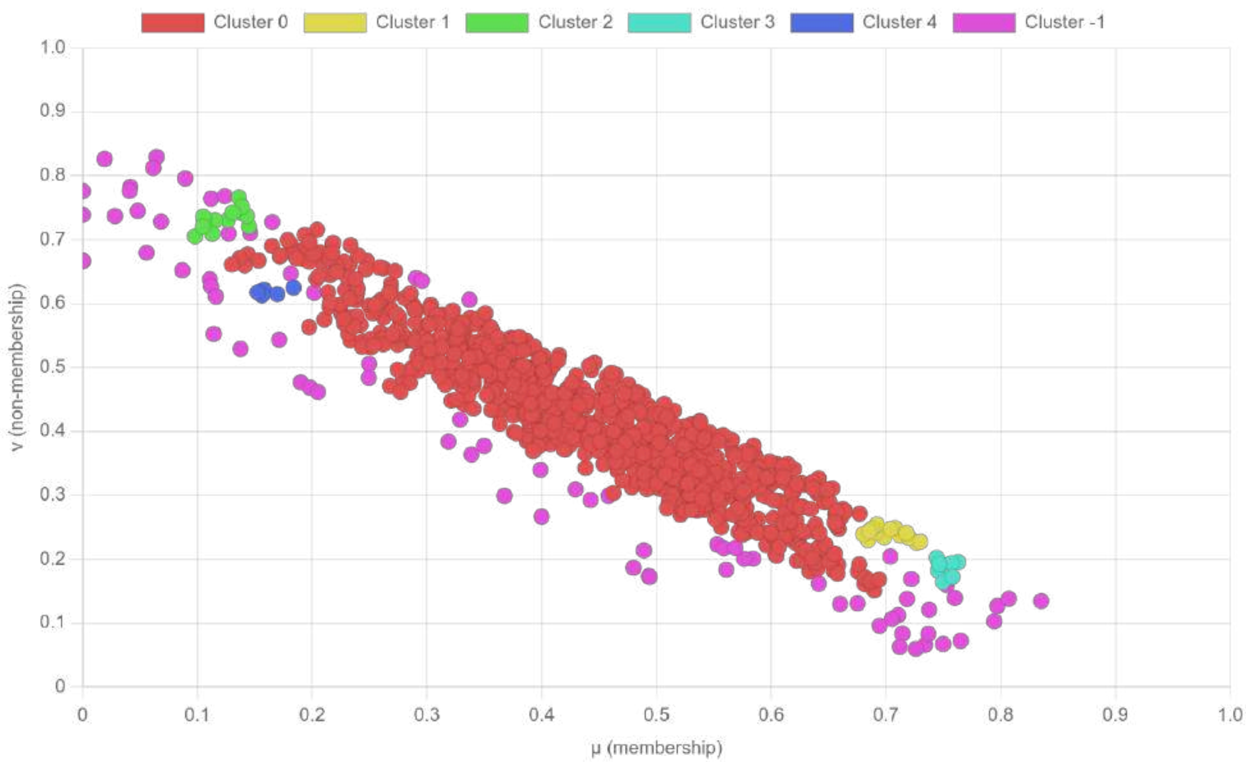

Fig. 13. *Result from DBSCAN at values eps = 0.02 and min_samples = 5 without segmentation by degree of membership*

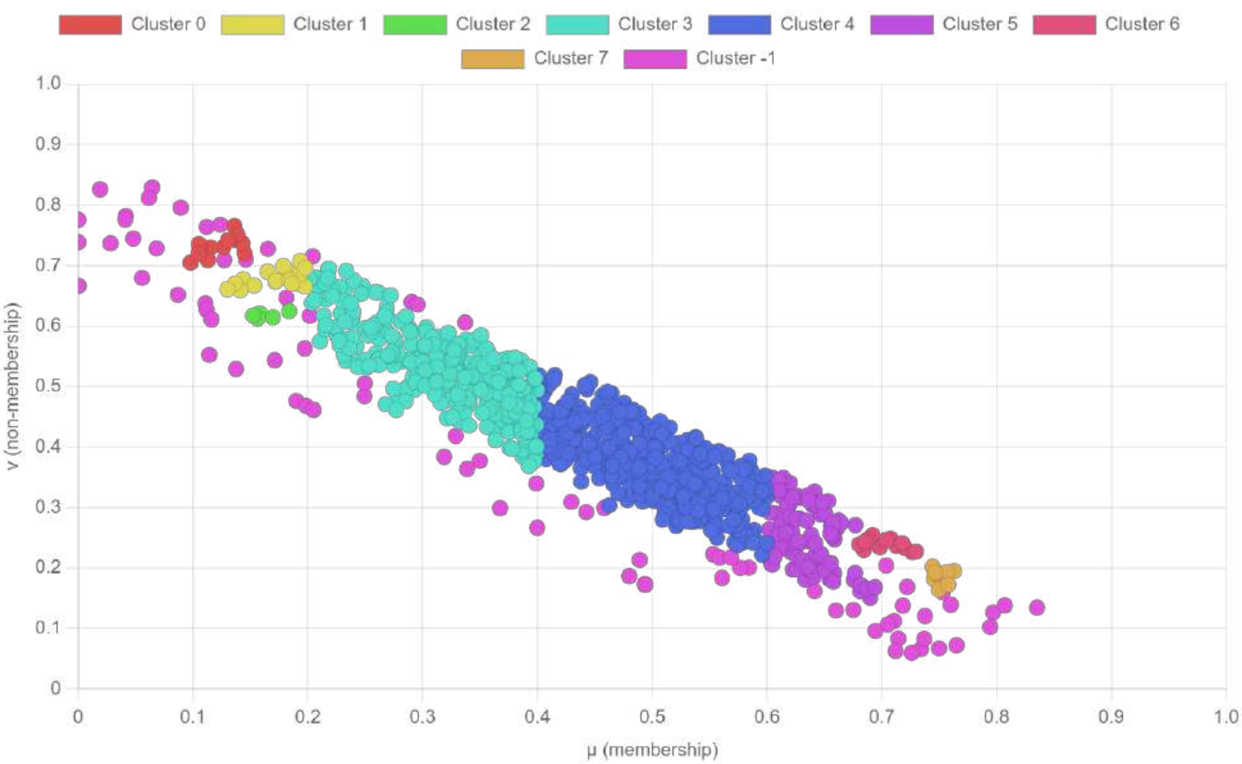

Fig. 14. *Result from DBSCAN at values eps = 0.02 and min_samples = 5 with segmentation by degree of membership*

DBSCAN is implemented through the scikit-learn library in Python, and the graphs are visualized by the client side through JavaScript and the chart.js library.

The client side visualizes the data using JavaScript and the Chart.js library. The graph is interactive and allows you to see which school corresponds to a given point, as well as to remove points from a given cluster. This creates an easy-to-use and user-friendly graph that allows for easier analysis of the clustering procedure.

Based on the results obtained from DBSCAN (Fig. 14), the following conclusions can be drawn for the schools in each of the clusters:

- Cluster 0 (top left, red) – schools in this cluster show low results with a high degree of confidence in the assessment (low uncertainty value). These are schools that have consistently low results and require action to be taken;
- Cluster 1 (top left, yellow) – schools that are slightly better than those in cluster 0, but have a moderate level of non-belonging and a fairly low level of belonging. These are

schools with low results but potential for improvement. Again, they require specific measures to improve results and the quality of the educational process.

- Clusters 2 and 3 (in the middle, green and cyan) – schools with good average results. Although they can be considered rather average, these are certainly schools that have the potential to improve their results. There are quite a few schools in these clusters;
- Cluster 4 (middle right, blue) – schools with good results that are above average but have potential for growth and improvement. With appropriate measures, these schools can become top schools;
- Cluster 5 (right side, pink) – schools with top results but mixed levels of uncertainty. This cluster includes schools with good results but often with a higher degree of uncertainty. These schools have good results, but appropriate measures for growth can also be identified;
- Cluster 6 (right, red) – Schools with top results and a lower degree of uncertainty. These are schools with consistently good results;
- Cluster 7 (right, orange) – Schools with top results and very low non-belonging values. These are elite schools with mainly excellent SME scores.
- Cluster -1 (noise) – These are schools that either have exceptionally high results or have a small number of students but lower results. Most of the schools classified as noise are among the most elite in the country. Further segmentation of schools classified as noise can provide additional and useful information.

The software developed within the scope of this dissertation is easy to use and does not require specific settings, resources, or installation. At the same time, it can be easily upgraded and expanded in terms of functionality, which makes it a great opportunity for implementation in the education system for the purposes of the Regional Education Administration, the Ministry of Education and Science, and other organizations involved in education management and analysis.

Contributions in the dissertation

The contributions in this dissertation are of a scientific, scientific-applied, and applied nature.

The scientific contributions are as follows:

- Development of an algorithm with intuitionistic fuzzy evaluations for data analysis and recommendations for student academic achievement, and development of an algorithm with intuitionistic fuzzy evaluations for data analysis from state matriculation exams in the secondary education system in Bulgaria;
- Defining intuitionistic fuzzy assessments in the process of sorting waste using a robotic arm and for assessing the frequency of consumer requests;
- Development of algorithms for solving the knapsack problem based on circular and elliptical intuitionistic fuzzy sets;

The scientific and applied contributions are as follows:

- A generalized net model has been created, reflecting the process of clustering large data sets using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. The basic concept of DBSCAN is based on recognizing areas with high object density, which are considered clusters, and separating objects with low density as noise or anomalies. Compared to other cluster analysis algorithms, DBSCAN does not require the number of clusters to be specified at the outset. Another valuable quality of the algorithm is that it provides better cluster distribution when the data is unevenly distributed.
- A generalized net model has been created, reflecting the clustering process using the BIRCH method—one of the most effective methods for processing large data sets. BIRCH is specifically designed to process large volumes of information, minimizing the need for data storage in memory. Its main idea is to build a compact summary of the data using a special hierarchical structure called a CF-tree (Clustering Feature Tree), which allows for incremental and dynamic processing of input objects.
- Implementation of a support vector method with intuitionistic fuzzy evaluations for recognizing scientific terms and determining keywords in a system for offering educational resources and scientific articles;
- Optimization solution for shipment delivery based on data analysis and application of the elliptical intuitionistic fuzzy knapsack decision-making problem;

The applied contribution consists of testing basic techniques in the field of data mining. The constructed models are illustrated with actual data in order to demonstrate and examine the practical part of working with the methods mentioned. For the implementation of the tests of the algorithms with intuitionistic fuzzy evaluations and the knapsack problem, products with a free distribution license were used.

Publications

1*. Petrov, P., Bureva, V., Kacprzyk, J. (2023). Generalized Net Model of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Algorithm and Its Application on Diabetes Dataset. In: Sotirov, S., Pencheva, T., Kacprzyk, J., Atanassov, K.T., Sotirova, E., Ribagin, S. (eds) Recent Contributions to Bioinformatics and Biomedical Sciences and Engineering. BioInfoMed 2022. Lecture Notes in Networks and Systems, vol 658. Springer, Cham., pp 264–271 https://doi.org/10.1007/978-3-031-31069-0_26

2*. Bureva, V., Petrov, P., Popov, S. (2022). Generalized Net Model of Balanced Iterative Reducing and Clustering Using Hierarchies (Birch) with Intuitionistic Fuzzy Evaluations. In: Kahraman, C., Tolga, A.C., Cevik Onar, S., Cebi, S., Oztaysi, B., Sari, I.U. (eds) Intelligent and Fuzzy Systems. INFUS 2022. Lecture Notes in Networks and Systems, vol 504. Springer, Cham., pp 673–680 https://doi.org/10.1007/978-3-031-09173-5_78

3*. Petrov, P., Bureva, V., Atanassov, K. (2023). Intuitionistic Fuzzy Evaluations of Garbage Sorting Using a Robotic Arm. In: Kahraman, C., Sari, I.U., Oztaysi, B., Cebi, S., Cevik Onar, S., Tolga, A.Ç. (eds) Intelligent and Fuzzy Systems. INFUS 2023. Lecture Notes in Networks and Systems, vol 758. Springer, Cham.,pp 259–264 https://doi.org/10.1007/978-3-031-39774-5_32

4*. Bureva, V., Petrov, P., Andonov, V., Atanassov, K. (2023). Intuitionistic Fuzzy Evaluation of User Requests Frequency. In: Atanassov, K.T., *et al.* Uncertainty and Imprecision in Decision Making and Decision Support - New Advances, Challenges, and Perspectives. IWIFSGN BOS/SOR 2022 2022. Lecture Notes in Networks and Systems, vol 793. Springer, Cham., pp 15–21 https://doi.org/10.1007/978-3-031-45069-3_2

5*. Petrov P. , Bureva V. Software System for Monitoring of Student Academic Performance in School using Intuitionistic Fuzzy Evaluations, FQAS: Conference on Flexible Query Answering Systems, 11 – 13 September 2025, Burgas, Bulgaria (in press)

6*. Traneva, V., Petrov, P., Tranev, S. (2023). Intuitionistic Fuzzy Knapsack Problem Through the Index Matrices Prism. In: Georgiev, I., Datcheva, M., Georgiev, K., Nikolov, G. (eds) Numerical Methods and Applications. NMA 2022. Lecture Notes in Computer Science, vol 13858. Springer, Cham. pp 314–326, https://doi.org/10.1007/978-3-031-32412-3_28

7*. Traneva, V., Petrov, P., Tranev, S. (2024). Circular Intuitionistic Fuzzy Knapsack Problem. In: Lirkov, I., Margenov, S. (eds) Large-Scale Scientific Computations. LSSC 2023. Lecture Notes in Computer Science, vol 13952. Springer, Cham. pp. 279- 287, https://doi.org/10.1007/978-3-031-56208-2_28

8*. Traneva, V., Petrov, P., Tranev, S.: An Elliptic Intuitionistic Fuzzy Portfolio Selection Problem based on Knapsack Problem. Annals of Computer science and information systems 37, 335- 342 (2023). doi: 10.15439/2023F4882

9*. Traneva, V., Petrov, P., Tranev, S. (2024). A Software Solution for Package Delivery Optimization in Courier Business Using an Elliptic Intuitionistic Fuzzy Knapsack

Problem Recent advances in computational optimization: results of the Workshop on Computational Optimization WCO 2023, 2025

10*. Petrov P. , Bureva V. From Data to Insight: An Intuitionistic Fuzzy Clustering Framework for Evaluating School Performance in Bulgaria, SISNEXUS 2025, Extended Abstracts of the International Conference “Science-Industry-Society Nexus for Sustainable Regions”, June 12-14 2025, Burgas, Bulgaria -> Bioautomation