



Даниела Ананиева Орозова

Приложение на науката за данните  
във виртуалното образователно пространство

АВТОРЕФЕРАТ

на дисертация  
за придобиване на научната степен „доктор на науките“

по професионално направление 4.6. Информатика и компютърни науки

София, 2023 г.

## **СЪДЪРЖАНИЕ**

Съответствия на използваните термини от английски език	2
<b>УВОД</b>	3
1. Науката за данните в обучението	3
2. Цел и задачи на дисертационния труд	4
<b>ГЛАВА 1. ВИРТУАЛНО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО</b>	5
1.1. Архитектура на данните на виртуално образователно пространство	5
1. 2. Виртуално образователно пространство в концепцията Интернет на нещата	5
1. 3. Виртуалното образователно и големите данни	6
1. 4. Виртуалното образователно пространство на нововъзникващите технологии	6
1. 5. Модели за анализ на предметната област	7
<b>ГЛАВА 2. СРЕДСТВА НА ВИРТУАЛНОТО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО</b>	8
<b>ГЛАВА 3. МОДЕЛИ ЗА АНАЛИЗ НА ДАННИ ВЪВ ВИРТУАЛНОТО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО</b>	9
3.1. Използване на електронни таблици за аналитични изследвания на данните	9
3.2. Аналитични изследвания на данните в обучаващото пространство с „Orange“	10
3.2.1. Проблеми при подготовка на данните за анализ	10
3.2.2. Оценяване и прогнозиране на знанията на обучаемите във виртуалното образователно пространство	11
3.3. Уеб метрики за оценяване дейностите на обучаемите във виртуалното образователно пространство	20
3.4. Моделиране на потребителя в образователното пространство	23
3.5. Когнитивен агент в среда за електронно обучение	27
3.6. Модели за йерархично многокомпонентно оценяване	29
<b>ГЛАВА 4. ОБОБЩЕНОМРЕЖОВИ МОДЕЛИ ВЪВ ВИРТУАЛНО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО</b>	34
4.1. Модел на процесите на персонализиране и използване на среда за обучение	34
4.2. Модел на процеса на прилагане на средства за извличане на знания от данните в среди за обучение чрез апарата на обобщените мрежи	35
4.3. Обобщеномрежов модел на процес за многокомпонентно оценяване	37
4.4. Обобщеномрежов модел на процесите в проектно-базирано обучение	38
4.5. Модел за игровизация на електронен курс за обучение	41
4.6. Модели за осигуряване на качеството и акредитация на висшето образование	43
4.6.1. Слабо-централизиран модел за акредитация	44
4.6.2. Силно-централизиран модел за акредитация	44
4.6.3. Обобщеномрежов модел на процесите на обработка на данни при самооценяване във висше училище	45
<b>ГЛАВА 5. РЕШЕНИЯ НА НАУЧНОПРИЛОЖНИ ПРОБЛЕМИ В ОБУЧЕНИЕТО</b>	46
5.1. Учебна програма за анализ на данните	46
5.2. Инструменти за извличане на знания чрез примери	47
5.3. Въвеждане на онтологии в обучението на студенти	48
<b>ЗАКЛЮЧЕНИЕ</b>	49
Приноси на дисертационния труд	50
Списък на публикациите	51
<b>БИБЛИОГРАФИЯ</b>	56

## Съответствия на използваните термини от английски език

Термин от английски език	Пояснение
Big Data	големи данни
Content Management System (CMS)	система за управление на съдържанието
Cyber-Physical Social System (CPSS)	кибер-физическа социална система
Cyber-Physical Social Educational System (CPSES)	кибер-физическа социална образователна система
Data Analytics	анализи на данни
Data Science	наука за данните
Data Mining	извличане на знания от данните
Data Cleaning	почистване на данните
Data Integration	интеграция на данните
Data Reduction	редуциране на данните
Data Transformation	трансформация на данните
E-Learning	електронно обучение
Educational Data Mining	извличане на образователни данни
Emerging technologies	нововъзникващи технологии
Executable models	изпълними модели
Fuzzy sets and systems	размити множества и системи
Generalized Net (GN)	обобщени мрежи (OM)
Internet of Things (IoT)	интернет на нещата
Knowledge based systems	системи, основани на знания
Knowledge representation	представяне на знания
Learning Analytics	учебни аналитики
Learning Management System (LMS)	система за управление на учебния процес
Learning System	учебна система
Machine Learning	машинно самообучение
Multi-agent systems	мултиагентни системи
Natural language processing	обработка на естествен език
Ontologies engineering	онтологично инженерство
Orange Data Mining system	софтуер за машинно обучение и анализ на данни
Outliers	отличителни стойности
Property-based models	модели, основани на свойства
Reinforcement Learning	подсилващо обучение
Recommendation systems	препоръчващи системи
Sharable Content Object Reference Model (SCORM)	спецификация за стандартизирано електронно учебно съдържание
Semantic Web	семантичен уеб
Supervised and Unsupervised learning	контролирано и неконтролирано обучение

## УВОД

### 1. Науката за данните в обучението

Виртуалните образователни пространства [3] са информационни и социални пространства, които интегрират хетерогенни технологии и различни педагогически подходи. Те са среда за доставка на учебни материали и образователни услуги за различни целеви групи, независимо от времето и пространството. Данните натрупвани от работата на образователното пространство постоянно се увеличават. Много страни и университети по целия свят изграждат инфраструктури за анализиране на тези данни. Siemens & Long [90] определят *Data analytics in Education* като "измерване, събиране, анализ и отчитане на данните за учащите се и техните контексти с цел разбиране и оптимизиране на обучението и околната среда, в която се случва." Chatti [23] определя целите на този анализ като: мониторинг, прогнозиране, индивидуализация, намеса в обучаващия процес, оценка и препоръки на обучаемия.

Термините *Big Data*, *Data Analytics* и *Data Mining* описват както самите данни, така и технологиите за събиране, обработка, управление на данните и методите за анализ [24]. *Data Mining* е процеса на търсене на скрити данни и закономерности, предварително неизвестни, нетривиални и практически полезни, необходими за взимане на решения в различни сфери на човешките дейности. *Big Data Analytics* се явява развитие на концепцията *Data Mining*. Също така е и развитие на решаваните задачи, сфери на приложение, източници на данни, методи и технологии на обработка. През 1998 г. Чикио Хаяши [44] въвежда термина *Data science* като нова, интердисциплинарна концепция, с три аспекта: проектиране на данни, събиране и анализ. *Data science* съчетава множество подходи и техники, свързани с анализ на данни, откриване на знания, машинно обучение, изкуствен интелект, програмиране, комуникация др. *Data science* е „сплав“ от различни дисциплини, технологии и средства за анализ на данни и се явява основно средство за откриване и оползотворяване на потенциала на Big Data [26].

Терминът „*Data analysis*“ се отнася до обработката на данни чрез конвенционални (класически статистически, емпирични или логически) теории, технологии и инструменти за извличане на полезна информация и за практически цели. Терминът „*Data analytics*“, от друга страна, се отнася до теориите, технологиите, инструментите и процесите, които позволяват задълбочено разбиране на изследваните данни [21].

Събирането на данни за анализ на обучението се отнася до целия процес и включва всички данни, получени по време на учебните дейности. Това са много и разнообразни типове данни. Международната организация по стандартизация, IMG Global Learning Consortium (IMG Global) класифицира данните, които могат да бъдат събрани и анализирани в областта на образованието в пет вида: данни за учебното съдържание; данни за учебната дейност; оперативни данни; данни, свързани с кариерното развитие; данни за профила на обучаемия. Това е предпоставка за появата на *Learning Analytics* [23], нов клон на анализ на данни, който включва начертани цели и методи от образователни и психологически изследвания.

През 2009 г. Avinash Kaushik публикува книгата "Уеб анализ 2.0", в която дефинира *Digital Analytics* като „непрекъснато анализиране на количествени и качествени данни от уеб пространството, което да подобри онлайн преживяването на потенциалните клиенти и така да доведе до желаните резултати“ [51]. Въведен е подход на разделяне на данните на различен принцип, което им придава специфично значение. Навлизането и популяризирането на социалните мрежи поставя началото и на съвременния социален анализ *Social Analytics*, определен като "наблюдение, анализ, измерване и тълкуване на цифровите взаимодействия и взаимоотношения на хора, теми, идеи и съдържание" [92]. Този анализ изследва ролята на социалните взаимодействия в процеса на учене и тяхното влияние върху напредъка на обучаемите.

Извличането на образователни данни, наричано още *Educational Data Mining*, е

дефинирано като „дисциплина, занимаваща се с разработването на методи за изследване на уникалните и все по-мощни данни, които идват от образователните среди, и използването на тези методи за по-добро разбиране на обучаемите и на средата, в която те учат“ [47]. Интегрирайки средствата за извличане на знания от данните в обучаващите среди, могат да се търсят тенденции в развитието на процесите на електронното обучение и неговото обслужване. Проследявайки процеса на избор и използване на различни средства могат да се изградят оптимални учебни среди с възможности за персонализирано овладяване на ключови знания, умения и компетенции. Могат да предвидят потенциалната загуба на обучаеми, да се определят кои дейности ще се предпочитат и каква ще е тяхната ефективност. Новите информационни технологии допринасят за изследване, анализиране и развиване на обучаващите среди като наблюдават и измерват различни аспекти на виртуалната и физическата среда, в която се провежда обучението.

## 2. Цел и задачи на дисертационния труд

**Целта** на изследването са теоретични обобщения на процесите на наблюдение и анализ на данните от динамичното взаимодействие на обекти в обучаваща среда и създаване на методи и модели за решаване на научни или научноприложни проблеми във виртуалното образователно пространство.

За постигане на целта са поставени следните четири **задачи**:

- Анализ на прилагането на средства за извличане на знания в обучаващите пространства и търсене на решения за персонализиране на електронното и дистанционно обучение.
- Създаване на методи за оценяване и прогнозиране на знанията, уменията и компетентностите на обучаемите във виртуалното образователно пространство.
- Създаване на модели, като резултат от теоретични обобщения на процесите на наблюдение и анализ на дейностите на обучаемите, свързани с Big Data Analytics, Data Mining, Web метрики, обобщени мрежи, машинно обучение и размита логика.
- Въвеждане на базови модули и инструменти на науката за данните за решения на научноприложни проблеми в обучението.

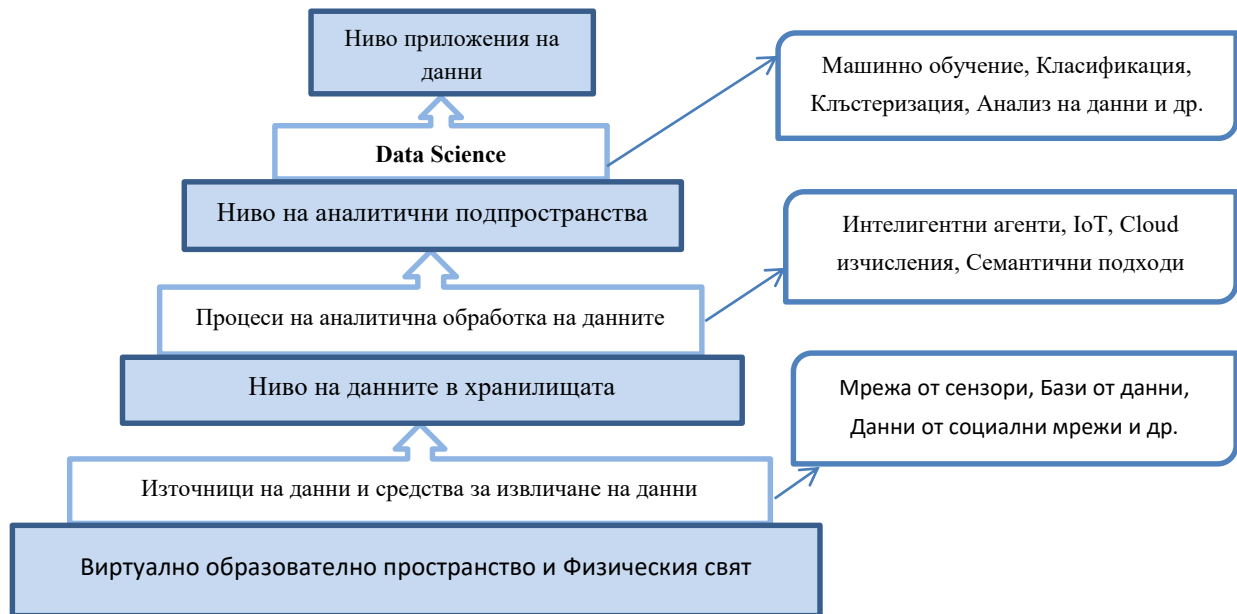
Дисертационният труд на тема „Приложение на науката за данните във виртуалното образователно пространство ” включва 49 фигури, библиографията обхваща 220 източника, от които 197 на английски език.

Текстът е организиран в увод, пет глави и заключение. В **увода** се прави преглед на основните понятия и концепции на изследваната област. Дефинира се целта на дисертационния труд и свързаните с нейното постигане задачи. В **първа глава** се разглежда областта на виртуалното образователно пространство, основните характеристики и възможности на такъв тип инфраструктура и мотивацията за проведената научноизследователска работа. **Втора глава** представя състоянието, предлагани решения, обща характеристика и възможности за интеграция на средства и инструменти към виртуалното образователно пространство. В **трета глава** са предложени методи и техники за моделиране на процесите в образователното пространство, с приложение на инструменти за извличане на данни в образователен контекст (Educational Data Mining). В **четвърта глава** се разглеждат работни процеси във виртуалното образователно пространство и техни формални модели чрез апарата на обобщените мрежи. **Пета глава** представя практики и тенденции при навлизането на науката за данните (DataScience) във висшето образование. **Заключението** обобщава резултатите и дава насоки за продължаване на изследванията по темата на дисертационния труд. Приложени са: Декларация за оригиналност на резултатите, Приноси на дисертационния труд, Списък на публикациите по темата на дисертационния труд.

# ГЛАВА 1. ВИРТУАЛНО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО

## 1.1. Архитектура на данните на виртуално образователно пространство

Разглеждайки архитектурата на виртуалното образователно пространство от гледна точка на данните, акцентът се поставя върху различните нива за представяне на данни и свързаните с тях механизми за събиране и анализ на данни, за извличане на информация и прогнозиране поведението на средата и обучаемите. На фигура 1 е направено обобщение в четири базови нива на представяне на данните, но могат да бъдат разглеждани и други нива в зависимост от прилаганите критерии за съхранение и достъп.



Фигура 1. Функционални нива във виртуалното образователно пространство

**Първо ниво:** Данни от виртуалното образователно пространство и физическия свят. Източниците обхващат широко поле, обединявайки данни получавани от сензорни датчици, данни от социални мрежи, база от данни за областта на обучението и много други.

**Второ ниво:** Хранилища на данните. Това ниво предоставя всички данни за състоянието на физическия свят, пренесени във виртуалната среда на пространството. Включва използването на технологии за събиране, съхраняване, обработка на данни и инструменти за анализ.

**Трето ниво:** Хранилища с аналитични подпространства на данните. Данните във виртуалните образователни пространства могат да бъдат разглеждани от различни гледни точки и подлагани на различни видове анализи.

**Четвърто ниво:** Приложният слой предоставя на потребителите набор от инструменти и интерфейси за управление, обработка и визуализиране на данни от предходни слоеве.

Новите технологии за работа с големи данни и техническата мощ на съвременните компютри създават възможност за иновативни приложения, базирани на интегриране на разнородни данни.

## 1.2. Виртуално образователно пространство в концепцията Интернет на нещата

Концепцията за Интернет на нещата (IoT) дефинира устройствата или обектите (нещата), съставляващи екосистема, която обитават и обогатяват чрез споделяне на своите функционалности, съдържание и знания. Електронното обучение може да бъде включено в условията на IoT. Всеки аспект на онлайн образованието е нещо или комбинация от неща в тази екосистема: учене, оценка, интелигентна образователна помощ и др. Навлизането на Интернет

повсеместно в нашия живот и постепенната му трансформация в Интернет на нещата е предпоставка за създаване на “кибер-пространство” [16]. В [14] се дефинира „кибер-физическа система“ (Cyber-Physical-System, CPS) като инженерна система, изградена от изчислителни и физически компоненти. "Физически" са елементите на системата, заемащи физическо пространство, "кибер" са изчислителните и комуникационни елементи на системата. Отчитането на социалната компонента в CPS я трансформира в кибер-физическа социална система (Cyber-Physical Social System, CPSS) [102]. Реализацията в сферата на образованието се явява като CPSES (Cyber-Physical Social Educational System) и представя подмножество на CPSS в домейна „образование и обучение“.

Могат да се посочат успешни резултати за интеграция на виртуалната среда с физическия свят. В лабораторията „Център за електронно обучение DeLC (Distributed eLearning Center)“ на Пловдивския университет се разработва кибер-физическа-социална система, която се изгражда като наследник на системата за електронно обучение DeLC [93]. Тя е наречена VES (Virtual eLearning Space) и се реализира чрез агентно-ориентиран подход. Изгражданото образователното пространство [95] се "населява" от активни автономни компоненти, наречени асистенти. Те са имплементирани като интелигентни софтуерни агенти, които поддържат планирането, организирането и изпълнението на учебния процес [86, 96]. Пространство интегрира виртуалния свят на електронното обучение с интелигентни устройства, сензори за събиране на данни от физическия свят, комуникационни инфраструктури и системна архитектура.

Повечето университети в страната реализират проекти за прилагане на информационни технологии и новаторски подходи, базирани на готови системи за управление на обучението. При тези среди липсва тясна интеграция на виртуалната среда с физическия свят на учебния процес.

В рамките на този дисертационен труд виртуалното образователно пространство се дефинира като среда, която интегрира различни информационни технологии и педагогически подходи за доставка на учебни материали и образователни услуги в абстрактен университет (без да реализира интеграция с физическия свят). Обхваща различни образователни инструменти и ресурси, които се използват в дигитална форма, за да се подобри образователния процес. Усилията са в посока разработване на методология за структуриране и моделиране на процесите на обучение, детайлизиране и формализация на процесите.

### **1. 3. Виртуалното образователно пространство и големите данни**

Виртуалното образователно пространство ежедневно събира огромно количество данни, което изисква нов подход в тяхната обработка. С термина "BigData" се определят усилията в организирането, съхранението, обработката и анализа на големи данни, които са толкова обемни и сложни, че е невъзможна тяхната обработка с традиционните приложения за работа с данни [12]. В [62] се предлага модел, описващ процесите свързани със съхранение, обработка и анализ на големи данни, които изискват нов поглед и съвместно прилагане на редица утвърдени технологии. В [63] се предлага модел на процесите, свързани с разпределени изчисления, базирайки се на Map/Reduce парадигмата, при използване на различни алгоритми и инструменти за реализация.

### **1. 4. Виртуалното образователно пространство на нововъзникващите технологии и рискове**

Водещите технологии се характеризират с радикална новост, бърз растеж и въздействие върху останалите технологии е прието да се наричат нововъзникващи технологии (*emerging technologies*) или подривни технологии (*disruptive technologies*) [19]. В американски източници тези технологии се определят и като *брилянтни технологии* [20], които се развиват във времето и имат потенциал да повишат влиянието си върху процесите на производство на знания. **Изкуственият интелект е „сърцето“** на нововъзникващите технологии, защото свързаните с него научни пробиви формират направления, чието функциониране зависи в най-голяма степен от

представянето на знанията и имитиране на способностите на човешките разсъждения.

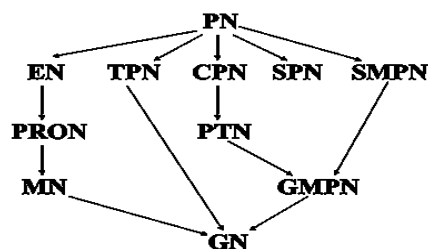
Цифровите технологии стават все по-сложни и интегрирани и като такива те предизвикват значителни трансформации в обществото и икономиката. Тази промяна има **положителни въздействия**, но трябва да отчитаме и **отрицателните въздействия** като например: загуба на работни места в традиционни професии; киберпрестъпност и хакерски атаки; грешки и неразбираемост; необходимост от нови концепции за отговорност, отчетност и управление; повишаване на неравенството между хората. Изложените на най-голям брой рискове са технологиите на „Изкуствен интелект и вземане на решения“, „Големи данни и решения“, „Икономика на споделянето“, „3D печат и производство, потребителски продукти и здравеопазване“, „Нашето цифрово присъствие“ и „Автономни автомобили“ [84]. За да продължи да се развива концепцията на виртуалното образователно пространство в тази динамична и хетерогенна среда, тя трябва да се адаптира към характеристиките и изискванията, които средата налага. Но всички промени трябва да са подчинени на политика и инвестиции за надежден изкуствен интелект [85] и да са основани на етичен и ориентиран към човека подход [31].

### 1. 5. Модели за анализ на предметната област

В литературата [4] формалните модели за анализ на програмно осигуряване се разделят на: модели, основани на свойства, изпълними модели и модели интегриращи двата подхода.

Към изпълнимите модели се разглеждат мрежите на Петри като просто и ясно средство за моделиране [76]. Те имат разширения, добавящи нови свойства и възможности за моделиране. Независимо от техните спецификации, всички тези инструменти имат позиции и преходи (дъги) и ядра, които се движат в мрежата. Основни разширения на мрежите на Петри [8, 9]:

- o EN – *Evaluation nets* добавят продължителност на движението на ядрото;
- o TPN – *Temporal PN* добавят момент за активиране на прехода;
- o CPN – *Color PN*. Всяко ядро има цвят и може да се движи само по дъги от същия цвят;
- o SPN – *Stochastic PN*. Изборът на дъга се основава на произволно генерирано число;
- o SMPN – *Self Modifying PN*. При самомодифициращите се мрежи на Петри, изборът на дъга се основава на генерирано двоично число 0 или 1, (като при 0 преходът не се извършва);
- o PRON – *Pro-Net* въвежда тип на прехода;
- o PTN – *Predicate/Transition Net*. Тези мрежи дефинират условие на прехода;
- o MN – *M Net*, при които ядрата изпълняват и допълнителни процедури;
- o GMPN – *Generalized Modifying PN* – ядрото се поглъща, когато предикатът генерира 0, в противен случай при генериране на 1 ядрото извършва прехода.
- o GN *Generalized Net* – обединяват различните разширения в единен формализъм за описание на паралелно протичащи процеси. Различните разширения на мрежите на Петри и връзката между тях се представя на фигура 2.



Фигура 2. Разширения на мрежите на Петри

В четвърта глава са представени формални модели на процеси и компоненти на виртуалното образователно пространство чрез апарата на обобщените мрежи. Те са обобщение на създадена серия от модели, свързани с наблюдение и анализ на дейностите на обучаемите в образователна среда, представени в публикациите [59, 62, 63, 68, 40, 41, 70, 71, 72, 74, 69, 97, 99].



## ГЛАВА 2. СРЕДСТВА НА ВИРТУАЛНОТО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО

**Виртуалното образователно пространство се разглежда като среда, която интегрира различни информационни технологии, педагогически подходи, образователни инструменти и ресурси, които се използват в дигитална форма, за доставка на учебни материали и образователни услуги в абстрактен университет.**

Науката за данните може да даде множество ползи в управлението на университет и да помогне за подобряване на ефективността и вземането на решения. Като се имат предвид конкретните нужди и цели на образователната институция могат да се изберат подходящи инструменти и методики за анализ на данните. Пространството се разглежда като активна система, отворена за компоненти с общ подход за достъп и използване на информацията [1, 55]. Такава система осигурява базови компоненти като:

- Средства за събиране, пречистване и ефективно съхранение на данните. Съхраняваната информация да бъде достъпна на различни нива на достъп и от различни приложения за извличане, показване и вземане на решения, както и контрол на достъпа до данните. Възможности за свързване на информацията, получена от различни източници.
- Средства за управление на знания с възможности за интелигентно търсене и четене, контрол за цялостност и коректност на данните, автоматизирано извличане на информация и др. Тук се включат инструменти за управление на качеството на данните, автоматизирано откриване на модели в данните и възможности за самообновяване на данните.
- Средства за оптимизиране на процеса на изпълнение на заявки за търсене и достъп до информацията. Средства за преобразуване и визуализиране на информацията, за по-лесно възприемане и използване от потребителя.
- Възможности за интегриране на компоненти с разнородни технологии и педагогически методи, посредством възприемане на общи архитектурни подходи за системата.

Освен базовите средства, начин за ангажиране на вниманието на обучаемите е да се интегрират в учебния процес онези средства, които те използват ежедневно в личното си виртуално пространство.

Интегрирането на средства за извличане на знания в средите за обучение може значително да подобри образователния процес като предостави ценна информация на обучаемите, преподавателите и администраторите, но от своя страна изисква ефективно управление на образователните данни. Това включва съхранение, обработка и защита на данните, като се осигурява тяхната сигурност и поверителност. Те трябва да бъдат използвани съобразно етични принципи и да зачитат правата на личността и поверителността.

В тази глава от дисертационния труд са обобщени резултати от проведени изследвания и анализи на основни характеристики на виртуалното образователно пространство на Университет, представени в публикациите, свързани с темата на дисертационния труд [61, 79, 82, 94, 96, 97]. Обосновани са предимства на стандартизираното електронно учебно съдържание при доставка на образователни услуги. Мотивирана е необходимостта от прилагане на средства за извличане на знания от данните в обучаващите пространства и търсене на решения за персонализиране на електронното и дистанционно обучение.

### ГЛАВА 3. МОДЕЛИ ЗА АНАЛИЗ НА ДАННИ ВЪВ ВИРТУАЛНОТО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО

В тази глава от дисертационния труд са обобщени резултатите от проведените експерименти и анализи на събраните данни за обучаемите в различни форми на електронно обучение, подробно представени в авторските публикации [75, 79, 80, 82]. В проведените изследвания се търси връзка между статистически методи, машинно обучение, откриване на поведенчески модели и анализ на данни. Разработените алгоритми и софтуер са представени в [57, 58, 83]. Предложени са модели, като резултат от теоретични обобщения на процесите на наблюдение и анализ на дейностите на обучаемите, свързани с Big Data Analytics, Data Mining, уеб метрики, машинно обучение и размита логика.

#### Анализи на данни във виртуално образователно пространство

За оценяване на знанията на обучаемите се използват различни форми и методи на изпитване като: online или offline тестове, отговори на отворени въпроси, тестване и разчитане на код, решаване на задачи и др. Крайната оценка по дисциплината често е комплексна и включва няколко компонента, участващи във формирането ѝ с различни тегла. Тя трябва да отразява различните страни от подготовката на обучавания (теоретични знания и практически умения), да е съобразена със спецификата на оценяването, с възрастови и индивидуални особености и др.

Експеримент за оценяване и аналитични изследвания на данните за обучаемите, е представен в [75]. За целите на анализа разглеждаме конкретен подход за оценяване на обучаемите, при който оценката в края на семестъра се определя като претеглена средна стойност (weighted mean  $\bar{x}$ ) от всички оценки през семестъра. Всяка оценка има важност или тегло (weight). Използваната в конкретния случай функция има вида:

$$(1) \quad \bar{x} = \frac{\sum xw}{\sum w}$$

където  $x$  е оценката,  $w$  е теглото на всяка оценка.

Оценката на студентите по предмета Математика в края на семестъра се определя въз основа на следните изпитвания: Тест 1 Множества (*Test 1 Sets*), Тест 2 Логика (*Test 2 Logic*), Тест 3 Геометрия (*Test 3 Geometry*), Тест 4 Статистика (*Test 4 Statistics*), Тест 5 Вероятности (*Test 5 Probability*), изпит в края на семестъра (*Final Exam*), контролни (*Quizzes*), домашна работа (*Homework*), проекти (*Projects*) и работа в часа (*Class Activities*). Всеки критерий има степен на важност и функцията има вида:

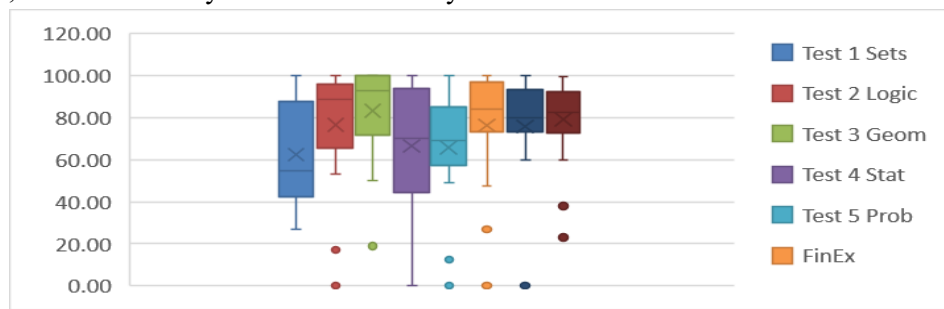
$$(2) \quad \bar{x} = \frac{\sum T1*0.09+T2*0.09+T3*0.09+T4*0.09+T5*0.09+FinEx*0.15+Act*0.40}{\sum w}$$

където  $\bar{x}$  е оценката в края на семестъра,  $w$  е теглото на всяка оценка. Разглеждаме как получената оценка по отделните тестове влияе на оценяването и как дейностите в часа и извън часа се отразяват на оценката в края на семестъра. Освен точките, натрупвани по отделните компоненти, съхраняваме и данни за всеки студент: Age (години е студена), Gender (пол), New Student Experience (NSE), FullTime-PartTime (кредитите за предмета са 12 или повече), Student Program (тип на обучението), Times Taken Course (за кой път студентът изучава предмета).

#### 3.1. Използване на електронни таблици за аналитични изследвания на данните

Електронните таблици (в частност Excel) предлагат ефективни функции за обобщаване на данните. Филтрите (*Filters*) дават възможности за подбор на записи. Условното форматиране (*Conditional formatting*) оцветява данните, които отговарят на определени критерии и ни помага да откриваме отклонения и тенденции в данните. Графиките (*Charts*) визуално показват отклонения и тенденции [30]. Използвайки инструментите на Excel се извърши първоначален анализ на резултатите от представянето на студентите по математика в продължение на пет семестъра. Графиката от фиг. 3 показва, добро усвояване на учебния материал. Най-ниски са оценките по Test

4 Statistics, където 75% от обучаемите имат оценки по-ниски от 50% от обучаемите по Test 1 Sets, Test 2 Logic, Test 3 Geometry и Test 5 Probability.



Фигура 3. Графика, относно разпределението на оценките по модули

В резултат на изследването бе направена препоръка за промяна на последователността в изучаването на две дисциплини. Промяната на стратегиите за преподаване оказва влияние върху начина, по който обучаемите усвояват материала.

### 3.2. Аналитични изследвания на данните в обучаващото пространство със системата „Orange“

Достъпни са различни инструменти за извличане на данни. Сред най-популярните са: RapidMiner, RapidAnalytics, WEKA, PSPP, KNIME, Orange, Apache Mahout, jHerWork, Rattle, GhostMiner, XENO, SAS Miner, Polyanalyst, IBM SPSS modeler. За целите на това изследване е разработен инструмент за идентифициране и прогнозиране на причините за изоставане или отпадане на обучаеми, представен в [78]. Използвана е системата “Orange Data Mining system” [56], софтуер с отворен код за машинно обучение и извличане на данни, написан на Python.

#### Контролирано обучение

Основната задача при контролираното обучение (*Supervised learning*) е да се създаде модел от обозначени (етикирани) данни, което позволява да се правят прогнози за бъдещи данни. Основни техники са: класификация, когато етикетите на класовете са дискретни и регресия, когато резултатът е непрекъснатата стойност. В системата *Orange Data Mining System* са създадени редица инструменти за изграждане на модели за класификация и регресия. Следва представяне на експерименти за оценяване и прогнозиране на знанията на обучаемите в образователно пространство с прилагане на инструментите: Logistic Regression, Naïve Bayes Classifier, Support Vector Machines (SVM), Decision Trees and Artificial Neural Networks и др.

#### 3.2.1. Проблеми при подготовка на данните за анализ и подходи за решаване

За анализ на данните в образователно пространство, основни входни данни са точките получени от обучаемите на различни оценъчни компоненти, а изходни са съответните крайни оценки, поставени от преподавател. Основни задачи за подготовка на данните са [22]: почистване на данните (Data Cleaning), интеграция на данните (Data Integration), трансформация (Data Transformation), редуциране на данните (Data Reduction). Почистването на данните е процесът за гарантиране, че данните са правилни, последователни и използваеми. *Дублиращи* се данни могат да възникнат при комбиниране на набори от данни от множество източници. *Неподходящи* данни са такива, които не отговарят на решавания проблем. *Структурни грешки* в данните могат да възникнат по време на измерване или прехвърляне на данни. *Outliers* (отличителни стойности) могат да причинят проблеми с някои видове модели. Когато не разполагаме с пълното множество от данни за дадена характеристика имаме *липсващи данни* (Missing Data). Процесът на подготовка на данни, включва идентифициране на грешки, коригиране, изтриване, подредба или друга обработка, като се запазят само потенциално полезните данни.

### 3.2.2 Оценяване и прогнозиране на знанията на обучаемите във виртуалното образователно пространство

Тук се предлага като **многостъпков процес**, метод за оценяване и прогнозиране на знанията, уменията и компетентностите на обучаемите във виртуалното образователно пространство [66]. **Методът включва пет стъпки** и може да бъде адаптиран към различни курсове на *Emerging technologies* във виртуалното образователно пространство.



Фигура 4. Многостъпков процес на анализ и прогнозиране на знанията на обучаемите

**Стъпка 1. Избор на начин за оценяване.** Определяне на ключови знания и компетентности за изучаваната технология. Определяне на степен на тежест (важност) на всяка анализирана компетентност.

- Основните **теоретични знания** се оценяват с помощта на компоненти като: *междинни тестове, решаване на задачи и казуси, изпити, обобщена дискусия* и др. Тези компоненти оценяват придобитите знания и способностите за разбиране на изучавания теоретичен материал. Използват се тестови въпроси, изброяване и сравняване на обекти, примери за понятия и използване на алгоритми, въпроси, свързани с обясняване и визуализиране на решения.

- Основните **компетентности** се оценяват от възможностите на студентите да прилагат придобитите знания за вземане на нестандартни решения в: *контролни и домашни работи, курсови задачи и проекти*, в които се решават нови задачи, извършва се критичен анализ на решения, определят се потенциални рискове, правят се самостоятелни изводи и заключения.

**Стъпка 2. Процес на оценяване.** По време на цялостния процес на обучение се провеждат и процедурите за оценяване. Резултатите се натрупват и съхраняват с цел профилиране на студентите, последващи анализи и прогнозиране в оценяването на нови студенти. В получените резултати се търсят зависимости в отделните оценъчни компоненти между теоретичните и практическите знания, умения и компетентности. В някои случаи студент получава висока оценка на тестовете и слаба оценка на контролната работа или обратно. Създаването на автоматизиран алгоритъм за оценяване при налични специфични стойности на оценъчните компоненти изисква нестандартно решение.

**Стъпка 3. Анализ на натрупаните данни от проведените обучения в реална среда.** Налични са различни алгоритми за машинно обучение. Основни входни данни са точките, получени от оценъчните компоненти, а изходни са съответните крайни оценки на студентите по дисциплината. При методите за обучение, част от примерните данни се ползват за обучение, а друга – за тестване. При недобри резултати от тестването, процесът на обучение може да се повтори или да се прецени, че избраният подход не е добър за решение на проблема.

**Стъпка 4. Създаване на модел за предвиждане.** На базата на натрупаните данни се създава класификатор, който извършва прогнози за текущи оценки по дисциплината. Системите за анализ на данни като: SPSS, Orange, Weka и други предлагат средства, на базата на алгоритми за: дърво на решенията, логистична регресия, теорема на Бейс, невронни мрежи и др. След приключване на обучението, моделите могат да се прилагат върху нови данни.

**Стъпка 5. Оценка на точността на предвиждане на алгоритмите.** След обучението се извършва тестване за точност и прецизност на работата на модела.

Доколко е добър един класификатор се определя от стойностите на съответни метрики за оценка на качеството. Не е достатъчно да се гледа само една от тях. Трябва да се вземат под внимание няколко метрики, като избора на това кои са по-важни е в зависимост от задачите и съответните цели.

*Confusion Matrix* е един от най-популярните начини за оценка на качеството на класификацията [89]. Представява матрица с размер  $N \times N$ , където  $N$  е броят класове на целевата променлива. Чрез прилагане на този инструмент могат да се съпоставят определените от класификатора и действителните стойности по 4 метрики: True Positive (*TP*), False Positive (*FP*), False Negative (*FN*) и True Negative (*TN*).

- *Общата точност* е показател, който дава информация за това каква част от всички случаи са правилно класифицирани:

$$(3) \quad Accuracy (ACC) = (TP+TN) / (TP+FP+FN+TN)$$

Общият брой правилно класифицирани обекти се разделя на всички случаи.

- *Общата грешка* показва каква част от всички обекти класификаторът е разпределил в грешните класове:

$$(4) \quad ERR = (FP+FN) / (TP+FP+FN+TN) = 1 - ACC$$

Може да се изчисли като общият брой грешно класифицирани обекти се раздели на всички случаи или ако от 1 се извади оценката *Accuracy*.

Нужно е да се вземат под внимание и метриците: *Precision*, *Recall*, *Specificity* и *F1-score*. Тези метрики са най-популярни и често използвани при оценка на класификатор:

- *Precision* - метриката показва каква част от обектите, класифицирани като положителни са в действителност положителни:

$$(5) \quad Precision = TP/(TP+FP)$$

- *Recall* или True Positive Rate (TPR) - дава информация за това каква част от положителния клас е била открита от класификатора:

$$(6) \quad Recall = TP/(TP+FN)$$

- *Specificity* или True Negative Rate (TNR) - показва каква част от отрицателния клас е била намерена от класификатора:

$$(7) \quad Specificity = TN/(TN+FP)$$

- *F1-score* - обобщава *Precision* и *Recall* в единствена стойност. Стойността на този показател е максимална, когато *Precision* и *Recall* са равни:

$$(8) \quad F1-score = 2*(Recall*Precision)/(Recall+ Precision)$$

Анализът на резултатите е свързан с решения от страна на ръководителите на обучаващия процес, как да се стимулират обучаемите и как да се помага на обучаемите в **риск**: допълнителни упражнения и задачи, допълнителна индивидуална и екипна работа, обяснения стъпка по стъпка, работа със софтуер като инструмент за моделиране и т.н.

**Приложение на метода за оценяване.** Изследвания и анализи на метода за оценяване и прогнозиране са извършени по дисциплината „Искусствен интелект“ с използване на инструментите на *Orange system* за експериментиране и изводи [66]. Представяната дисциплина (с преподаватели акад. Иван Попчев и проф. д-р Даниела Орозова) е част от виртуалното образователно пространство. Тя е структурирана в четири модула: Искусствен интелект – характеристики и проблеми; Търсене на решение в пространството на състоянията; Представяне на знания; Интелигентно вземане на решения. В средата за електронно обучение по дисциплината има предоставени материали с обем 587 МВ текстово описание, примери и 14 линка с полезни връзки. В преподаването се използват средите Moodle и Microsoft Teams. Тази дисциплина е

основа на курсовете: „Анализ и проектиране на бази от данни и знания“, „Управление на знания в компютърни системи“ и др.

**Съпка 1.** За оценяване на студентите по дисциплината “Изкуствен интелект” са определени оценъчни компоненти, всеки, от които инспектира теоретични знания, практически умения и компетентности с различно познавателно ниво. Дефинирани са 3 теста (3 x 5 = 15 точки), Контролна работа qwizz (15 точки), Проект - 45 точки и Обобщена дискусия - 25 точки.

В **проекта**, всеки студент сам избира тема от основните направления на изкуствения интелект, като например: Ontologies engineering, Semantic Web, Knowledge representation, Computational intelligence, Robotics, Natural language processing, Machine Learning, Deep learning, Soft computing, Pattern recognition, Multi-agent systems, Artificial neural networks, Genetic algorithms, Knowledge based systems, Decision support systems, Business intelligence, Data Science, Fuzzy sets and systems, E-learning и т.н. Проектът задължително включва: състояние, тенденции на развитие, идентификация, анализ и оценка на избрания инструментариум за въздействие върху потенциалните рискове, нерешени проблеми, изводи, заключение и библиография.

**Обобщената дискусия** със студента е върху темата на проекта, риск мениджмънт, мониторинг, контрол и оценка на риск мениджмънта на потенциалните рискове и възможностите за решаване на нови задачи с нестандартни решения.

Използваният примерен модел за оценяване по дисциплината е представен на фиг.5. Този модел на оценяване може да бъде динамично променян и адаптиран към конкретната дисциплина. Например в дисциплината „**Управление на знания в компютърни системи**“ се прилага алтернативен модел.

Оценъчен компонент	Означение	Оценъчна скала
Теоретични знания модул 2	Тест_1	до 5 точки
Теоретични знания модул 3	Тест_2	до 5 точки
Теоретични знания модул 4	Тест_3	до 5 точки
Практически компетентности	Контролна работа	до 15 точки
Практически знания и умения	Проект	до 45 точки
Теоретични знания и умения	Обобщена дискусия	до 25 точки
Крайна оценка	Окончателна оценка	Точките в оценка [2, 6]

Фигура 5. Примерен модел за оценяване по дисциплината

**Стъпка 2.** *Провеждане на обучение и процес на оценяване.* Задачата е свързана с намиране на общ подход за автоматизирано оценяване и прогнозиране на оценките на обучаемите. За намаляване субективността при оценяването на практическите умения се допуска оценяването да се извършва от външен оценител от фирми в областта на информационните технологии като: Технологика, Скейл Фокус и др. По всяка дисциплина студентите могат да получат до 100 точки и окончателната оценка се формира по следната скала: от 54 до 60 точки - *Среден* (3); от 61 до 70 точки - *Добър* (4); от 71 до 80 точки - *Много добър* (5); от 81 до 100 точки - *Отличен* (6).

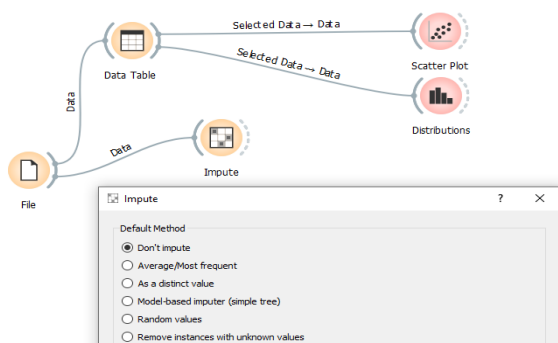
**Стъпка 3.** Анализ на натрупаните данни от проведените обучения в реална среда.

В процеса на работа са проведени множество експерименти. Основната цел е да се решава класификационен проблем като се установи дали е възможно да се предвиди оценката (изходна променлива) с помощта на входните променливи (точките от отделните оценъчни компоненти), които се запазват в модела. За решаване на класификационния проблем се прилагат различни техники, чрез инструментите на *Orange Data Mining System*.

В началото се създава работен процес и чрез инструмента „*File*” **зареждаме данните** за оценките на студентите по различните оценъчни компоненти в брой точки. Те могат да бъдат въведени от Excel (.xlsx), от текстов файл с раздели (.txt), файл с данни разделени със запетая (.csv)

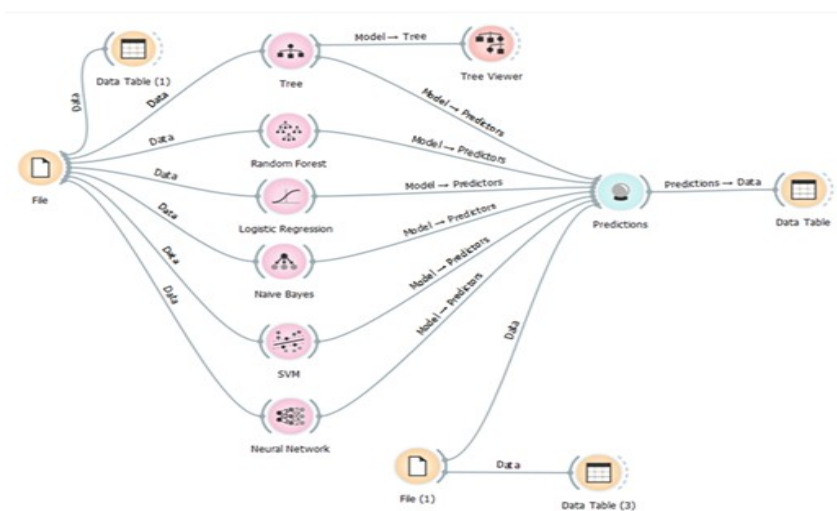
или URL адрес. За по-доброто разбиране, данните могат да се визуализират по някои колони или извадки от тях. Например, може да се свърже файла с данни с инструмента „Scatter Plot“ и се избераат колоните, чиито стойности ще се изобразяват по осите X и Y, използвани цветове, форми, размери и други параметри. Друг популярен инструмент за визуализиране на данните е „Distribution“, чрез който може да се покаже разпределение в набора от данни по даден атрибут.

В зависимост от целта и вида на данните избираме конкретен инструмент за регресия или класификация и **задаваме целевата променлива**. Следва **почистване на данните**. Могат да се приложат различни подходи: изтриване на липсващите стойности или да се заместят по подходящ начин. Чрез инструмент „Impute“ (фиг. 6), се избира измежду различни методи за заместване. По подразбиране е опцията „Remove the rows with missing values“. Други възможни опции са: *Distinct Value, Random Values, Model-Based*.



Фигура 6. Прилагане на инструменти за зареждане и почистване на данните

**Стъпка 4.** Създаване и трениране на модели за предвиждане. В експеримента последователно се прилагат инструментите на системата: **Tree, Random Forest, Logistic Regression, Naïve Bayes, Support Vector Machines (SVM) и Neural Network**. Работния поток за създаване и трениране на моделите е показан на фиг. 7.



Фигура 7. Работен поток за трениране на моделите и предвиждане

**Предвиждане на нови данни.** Чрез обучените алгоритми се очаква при задаване на нова комбинация от стойности за избраните компоненти да се определя изходната променлива. На тази стъпка считаме, че моделът е готов за практическо приложение. Моделът придобива независимост и прави свои заключения въз основа на набори от данни и обучение. На фигура 7 е представен работния поток за предвиждане чрез инструмента *Prediction* на системата Orange. Новите данни се подават чрез файл *Test.xlsx*, който има същата структура като началната таблица с данни, но колоната с изходната променлива не е зададена. Общ вид на получавания резултат от



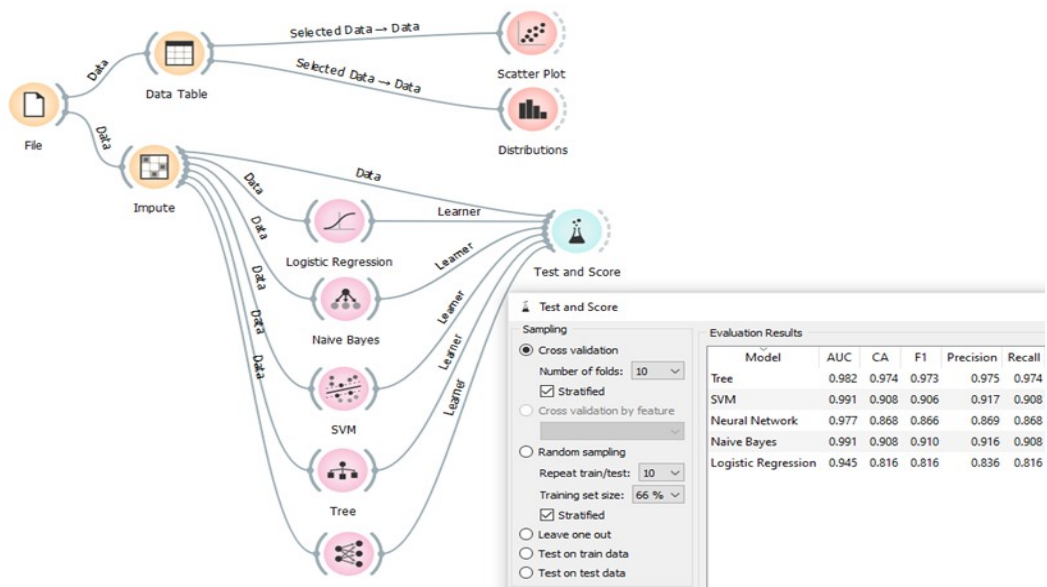
предвиждането е показано на фигура 8. Това са резултатите от предвижданията за крайната оценка на обучаемите, получени от различните модели.

Data & Predictions					
Tree	Logistic Regression	Random Forest	SVM	Naive Bayes	
1 0.00 : 0.67 : 0.33 : 0.00 : 0.00 → Excellent	0.00 : 0.49 : 0.02 : 0.06 : 0.43 → Excellent	0.00 : 0.57 : 0.11 : 0.00 : 0.32 → Excellent	0.01 : 0.18 : 0.03 : 0.01 : 0.77 → Very Good	0.00 : 0.85 : 0.01 :	
2 0.25 : 0.00 : 0.50 : 0.25 : 0.00 → Good	0.01 : 0.00 : 0.45 : 0.22 : 0.32 → Good	0.20 : 0.10 : 0.20 : 0.50 : 0.00 → Middle	0.22 : 0.02 : 0.23 : 0.51 : 0.03 → Middle	0.39 : 0.00 : 0.04 :	
3 0.25 : 0.00 : 0.50 : 0.25 : 0.00 → Good	0.00 : 0.00 : 0.20 : 0.30 : 0.50 → Very Good	0.17 : 0.00 : 0.76 : 0.05 : 0.02 → Good	0.03 : 0.05 : 0.42 : 0.08 : 0.42 → Very Good	0.57 : 0.00 : 0.22 :	
4 0.00 : 0.67 : 0.33 : 0.00 : 0.00 → Excellent	0.00 : 0.72 : 0.03 : 0.04 : 0.22 → Excellent	0.00 : 0.60 : 0.20 : 0.00 : 0.19 → Excellent	0.01 : 0.43 : 0.03 : 0.01 : 0.52 → Very Good	0.00 : 0.25 : 0.04 :	
5 0.00 : 0.67 : 0.33 : 0.00 : 0.00 → Excellent	0.00 : 0.03 : 0.31 : 0.37 : 0.29 → Middle	0.00 : 0.20 : 0.44 : 0.02 : 0.33 → Good	0.03 : 0.06 : 0.58 : 0.08 : 0.24 → Good	0.01 : 0.04 : 0.86 :	
6 1.00 : 0.00 : 0.00 : 0.00 : 0.00 → Bad	0.64 : 0.00 : 0.07 : 0.28 : 0.01 → Bad	0.92 : 0.00 : 0.00 : 0.08 : 0.00 → Bad	0.87 : 0.03 : 0.02 : 0.06 : 0.02 → Bad	0.86 : 0.00 : 0.00 :	
7 0.00 : 0.00 : 0.00 : 0.00 : 1.00 → Very Good	0.00 : 0.00 : 0.00 : 0.20 : 0.79 → Very Good	0.10 : 0.15 : 0.00 : 0.00 : 0.75 → Very Good	0.02 : 0.20 : 0.04 : 0.03 : 0.71 → Very Good	0.01 : 0.60 : 0.00 :	
8 0.00 : 1.00 : 0.00 : 0.00 : 0.00 → Excellent	0.00 : 0.80 : 0.01 : 0.01 : 0.18 → Excellent	0.00 : 1.00 : 0.00 : 0.00 : 0.00 → Excellent	0.02 : 0.87 : 0.03 : 0.03 : 0.06 → Excellent	0.00 : 1.00 : 0.00 :	
9 1.00 : 0.00 : 0.00 : 0.00 : 0.00 → Bad	0.11 : 0.00 : 0.01 : 0.08 : 0.79 → Very Good	0.47 : 0.25 : 0.14 : 0.13 : 0.00 → Bad	0.29 : 0.09 : 0.27 : 0.25 : 0.10 → Good	0.46 : 0.16 : 0.33 :	

Фигура 8. Резултати от предвиждането чрез различни модели

### Стъпка 5. Оценяване на работата на моделите, върху данните.

В работния поток се свързва всеки от създадените модели с “*Test and Score*” инструмент. След като се оценят моделите, трябва да се види дали тяхната точност може да бъде подобрена, чрез настройка на параметрите, присъстващи във модела.



Фигура 9: Работен поток за оценяване с инструмент “*Test and Score*”

Резултат от работата на инструмент “*Test and Score*” е таблица с оценки за: *Accuracy*, *Precision*, *Recall* и *F1 Score* за създадените модели. Конкретни оценки за метриците за оценка на качеството за създадените модели в експеримента са дадени на фиг. 10.

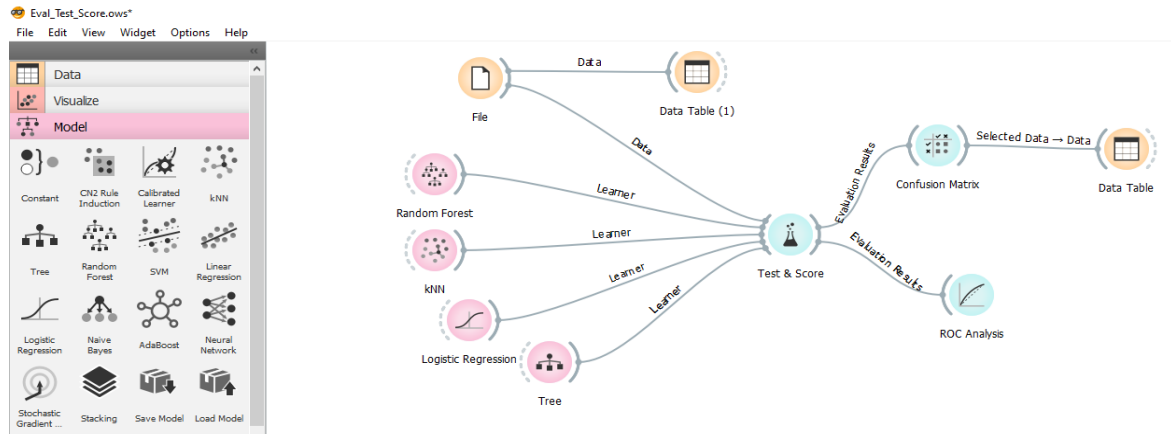
Scores					
Method	AUC	CA	F1	Precision	Recall
Tree	0.965	0.855	0.884	0.864	0.905
Logistic Regression	1.000	0.776	0.976	1.000	0.952
Random Forest	1.000	0.895	0.930	0.909	0.952
SVM	0.975	0.842	0.900	0.947	0.857
Naive Bayes	0.994	0.816	0.952	0.952	0.952
Neural Network	0.941	0.803	0.864	0.826	0.905

Фигура 10. Резултат от работата на инструмента *Test&Score* върху създадените модели

Разглеждайки прогнозната точност за всеки един от класовете оценки, може да се обобща, че тя е най-лоша за класа оценки *Middle* (оценка Среден). Най-висока точност се постига за класовете *Bad* (Слаб) и *Excellent* (Отличен). Прогнозите за оценка *Good* (Добър) и *Very Good* (Много добър) при всички разгледани модели се представят с точност около 60-75%. Моделът на *Random Forest* е най-надежден, защото се представя с най-висока точност за всички класове оценки. Моделът *Naive Bayes* е с най-ниска оценка за точността спрямо другите модели при разглежданите данни.

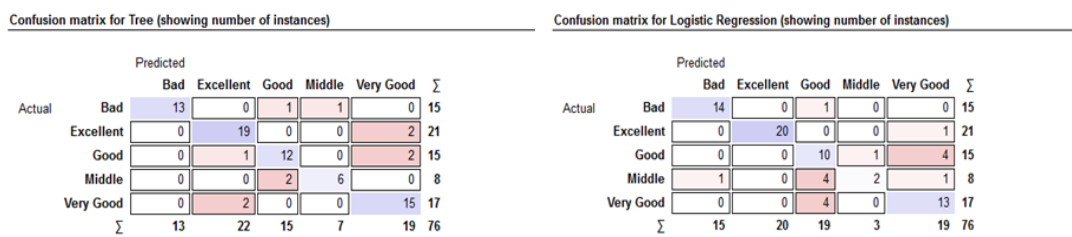


Тестване на различните модели и визуализация на точността на работата на моделите може да се извърши чрез „Confusion Matrix”, ROC Analysis или друг инструмент на системата. Работният поток е показан на фиг.11.



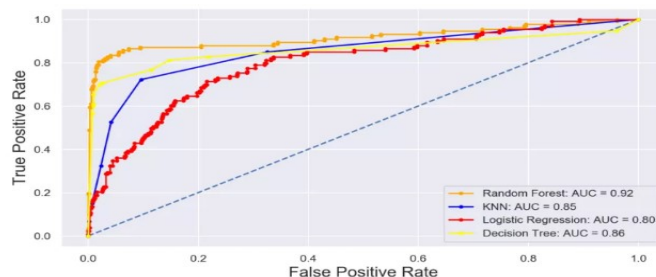
Фигура 11. Работен поток за създаване на модели и оценка с чрез Confusion Matrix и ROC Analysis

Резултати от работата на инструмента Confusion matrix, са дадени на фиг. 12, за сравнение на ефективността на създадените модели Tree и Logistic Regression.



Фигура 12. Резултати от работата на инструмента Confusion matrix за създадени два модели

Receiver Operating Characteristic (ROC) кривата съпоставя TPR (True Positive Rate) и FPR (False Positive Rate), което дава информация до каква степен моделът правилно разпознава съответните класове. На фиг. 13 е даден резултатът от съпоставяне на четири различни класификатори. Колкото по-близо до горния ляв ъгъл е ROC кривата, толкова по-високо е качеството на класификатора. На графиката се вижда, че в конкретния случай за предвиждане, моделът, използващ алгоритъм Random Forest се справя най-добре.



Фигура 13. Визуализация с инструмента ROC Analysis в Orange

Като се обобщят резултатите за средната оценка за прогнозната точност върху класовете оценки (Слаб, Среден, Добър, Много добър и Отличен) се получава: Decision Tree - 85.5%; Logistic Regression - 77.6%; Random Forest - 89.5%; Support Vector Machine - 84.2%; Naive Bayes - 81.6%; Neural Network 80.3.1%. Като подходящи за използване, при подобни задачи за оценяване, могат да бъдат класификационните алгоритми Decision Tree и Random Forest, които предвиждат с висока степен на точност елементите на Bad класа. Това са студентите със слаби оценки, които разглеждаме като „студенти в риск“.

В проведения експеримент инструментът обработва данните, които се събират по време на обучението на студентите в техните курсове. Освен това се използват и данни от анкетно проучване. Анкетата се изпраща по имейл през третата седмица от курса. В основата на анкетата са въпроси, с които студентите дават мнение относно задачите, материалите и нивото на трудност на учебния предмет. Данните, получени от анкетите, добавят редица нови характеристики, които са пряко свързани с отпадането на обучаемите, като липса на интерес, липса на време, организационни пречки по време на обучението и др. Първите експерименти са направени от малка извадка от данни, ограничена от първоначално налични реални данни за студенти. С помощта на разработената система за наблюдение и събиране на данни за обучаемите и тяхната дейност (решени задачи, преглеждане на електронни текстови и видео материали, участие във форум и/или групи по интереси и др.) от 4580 записа за обучаеми, 15 резултата от обучаемите са идентифицирани в риск. Това е многостъпкова процедура, която е пряко свързана с Big Data Analytics [101] в пространството за електронно обучение. Ежедневно се събира огромно количество данни, което сега изисква нов подход към обработката на данни, прилагайки рамка за разпределена обработка с отворен код Hadoop, базирана на алгоритъм Map/Reduce.

**ИЗВОД:** Проведените емпирични изследвания потвърждават, че предложеният **петстъпков метод за оценяване е перспективен** за разработване на система за ранно предупреждение за различни заинтересовани страни от обучаващия процес.

Базирайки се на натрупани данни от работата на система за електронно обучение с различни потребители, прилагайки средства от областта на науката на данните, могат да се взимат различни решения относно обучението. Анализът на данните, натрупвани в курсовете за електронно обучение дава възможност да се промени модела на изпитване и да се проектират модули, които да отговарят на индивидуалните нужди на обучаемия.

### **Подсилващо обучение (обучение с утвърждение)**

Към алгоритмите за машинно обучение принадлежи и подсилващото обучение (*Reinforcement Learning*). Характерното за този тип алгоритми е, че имитира психологически модел, при който на системата се подават „награждаващи“ и „наказващи“ сигнали с цел да се максимизира вероятността за получаване на „награда“ и да се минимизират „наказанията“ [88]. Това обучение се прилага обикновено при липса на предварително зададен „правилен“ набор от обучителни данни. Такъв подход е различен от начина на действие при контролирано обучение, където целта е намаляване на отклонението, съобразно предварително зададени правилни данни (вход/изход).

### **Неконтролирано обучение**

Алгоритмите от този тип приемат съвкупност от данни, съдържаща само входящи стойности и намират структура или разпределение на данните, без указания за известна променлива или функция за възнаграждане, данните не са етикетирани, няма примери за обучение [100]. Основни типове решавани задачи са: Dimensionality reduction, Density estimation, Clustering.

Клъстерният анализ представлява разпределяне на съвкупност от наблюдения в подмножества (clusters), така че наблюденията в един и същ клъстер да са сходни според един или няколко предварително зададени критерии, а наблюденията от различни клъстери да са различни [46]. Различните техники за клъстериране работят с различни предположения за структурата на данните, които често се дефинират чрез някаква метрика за сходство и се оценяват, например, по вътрешна компактност, или близост между членовете на един и същ клъстер, и разлика между клъстерите.

Разработените методи за групиране използват различен индукционен принцип. Farley and Raftery (1998) [32] предлагат разделянето на методите на групиране на две основни групи: йерархични и разделящи методи. Han and Kamber (2001) [45] предлагат категоризиране на методите в допълнителни три основни категории: базирани на плътност методи, базирани на модели и мрежови методи. Алтернативна категоризация, основана на принципа на индукция на различни методи за групиране е представена в Estivill-Castro (2000) [29].

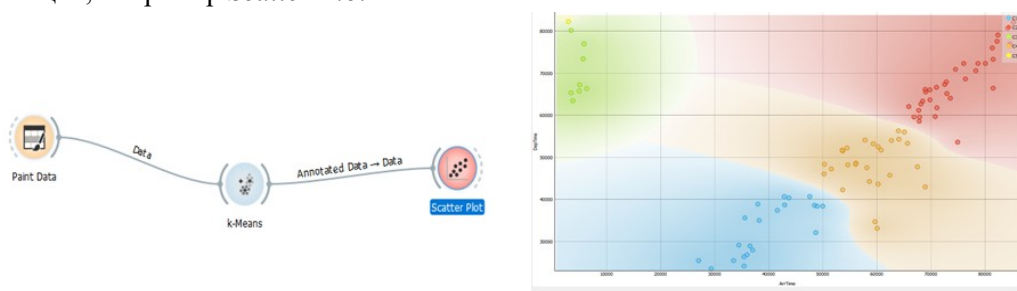
- Алгоритми за разделяне на клъстери

Тези алгоритми генерират различни дялове и след това ги оценяват по някои критерии. Те са наричани нейерархични, тъй като всеки екземпляр е поставен в точно един от  $k$  взаимно изключващи се клъстери. Изисква се предварително да се въведе желания брой клъстери.

Един от най-често използваните алгоритми за разделяне на клъстери е *k-means clustering* алгоритми. Този тип алгоритми принадлежат към *Exclusive Clustering*, защото данните са групирани, така че ако определена данни принадлежи към даден клъстер, тя не може да бъде включена в друг клъстер. Друг тип *Overlapping Clustering*, използва размити множества за групиране на данни, така че всяка точка може да принадлежи на два или повече клъстера с различна степен на членство.

Тристъпков алгоритъм: определя централната координата (центроид); определя разстоянието от всеки обект до центъра; групира обектите на базата на най-малкото разстояние. Прекратява се, когато центроидите спрат да се движат или някакъв праг е достигнат (например брой итерации). Следва примерен работен поток за прилагане на инструмент *k-means* върху набор от данни и визуализиране на резултата чрез инструмент *Scatter Plot* (фиг. 14).

За изграждане на информационния поток първо се извършва **събиране на данни**. Зарежда се работния плот и набора с инструменти и чрез инструмент „File” се зареждат данните за анализ. Може да бъде приложен и инструмент *Impute*” за почистване на данните. Следва **Избор на инструмент *k-means***. Този инструмент на системата *Orange* реализира *k-means clustering* алгоритъм. Последната стъпка е **визуализация на резултата**. Самият инструмент *k-means* не визуализира резултат, за тази цел трябва към информационния поток да бъде свързан инструмент за визуализация, например *Scatter Plot*.



Фигура 14. Работен поток с прилагане на инструменти *k-means* и *Scatter Plot*

- Йерархичната клъстериация

Когато броя на клъстерите не е предварително определен, се използват йерархични клъстеризационни процедури (*Hierarchical Cluster*). Тези алгоритми започват от деклариране на всяка точка за свой собствен клъстер и след това обединяват два най-подобни клъстера, докато не бъде удовлетворен критерий за спиране.

Голямото разнообразие на процедурите се поражда от използваната **метрика** между различните обекти. Йерархичната клъстериация изчислява йерархично групиране на произволни типове обекти от матрицата на разстоянията между тях и показва съответната *дендрограма*.

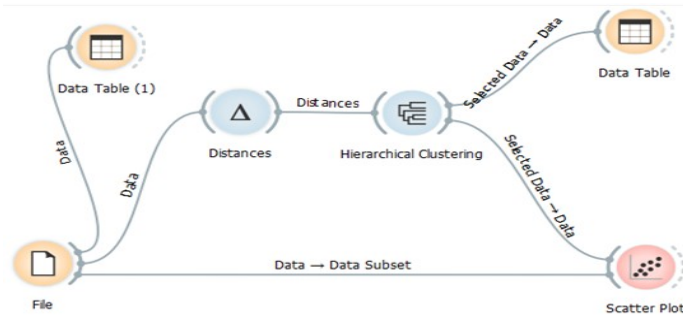
Метод за изграждане на информационен поток при йерархична клъстеризация е свързан с прилагане на следните *стъпки*:

1. **Събиране на данни.** Зарежда се работния плот и набора с инструменти и се зареждат данните за анализ. Може да бъде приложен и инструмент за почистване на данните.

2. **Избор на инструмент *Distance*** за изчисляване на разстоянието между елементите. За групиране на данните могат да се използват различни геометрични показатели за изчисляване на разстояние. *Euclidean metric* е мярка за разстоянието между точките, изобразени на евклидовата равнина. *Manhattan metric* е мярка, където разстоянието се изчислява като сума от абсолютната стойност на разликите между две точки, изобразени на декартовата координатна система. *Minkowski distance metric* е обобщение на показателите от *Euclidean metric* и *Manhattan metric*.

3. **Избор на инструмент *Hierarchical Clustering***. Този инструмент реализира алгоритъм за йерархичното групиране на произволни типове обекти чрез изчислените разстояния и показва съответната дендрограма. *Дендрограмата* е граф-дърво, в което всеки възел отразява една стъпка от процеса на клъстеризиране.

4. **Визуализация на резултата.** Чрез различни инструменти може да се визуализира резултата по различни начини, например чрез инструментите *Data Table* и *Scatter Plot*.



Фигура 15. Работен поток за йерархично клъстеризиране

Други популярни софтуерни средствата за извличане на знания, които могат да бъдат използвани за анализ на данни са например: WEKA [103], RapidMiner [87]. KNIME [52], KEEL (<http://sci2s.ugr.es/keel/>), SPSS <http://www.ibm.com/analytics/us/en/technology/spss/> и др.

### 3.3. Уеб метрики за оценяване дейностите на обучаеми във виртуално образователно пространство

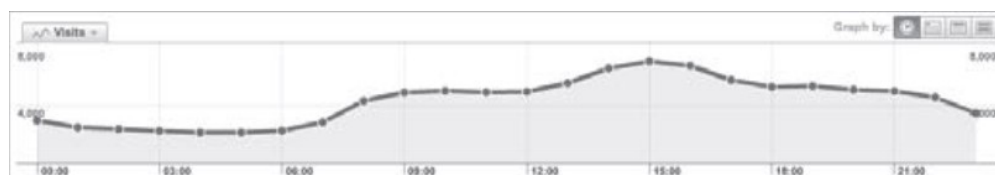
Анализът на данните, натрупани по време на обучението позволява проучвания и обратна връзка за това как обучаемите търсят информация, какви затруднения срещат и проектиране на модули, които отговарят на индивидуалните нужди на обучавания. На базата на веб метрики тук се предлага метод за оценка на степента на използване на веб сайтове, свързани с областта на обучението при моделиране на обучаемите, базирайки се на анализ на поведението им в обучаващата среда и веб пространството. Анализът на данните дава възможност да се проектират модули, които да отговарят на индивидуалните нужди на обучаемия за търсене на информация.

Уеб анализът е измерване, събиране, анализ и отчитане на интернет данни с цел по-добро разбиране на сложните взаимодействия между потребителите на уебсайтове. Процесът на уеб анализ включва различни уеб метрики, дефинирани в Google Analytics [25] като:

- **page views** - броя показвания на уеб страницата, достъпна от посетител (без паяци или роботи);
- **visitors** - броя на уникалните посетители на уеб сайта;
- **pages/visits** - броя на страниците, които са разгледани от посетител по време на посещение;
- **time on site** - продължителност на времето, прекарано от всички посетители на уеб сайта;
- **stickiness** - способността на уеб страницата да задържи посетителя на уеб сайта;
- **frequency** – (честота) брой посещения, от посетител на сайта (индикатор за лоялност);
- **recency** - брой дни, изминали от последното посещение на посетителя на сайта;
- **length of visit** - времето на посещение, прекарано от посетител на уебсайта (в секунди);
- **depth of visit** - брой страници, посетени от посетител за едно посещение и др.

В областта на уеб анализите се използват главно две техники за анализ на трафика на уеб сайта: събиране на данни от страна на сървъра и от клиентска страна. Методите за събиране на данни от сървъра извличат и анализират данни основно от регистрационните файлове (*Log files*) и включват информация за регистрационни файлове като IP адрес, час и дата, тип браузър и др. Информацията за поведението включва обща информация за сърфиране, като например брой прегледани страници, езикова настройка и др. Когато се събират данни от клиентски сайтове или страници с отметки, данните за посетителите на страницата се изпращат до проследяващ сървър с помощта на JavaScript код (или етикет), вмъкнат в HTML страницата. С този подход могат да се проследяват точно всички действия на посетителя, както и да се събира допълнителна информация. Бисквитите могат да се използват, за да се определи колко посетители за първи път или колко повторни посетители е получил даден сайт, колко пъти посетител се връща за всеки период и колко време минава между посещенията.

Google, WebTrends, Nedstat [5] и много други компании предоставят софтуер за уеб анализ с помощта на маркиране на страници. Google Analytics е най-използваната безплатна програма [25]. От гледна точка на анализа на дейността на обучаемите, особено интересна е свързаната с клиентите информация, историята на посещенията и поведението им, потребителския профил и др. Google Analytics изготвя анонимни и статистически отчети за уебсайтовете, които го използват. Показват данни като географско местоположение (на базата на общи кодове за геолокация, базирани на IP), време на посещение и т.н. Например: отчетът на фиг. 16 [25] дава възможност да се проследи по часове посетителския трафик на сайта.



Фигура 16. График за почасова тенденция с Google Analytics.

За да се направи анализ на най-посещаваните от обучаемия уеб сайтове, дълбочината и честотата на посещенията, се използва *Fuzzy Classification of the Web Metric* [107]. Известно е, че в теорията за размитите множества и размитата логика, се отчитат неточност, несигурност и неяснотата на човешкото мислене и език, определяйки функцията на принадлежност.

Размито множество  $A$  в  $X$  се дефинира като [105]:

$$(9) \quad A = \{(x, \mu_A(x))\}, \text{ където } x \in X, \mu_A : X \rightarrow [0, 1]$$

е функцията на принадлежност на  $A$  и  $\mu_A(x) \in [0, 1]$  е степента на принадлежност на елемента  $x$  в множество  $A$ . В нашите експерименти за броя на посещенията на даден уеб сайт, се въвеждат термините „low“, „medium“ и „high“ като езикови променливи. В общия случай, отчитайки стойностите за показване на страниците като true (1) или false (0) за месец, страниците се определят като „ниско посещавани“, ако посещенията са например между 0 и 25, между 25 до 55 посещения страниците са „средно посещавани“ и повече от 56 посещения са класифицирани като „високо показвани“. Въпреки това, ако посетител има 55 посещения на страница, той се определя в клас „средно“ активен, докато друг посетител с 58 посещения се определя като „много“ активен посетител. Въпреки разлика само от само от 3 посещения, те са разпределени в две различни групи. Чрез дефиниране на размити множества, представени с функция за принадлежност се въвежда непрекъснат преход между класовете „low“, „medium“ и „high“. Така посетител може да принадлежи отчасти на два класа (55% на „високо активен“ и 45% на „средно активен“) едновременно. Използването на размити класове позволява по-прецизна класификация на уеб метричните стойности [50]. Отчитайки това с цел а да се оцени трафикът на уебсайта, се разработва размита система от правила [28]. Измерването на броя на посещения на страница е число, което само по себе си няма голямо значение. Само контекстът на числото спрямо броя на прегледите на други страници предоставя знание, на базата на което можем да направим оценка.

За определяне на броя на посещенията на уеб страници се реализират правила за размита класификация. Прилага се метода на *индуктивна размита класификация (Inductive Fuzzy Classification IFC)*, при който групирането на елементите в размит набор, се извършва с функция за членство, извеждана чрез индукция от данни. Индуктивната размита класификация по процентилен ранг (IFC-PR) [50] генерира размита функция на членство, в случая използвайки общите езикови термини „low“, „medium“ и „high“, корелиращи с броя потребителски посещения:

- емпиричният ранг на стойността  $x$  с метриката  $M$  определя принадлежността към размития клас „high“:  $\mu_{\text{high}}(x) = P(M < x)$

- показателят  $M$  ще бъде класифициран като „low“ (отрицанието на принадлежността към класа „high“):  $\mu_{\text{low}}(x) = 1 - \mu_{\text{high}}(x)$

- класификация „medium“ се дефинира като:

$$\mu_{\text{medium}}(x) = 1 - \text{abs}(\mu_{\text{high}}(x) - 0.5) - \text{abs}(\mu_{\text{low}}(x) - 0.5).$$

В провежданото изследване [82] се прави анализ на посещенията на уеб сайтовете, указани като помощни материали в курса по дисциплината Изкуствен интелект. За всяка уеб страницата се оценява посещаемостта, като например: страницата  $W1$  има 135 посещения в рамките на обучаващия курс. Общо предложените и наблюдавани страници в курса са 80 и 56 от тях имат по-малко посещения от  $W1$ . За оценка на посещенията на  $W1$ , се пресмята:

$$\mu_{\text{high}}(\text{посещение } (W1)) =$$

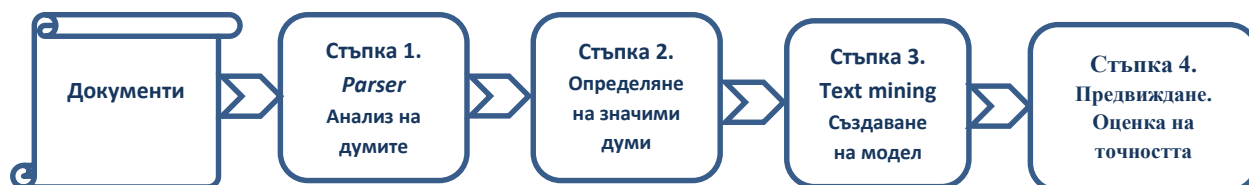
$$P(\text{Брой посещения} < \text{посещения } (W1)) = 56/80 = 0,7.$$

Така според размитата класификация се определя, посещаемост на страницата „висока“ над 0,7, за „ниска“ под 0,3, а за „средна“ до 0,6. На базата на получения анализ за всяка уеб страница, предлагана в обучаващия курс, се взимат решения за актуализиране на уеб източниците,

както и допълване на литературата с нови сайтове, свързани с най-търсените теми.

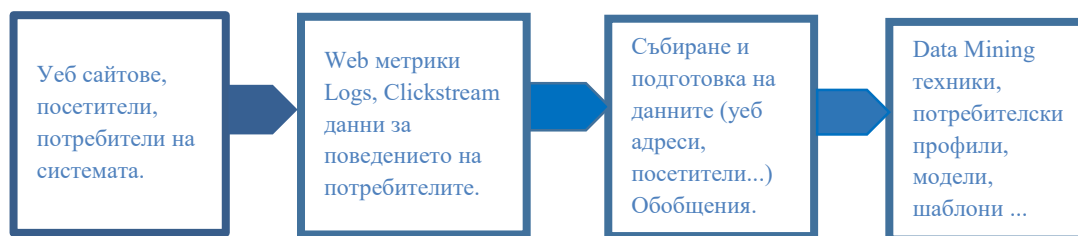
След като се направи анализ на най-посещаваните от обучаемия уеб-сайтове, дълбочината и честотата на посещенията, следва да се анализира и определи типа на тяхното съдържание. Подход за автоматизирано определяне типа на документ и доколко е свързан с областта на обучението е представен в [80]. Анализът се базира на идеята: определени думи в документа са значими за съдържанието му. За да се определят тези думи се търси честотата на появата им, използвайки софтуера *Wordcloud* [104]. Могат се налагат филтри, относно получавания списък с думи. Изчислява се близост на всяка дума от полученото множество, с всички думи от речника на предметната онтология, свързана с областта на обучение и се взима най-малката получена стойност за всяка дума. За определяне на близост на думите и фразите се използва подход, представен в [48] чрез q-gram метрики.

Чрез обработка на голямо множество от уеб-документи се натрупва набор от данни с извлечени от тях централни думи и степента им на близост с понятията от анализираната област. Тези данни са използвани за обучение на класификационни алгоритми, представени и описани под формата на многостъпков процес, като целевия атрибут е дали документа е свързан с областта или не. Предложен е метод за анализ на документи, показан с схемата на фиг. 17.



Фигура 17. Многостъпков процес на анализ и прогнозиране на документи от зададена област

От събираната информация за дейността на потребителите в уеб пространството с помощта на клъстерен анализ или прилагане на асоциативен анализ, може да се извършва потребителско сегментиране. Уеб съдържанието може да бъде анализирано на следните стъпки (фиг. 18).



Фигура 18. Архитектура за извличане на данни от уеб метрики

**ИЗВОД:** В етапа на анализ на резултатите от проведеното обучение, особено внимание се отделя на трудностите, които обучаемите срещат при изпълнение на отделните модули в дисциплината. В проведения експеримент, направените изводи са свързани с отчитане на факта, че студентите в изследваната дисциплина Изкуствен интелект срещат трудности при работа с литературните източници за изготвяне на проекта по избрана тема от *Модул 4*, в областите: „Blockchain“, „Ethics and Emerging Sciences“, „Policy and investment recommendations for trustworthy Artificial Intelligence“ както и при определяне на рискове, свързани с тези технологии.

Чрез представения метод в средата за обучение може да се проследи активността на обучаемите в интернет пространството, като се натрупват данни, относно тяхната активност в обучаващата среда. Чрез анализ на този тип данни се получава навременна информация за активността, напредъка и успеха на обучаемите. Резултатите показват, че формулираните решения могат успешно да се използват за различни задачи и могат да бъдат адаптирани към нови технологии и приложения. Предложеният метод насърчава създаването на иновативна обучаваща среда и е следващата стъпка в дигитализацията на образованието.



### 3.4. Моделиране на потребителя в образователното пространство

Чрез подходящи средства за следене и отчитане на дейността на потребителя се изгражда негов модел. Моделът служи за персонализиране на системата към знанията и уменията на потребителя и прилагане на адекватни подпомагащи стратегии в работата му. **Моделът на потребителя** по своята същност е структура от данни, представяща индивидуални характеристики за потребителя, на дадена програмна система. Той представя познавателните процеси и схващания на потребителя за приложната област [7].

Според Barr и Feigenbaum [10] съществуват два **базови подхода**:

- При *моделите с припокриване* (overlay approach) знанието на потребителя се представя като подмножество на общото знание, поддържано от системата за областта, която се изучава, знанието на експерта в областта или очакваното знание за обучаемия. Знанието на системата е декомпозирано на независими компоненти и е покрито със система от означения, показващи нивото на усвояване на всяка отделна компонента. Приемайки това становище тук се изгражда модел за оценяване на базовите знания на обучаемите относно основните понятия и зависимости между тях. В преподаваната предметна област е изградена йерархия на понятията, дефинирана в използваната домейн-онтология. След всеки тест на понятията се приписва относителна числова стойност (в проценти), която показва степента на сигурност на системата относно познаването на това понятие от конкретния обучаем. Оценката за всяко понятие може да се формира динамично като осреднена оценка на дъщерните му понятия по формулата:

$$(10) \quad \text{Mark\_term} = \frac{1}{k} \sum_{i=1}^k \text{Mark\_Subterm}[i],$$

където  $\text{Mark\_Subterm}[i]$  е оценката на  $i$ -тото дъщерно понятие, а  $k$  е броят на дъщерните (наследените) понятия на оценяваното основно понятие. Всяко дъщерно понятие от своя страна може да се разглежда като родителско на своите дъщерни понятия и да получава оценката си по същата формула. Тогава общата оценка за обучаваната тема може да се формира по формулата:

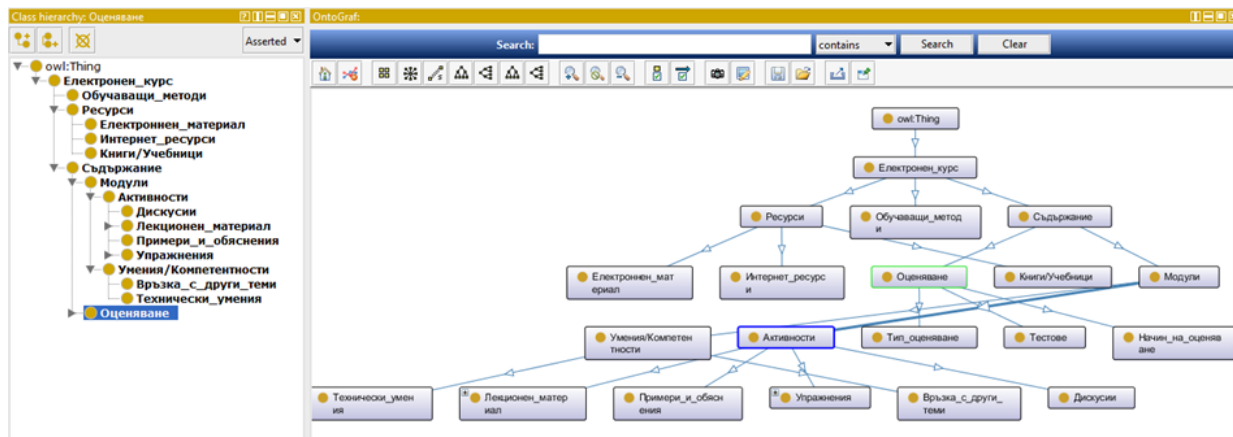
$$(11) \quad \text{Mark} = \frac{1}{n} \sum_{j=1}^n \text{Mark\_Term}[j],$$

където  $\text{Mark\_Term}[j]$  е оценката на  $j$ -тото основно понятие в изучаваната тема от предметната област, а  $n$  е броя на тези основни понятия.

- При подхода *моделиране с библиотека от грешки* (bugs theory) формално се диагностицират грешки, чрез списък от предварително дефинирани грешно усвоени и липсващи елементи от знания. Тук моделът на потребителя се състои от модел на експерта, допълнен със списък от допускани грешки в областта. Голяма част от тези данни се събират от средата за електронно обучение и се съхраняват в реляционна база от данни. Използвайки натрупаните данни, се генерират справки за участниците и работата им, например: справка за участниците в дадена дейност, справка за продължителност на дадено обучение, дата на започване и дата на завършване на участието; данни за обучаемите и курсовете, в които участват; общ брой участници; брой участници, преминали обучение в отделните курсове и др.

За да може обучението да бъде формално представено на мета-ниво, то неговите базови аспекти (знание, процеси и участници) трябва да бъдат моделирани и формализирани по начин, който осигурява дефиниране, намиране и използване на релации между тях. В дисертационния труд се разработва и предлага модел на учебния процес в образователното пространство, базиран на онтологии за представяне на знанията. На фигура 19 е представена схема на разработвания онтологичен модел на електронен курс. Прототип на онтологията е тествана чрез системата Protege (<http://protege.stanford.edu/>).

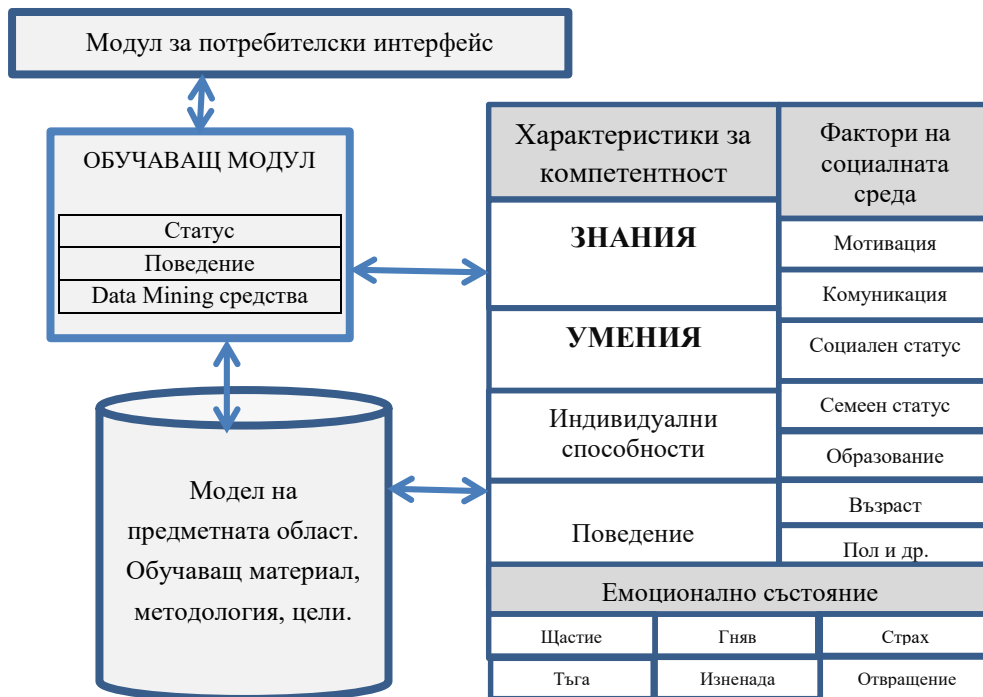




Фиг. 19. Онтологичен модел на електронен курс чрез системата Protege

Моделът на знанията от областта на обучение представя основните концепции и техните връзки, които трябва да бъдат преподавани и проследявани като ниво на усвоеност.

На базата на експерименти с обучаеми в електронна среда по различни дисциплини тук се предлага обща рамка на архитектура на модел на обучаемия, който включва три типа фактори: фактори на компетентност, емоционални фактори, фактори на въздействие на социалната среда.



Фиг. 20. Архитектура на обучаваща система с модел на обучаемия

Предложеният модел на потребителя е обособен като отделен модул, съдържащ данни за потребителя, и система от функции, грижещи се за събирането, съхранението, обработката и интерпретацията на тези данни. Този модул е свързан с модела на предметната област на обучение и обучаващият модул на системата. В индивидуалния модел (фиг. 20) се съдържат три типа данни.

- Първият тип представя оценки относно елементарни умения и типични грешки при работа в областта на обучение. Тези оценки се обобщават в общата оценка на обучаемия в областта.

- Вторият тип показва нивото на усвояване от обучаемия на всяка от единиците знание (ЕЗ). За обучаем се поддържа списък от елементи от вида: <номер на ЕЗ> : <оценка на ЕЗ>.

- Третият тип касае работата на обучаемия по време на обучаващата сесия: съдържа номера на учебна програма, текуща трудност на въпросите и задачите; нивото при оказване на помощ и др.

Тук се включват и данни за обучаемия като: пол, възраст, културен и социален статус (включва публикации във форум и брой отговори, групово взаимодействие), както и някои индивидуални особености. Тази компонента е в процес на експериментални изследвания и затрудненията произтичат от факта, че не е известна количествена теория за връзките между психологическите характеристики на обучаемия.

Оценките указващи нивото на усвоеност или неувоеност на дадено знание или оценките на въпроси, задачи или заявки на обучаемия, представляват реални числа в интервала  $[-1,1]$ .

Комбинирайки този подход с подхода за поддържане на библиотека от грешки, за всеки от предварително дефинираните типове грешки се поддържат броячи за случаите, когато е било възможно да бъде допусната грешка и броя на грешните опити. След приключване на събитието за всеки от типовете грешка се формира оценка по формулата:

$$(12) \quad Mark\_current = 1 - Count\_error / Count$$

Тук  $Mark\_current$  е междинна оценка, която е число в интервала  $[0,1]$  и всъщност представлява процента на опитите от страна на потребителя, при които не е допусната грешка.  $Count$  е общият брой опити, а  $Count\_error$  е броят на грешните опити. За да се получи окончателна оценка  $Mark$  за уменията в интервала  $[-1,1]$ , се пресмята по формулата:

$$(13) \quad Mark = 2 \cdot Mark\_current - 1.$$

Вземайки предвид предните две формули, за оценката получаваме:

$$(14) \quad Mark = 1 - 2 \cdot Count\_error / Count.$$

В случая, когато общият брой опити е 0, на оценката присвояваме стойност 0. След като е формирана оценка на типа грешка след събитието, тя се отразява на оценката за съответния тип грешка в модела на потребителя по формулата:

$$(15) \quad Mark\_new = (1 - K) \cdot Mark + K \cdot \text{Оценка\_на\_задачата}.$$

Тук  $K$  е число в интервала  $[0,1]$  и представлява влиянието, което оценката на типа грешка указва върху оценката в модела на обучаемия. Пресмята се по формулата:

$$(16) \quad K = 0.5 \cdot (Count - 1) / Count$$

Така в оценката за всяко от уменията се съдържа информация от предишни задачи и от последната задача. Оценката се умножава по 0.5, за да се ограничи влиянието на старата оценка.

Пресмятането на оценките от втория тип за нивото на усвоеност на отделните единици се извършва като, за всяка решавана задача или поставен въпрос, оценката се променя по формулата:

$$(17) \quad \text{Оценка\_нова} = (1 - K) \cdot \text{Оценка\_стара} + K \cdot \text{Оценка\_на\_проблема}$$

Но тук  $\text{оценка\_на\_проблема}$  е оценката на въпроса или задачата, задаена от преподавателя.  $K$  е влиянието, което оценката на проблема оказва върху оценката на единицата знание и се пресмята по формулата:

$$(18) \quad K = \text{Трудност\_на\_проблема} / \text{Трудност\_на\_ЕЗ}$$

колкото трудността на проблема (въпрос/задачата) е по-голяма, толкова по-голямо влияние оказва върху оценката на единицата знание.

Оценките са в интервала  $[-1,1]$ , когато оценката е по-близо до 1, то системата е по-сигурна, че обучаемият е усвоил вярно дадената единица знание. Колкото оценката е по-близо до -1, то системата е по-сигурна, че обучаемият е усвоил грешно дадената единица знание. Когато оценката е близо до нула, системата не е сигурна в знанията на обучаемия. Така се отчитат характерни черти и на подхода “моделиране с припокриване” и на подхода “моделиране с библиотека от грешки”, защото: ако оценката е положителна, тя представя нивото на знанията на обучаемия в сравнение със знанията на експерта, ако оценката е отрицателна, тя представя грешно усвоени знания.

Използват се броячи, с които се следи броя на използваните обучителни ресурси и типа на посетените интернет ресурси, които в различни комбинации могат да осигурят различно представяне на учебните материали. Информационните ресурси се оценяват по брой, видове и др.

## Разпознаване на емоционално състояние на обучаемия в среда за електронно обучение

За целите на провежданото изследване в дисертационния труд е разработено софтуерно приложение за разпознаване на човешки емоции по образа на лицето. Разработеният софтуер се базира на системата за кодиране на лицевото действие (FACS) на Ekman и Friesen [27]. Лицевият израз на емоциите се определя от включените мускули и жестовете, които го характеризират. Съществуват определени отличителни модели на емоционални реакции, обобщени и споделени от повечето хора. Те се считат за основни емоции: щастие (*Joy*), тъга (*Sadness*), гняв (*Anger*), изненада (*Surprise*), страх (*Fear*) и отвращение (*Disgust*). Други подобни кодиращи системи са EMFACS, MAX, AFFEX и CANDIDE-3. Разработеният софтуер и проведеното изследване са представени в [57].

*Методология.* Първата стъпка е категоризиране на зададените снимки със изображения на лица, в отделни групи (поддиректории). Те са седем на брой (включвайки и *Neutral*), като всяка за от тях отговаря за конкретна емоция и съдържа около петдесет примерни изображения, свързани със съответната емоция. Следваща стъпка е използването на *Dlib* библиотеката от алгоритми за създаване на софтуер с общо предназначение, която се използва в много области. Чрез използване на файла *shape\_predictor\_68\_face\_landmarks.dat*, лицата от изображенията се маркират със съответни основни точки. На базата на поставените точки, в следващата стъпка се извършват математически изчисления, с помощта на които се определят 6 нормализирани вектора на всяко лице. Всеки вектор с лицеви характеристики е изчислен като данните се нормализират и се съберат разстоянията между маркираните точки. Чрез средствата на framework „ML.Net“ се имплементира многокласов класификационен алгоритъм, базиран на ML агент от тип Supervised Learning.

След обучението на модела, той може да разпознава емоционалното състояние на подадено изображение (снимка) на лице. В резултат на работата, програмата връща предвиждане относно емоционалното състояние. Анализът на резултатите от метриците за оценка на точността на модела показва, че най-голяма точност на предвиждането се постига при емоционалните състояния *Joy*, *Surprise* и *Sadness*. Състоянието *Anger* често се бърка от алгоритъма със *Sadness*, *Joy* или *Neutral*. Алгоритъмът може да бъде подобрен чрез обучение с много по-голям обем тренировъчни данни от различните типове емоционални състояния. Но трябва да се използва балансиран набор от данни.

Алтернативен подход е прилагане на алгоритъм от тип *Unsupervised Learning*, при който се изследва структурата на данните без указания, данните за обучение не са етикетирани. Могат да се използват на *Clustering* алгоритми или *Deep Learning*. Прилагайки различни класификатори може да се експериментира с получаваните резултати за повишаване на точността, спрямо различни групи и категории данни. Целта е разработването на приложение да се интегрира към среда за e-Learning и да се търсят промените в емоционалните състояния на учащите се, докато се справят с когнитивните задачи в процеса на обучение.

## Устройство за подпомагане на потребители с увреден слух

Проектирано и експериментално изследвано е устройство [106], което произвежда широка гама от цветове и нива на интензитет, за да визуализира различни аспекти на звука. За хора с увреден слух, звуковата визуализация осигурява подобро усещане за заобикалящата среда и им позволява да интерпретират и да реагират на слухова информация в реално време.

Разработени са софтуерни инструменти за анализ на звуковите честоти и преобразуването им в съответните цветове в RGB модела. Създаден е модел на автоматизирана система, която реализира преобразуване на звуковите честоти в цвят. Постигнатата точност е оценена експериментално. Установено е, че точността на преобразуването зависи от честотата на звука. Доказано е, че процентът на грешки е по-висок при високи честоти на звуковите вълни (над 1000 Hz). Работа трябва да продължи с изследвания, насочени към обобщаване на моделите за преобразуване на звуковите честоти в цвят.

### 3.5. Когнитивен агент в среда за електронно обучение

Когнитивният агент [24] е автономен софтуер, който има способността да развива своите знания. Като се има предвид различната степен на интелигентност, когнитивните агенти могат да бъдат реализирани на базата на правила, които развиват знанията си с помощта на вградени в системата механизми за извод или извличат нови знания с помощта на техники на машинно обучение и Data Mining техники.

Общата рамка на архитектурата, на изгражданата за това изследване система, базирана на правила - *Rule Based Systems* включва компонентите [89]:

- o **База от знания** (множество от правила) - съдържа всички знания необходими на агента.
- o **База от данни**: съдържа данните, които са установени към текущия момент.
- o **Машината за извод (Inference Engine)** - интерпретаторът избира и прилага правилата, които могат да доведат до промяна на състоянието на базата от данни. Съществуват различни подходи: свързване напред, свързване назад и смесени подходи [89].

o **Базата от мета-знания** - мета-правила, задаващи връзката между правилата [24]. Начинът, по който се избират правилата от множеството на приложимите правила е стратегията за решаване на конфликти (*Conflict Resolution*). В случая се прилагат приоритети на правилата, които са динамични. Динамичните приоритети се определят въз основа на важността на действията в настоящата ситуация. Например: ако *оценката на обучаемия е висока*, то се задава нисък приоритет на правилото за търсене на следваща задача за решаване. Тогава обучаемият ще търси следваща задача за решаване само, ако няма други дейности за изпълнение. Но ако е със *слаба оценка*, правилото за решаване на следваща задача трябва да има висок приоритет. Обучаемият трябва да решава следваща задача, за да повиши оценката си.

От друга страна обучението с утвърждение (*Reinforcement Learning*) става все по-популярен тип машинно обучение. Това е техника, при която автономните агенти използват алгоритмите за проба-грешка и функция за кумулативно награждаване на обучението. Предимството се състои в калкулирането на оптимални действия, които агентите могат да предприемат в рамките на сценарии, обусловени от средата. Такъв подход е коренно различен от начина на действие при контролирано обучение, където целта е намаляване на отклонението съобразно предварително зададени правилни данни (вход/изход). Обучение протича като последователност от пробни действия, които постепенно водят до утвърждаване на добрите действия и избягване на неподходящите. Резултатът от обучението е оптимална стратегия за действие във всяка ситуация. Стратегията е оптимална, ако успява да максимизира сумата от всички награди, получени по време на изпълнението си.

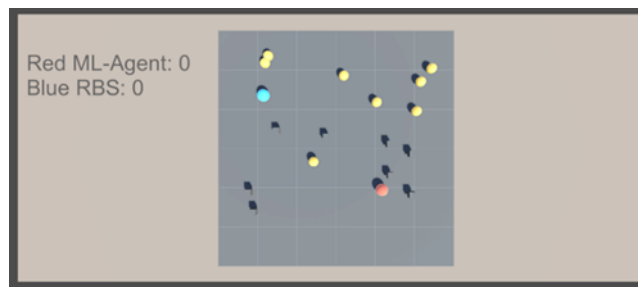
Формално обучението чрез *Reinforcement Learning* се разглежда като [54]:

- Множество  $S$ , включващо всички състояния на обкръжението, които агентът разпознава.
- Множество  $A$ , включващо всички действия, които агентът може да извърши.
- Множество от правила за преходи между състоянията.
- Правила, които определят  $R$  - наградите, която агентът може да получи при преходите.
- Правила, които описват какво спазва агента.

**Конкретна задача** в дисертационния труд е да се анализира и сравни поведението на два агента, реализирани на базата на двата различни подхода, базирани на машинно обучение. Резултатите от проведеното изследване и разработеният за целта софтуер са представени в [58]. За целите на анализа е създадена игра, която предоставя платформа и генерира  $x$  брой бомби и монети върху нея. Играта създава два агента, представени като две различни топки на платформата – червена и синя. Платформата няма оградящи стени, което означава, че топките (агентите) могат да паднат. Ако агент падне, той губи играта. Всеки агент може да събира монети, като всяка монета носи на агента точка. Ако агент докосне бомба, той губи точка. Играта

приключва, когато монетите свършат или когато агент падне от платформата. Агентът с повече точки печели играта. Един агент, представен със синя топка в играта, поддържа поведение въз основа на система от правила, създадени от програмиста и вградени в базата знания на системата. В този случай при изграждането на играта се прилага подходът *Rule-Based System*. От друга страна, агент, представен с червена топка, се реализира въз основа на подхода за машинно обучение с *Reinforcement Learning*. При него изкуственият интелект научава сам кое е най-доброто действие във всяка ситуация и с времето оптимизира решенията, които взема.

Разгледани са разликите в поведението на създадените решения чрез избраните два подхода. Хипотезата е, че агентът, реализиран с машинно обучение чрез *Reinforcement Learning* [54] ще има по-разнообразно поведение и след достатъчно време за обучение ще работи по-ефективно от агента, реализиран като система, базирана на правила.



Фигура 21. Изглед на разработения софтуер

Повече експерименти и разбиране на типа машинно обучение *Reinforcement Learning*, допринасят за разширяване на обхвата на използването му в съвременните среди за обучение. От друга страна алтернативата на системите, базирани на правила е подход доказал във времето своите предимства и ефективност за бързо постигане на поставените цели. Проектът [58] е създаден с помощта на платформата *Unity*, заедно с пакет на *Unity Technology* за разработване на агенти за машинно обучение *ML-Agents*. Този пакет предоставя сценарии за обучение и валидиране. Всеки агент има набор от състояния и наблюдения, може да предприема действия в средата и да получава награди.

- *Методология на изследването.*

Първият агент с *Reinforcement Learning*, е създаден чрез наследяване от клас *Agent*. Предефинира се методът *OnEpisodeBegin*, в който се задават първоначалните параметри за играта (монети и бомби и техните позиции). Предефиниране ба метода *CollectObservations*, осигурява предаване на текуща информация за посоката към най-близката монета. *OnActionReceived* метод задава разрешените движения на агента. Когато агент докосне обект, действието се засича и се извиква функцията "*OnTrigger*", която позволява да се види с кой обект агентът е влязъл в контакт.

Вторият агент, базиран на система от правила, е имплементиран чрез *RBS* класа и дейността на агента се задава от реализираните методи в класа. В зависимост от сценариите се прилагат правилата: „Ако има монети на платформата, отиди да вземеш най-близката“; „Ако RBS агента има повече монети от другия агент, той не трябва да се страхува от бомбите, за да завърши играта по-бързо и да не позволи на противника да събере повече монети. Изпълнението на правилата се основава на подхода свързване напред (*forward chaining*).

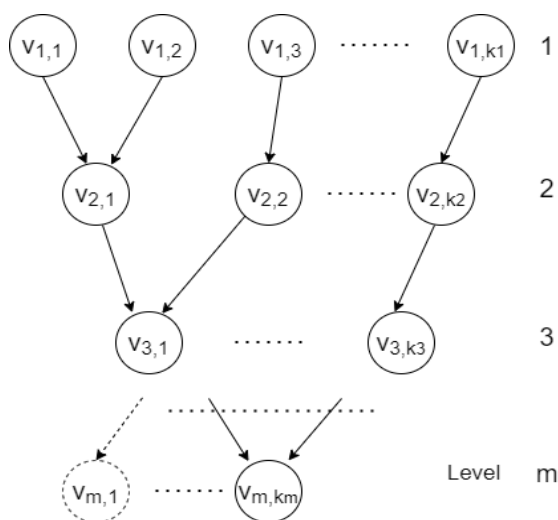
ИЗВОД: *Reinforcement Learning* агентът постепенно научава кое е най-доброто действие във всяка дадена ситуация и с времето оптимизира решенията, които взема. Той първоначално работи неефективно, но с течение на времето ще оптимизира действията си. От друга страна, при условие, че са създадени достатъчно добри правила, *Rule-Based System* е програмна задача, която може да поддържа много сложно поведение. Интересна тенденция в по-сложните приложения е съвместно прилагане на обучение с утвърждение с контролирано и неконтролирано обучение, ако използваните самостоятелно методи не дават достатъчно добър резултат.

### 3.6. Модели за йерархично многокомпонентно оценяване на обучаеми

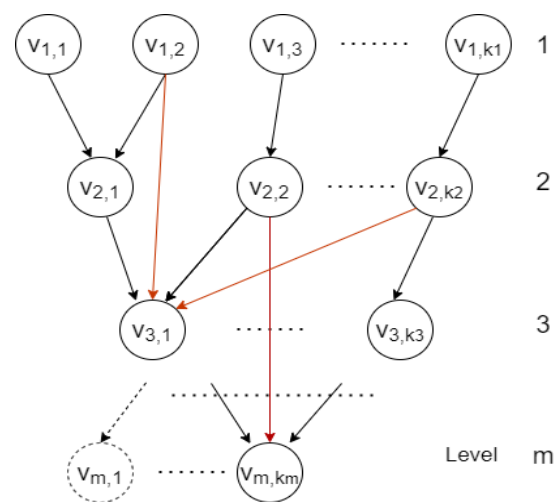
Представени са **нови модели** за йерархично многокомпонентно оценяване на обучаемите, които имат за цел комплексно оценяване на различни мисловни умения от висок и нисък порядък, теоретични знания и практически умения, и др.

Таксономията на Блум [13] дефинира йерархия на мисловните умения, в която по-високите нива на мислене включват всички познавателни умения от по-долните нива. Нивата са структурирани като: *Knowledge* (Знания), *Comprehension* (Разбиране), *Application* (Приложение), *Analysis* (Анализ), *Synthesis* (Синтез) и *Evaluation* (Оценяване). Всяко ниво се определя от множество когнитивни умения, дейности и методи за оценка. Счита се, че в процеса на обучение, обучаемият преминава последователно през всички нива. Една общоприета класификация определя уменията от горните три нива на Блум (*Analysis*, *Synthesis* и *Evaluation*) като *higher-order thinking skills* (HOTS), а тези от долните три нива (*Knowledge*, *Comprehension* и *Application*) – като *lower-order thinking skills* (LOTS). HOTS включва способности за критично мислене, умения за трансфер на знания и умения, умения за решаване на проблеми, и др. [15]. Оценката на студентите трябва да е обективна, а в много случаи е трудно, особено когато се отнася за оценяване на HOTS, които изискват креативно мислене. Това мотивира работа в насока създаване на комплексни многокомпонентни модели за оценяване на студентските постижения [36].

Предлага се **моделиране на йерархичната организация на оценъчните компоненти и зависимостите помежду им, които оценителят явно или неявно използва при оценяване**. Например, при множество дисциплини, основни оценъчни компоненти са „практика“ и „теория“, а в използваните термините тук: **thinking skills**, съответни техни под-компоненти могат да бъдат теоретични и практически **LOTS** и **HOTS**. От друга страна ако изберем за основни компоненти HOTS и LOTS, те биха могли да имат под-компоненти за оценяване на теоретични знания и практически умения. Създаването на конкретна йерархия от компоненти и подкомпоненти може да бъде извършено с помощта на предварително стандартизирани модели, избрани от оценителя. При достатъчно входно-изходни образци и ефективно обучение на невронна мрежа, получените резултати могат да отразяват тази йерархична организация, дори без тя да е указана явно.



Фигура 22. а) **Модел 1:** Йерархична дървовидна организация на оценъчните компоненти



Фигура 22. б) **Модел 2:** Йерархична графова организация на оценъчните компоненти

На фиг. 22 а) е представен общ **Модел 1 на йерархична дървовидна организация на оценъчните компоненти**, при който всеки основен компонент, в частност и крайната оценка, зависи от множество под-компоненти на предходното ниво. При това, всеки под-компонент влияе само върху един компонент от по-високо ниво.

**Модел 2 на йерархична графова организация на оценъчните компоненти** (фиг. 22 б), разширява и обобщава възможностите на Модел 1, с възможността един компонент да влияе на множество компоненти от произволни нива и да зависи от компоненти на различни по-ниски нива.

Представените модели могат да се използват при множество различни подходи за оценяване, при които крайната оценката е функция на множество оценъчни компоненти. Основни характеристики на предложените модели са:

- Ниво 1 на компонентната йерархия описва конкретни стойности от проведени оценявания, които могат да бъдат с различни оценъчни скали. Примери за такива стойности са: точки от тестове за теория и практика; точки от тестове за умения, оценки от задачи, проекти и др.

- На всяко следващо ниво, компонентите формират оценка, която е функция на оценките на подчинените му под-компоненти от предходното ниво. В частност, всички или част от функциите могат да бъдат с размитата логика.

- На последно ниво има един компонент, т.е.  $k_m=1$ , но не е задължително. Възможно е да има няколко крайни оценки, напр. за различни оценявани когнитивни умения. При това  $k_m \geq 1$ .

Възможен вариант за функции на ниво 2 е нормирането на всички стойности на компонентите от ниво 1 до една обща система за оценяване (например от 2 до 6). Аналогично, функциите на всяко следващо ниво биха могли да запазват нормираността на родителските компоненти в съответната скала за оценяване.

Обикновено, оценъчните под-компоненти имат стойности в пространството на реалните положителни числа  $R^+$ . При различни подходи за оценяване, стойностите могат да бъдат и в други пространства на реални, комплексни или други числа. Крайната оценка  $E$  задължително принадлежи на пространство от предварително определени възможни стойности – например тя трябва да е цяло число в интервала от 2 до 6, от 1 до 5, от А до F, и др.

При формалното математическо описание на **Модел 1** и **Модел 2**, всяко ниво с номер  $i$  притежава  $k_i$  на брой компоненти, определящи множеството:

$$(19) \quad V_i = (v_{i,1}, v_{i,2}, \dots, v_{i,k_i}) \in R^{k_i}, k_i \in N, i = 1, \dots, m$$

Компонентите на ниво 1 са конкретни стойности, получени при оценяване. На всяко следващо ниво  $i = 2 \dots m$  стойностите се получават в резултат на прилагането на функции върху компонентите от предходното ниво. В **Модел 1** броят на компонентите в дадено ниво намалява или се запазва равен на броя на компонентите в предходното ниво  $k_s \leq k_{s-1}, s = 2, \dots, m$ , което е следствие на обединяването на под-компонентите в компоненти. Броят на обобщаващите функции и съответно на компонентите, може да нараства в определени нива при **Модел 2**.

**Стойностите на компонентите**, след ниво 1, при **Модел 1** се изчисляват като **функция** върху стойности на компонентите от преките родителски нива:

$$(20) \quad v_{i,j} = f_{i,j}(P_{i-1,j}), P_{i-1,j} = (v_{i-1,s_1}, v_{i-1,s_2}, \dots, v_{i-1,s_j}) \subseteq V_{i-1}, \\ \forall i = 2, \dots, m, j = 1, \dots, k_i, s_j \leq k_{i-1} \text{ и}$$

компонентите на  $V_{i-1}$  се срещат само веднъж в някое  $P_{i-1,j}$ .

При **Модел 2**, параметри на функциите на дадено ниво, могат да бъдат произволни оценъчни компоненти от всички предходни нива, като при това отпадат множество ограничения:

$$(21) \quad v_{i,j} = f_{i,j}(Q), Q \subseteq \bigcup_{r=1}^{i-1} V_r, \forall i = 2 \dots m, j = 1 \dots k_i.$$

### **Експеримент с използване на fuzzy logic при йерархично многокомпонентно оценяване**

В проведения експеримент по дисциплината „Програмиране в Интернет“, чрез четири изпитни компонента, оценяваме теоретични знания и практически умения за програмиране:

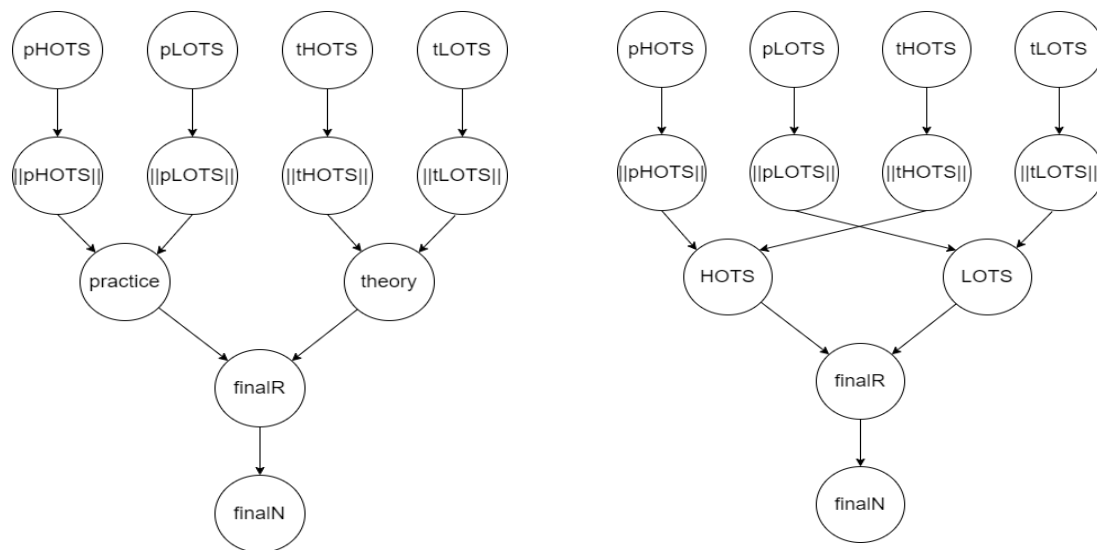
- tLOTS – теоретични LOTS, оценяват се чрез точки в интервал  $[0, 30]$ ;
- pLOTS – практически LOTS, оценката е цяло число в интервала  $[0, 15]$ ;

- tHOTS – теоретични HOTS, оценката е цяло число в интервала [0, 15];
- pHOTS – практически HOTS, оценката е реално число в интервала [2, 6].

Оценката pHOTS се формира от практическа задача и е в интервала [2, 6]. Точките на останалите оценъчни компоненти се формират при решаване на тест в хартиен или електронен вариант. Крайната оценка finalN е цяло число в скалата за оценяване [2, 6], т.е.  $finalN \in [2, 6] \subset N$ . Базирайки се на **Модел 1 на йерархична дървовидна организация на компонентите за оценяване**, дефинираме **два основни възможни варианта за формализиране логиката на оценителя** за оценяване по дисциплината (фиг. 23). При необходимост, чрез дефиниране на допълнителни връзки/зависимости между компоненти на различни нива, може да преминем към **Модел 2**. В практиката при различни ситуации могат да бъдат създадени и експериментирани и други варианти. Важното при всички случаи е, да се определи йерархията от компоненти и връзки между тях, както и множеството от съответните им агрегиращи функции, за даденото оценяване.

Основни компоненти, формиращи крайната оценка при **Вариант 1 за теоретико-практическото йерархично многокомпонентно оценяване** са **теория** и **практика** (фиг. 23 а). Под-компоненти на теорията са **нормализираните форми** на първичните компоненти **tLOTS** и **tHOTS**, а на практиката – нормализираните форми на **pLOTS** и **pHOTS**. Целта на нормализираните форми е свеждането на първичните компоненти до еднаква скала за оценяване (в случая от 2 до 6), така че да бъдат представени до по-близки за оценителя мисловни рамки за оценяването.

От друга страна, при **Вариант 2 за HOTS-LOTS йерархично многокомпонентно оценяване** (фиг. 23 б), за основни компоненти са приети HOTS и LOTS, а техни под-компоненти са съответно нормализираните форми на практическите и теоретични HOTS, и практическите и теоретични LOTS.



Фигура 23.а) **Вариант 1**: Теоретико-практическо йерархично многокомпонентно оценяване за **Модел 1**

Фигура 23.б) **Вариант 2**: HOTS-LOTS йерархично многокомпонентно оценяване за **Модел 1**

И в двата моделирани варианта на първо функционално ниво са нормализирани форми на основните оценъчни компоненти. Конкретните нормализиращи функции могат да следват **либерален** или **строг подход** за оценяване, или тяхна модификация, в зависимост от оценителя. При **либералния подход**, точките от първичните компоненти се свеждат до интервала [2, 6]:

$$(22.1) \quad ||tLOTS|| = 2 + 4 \frac{tLOTS}{30}$$

$$(22.2) \quad ||tHOTS|| = 2 + 4 \frac{tHOTS}{15}$$



$$(22.3) \quad \|pLOTS\| = 2 + 4 \frac{pLOTS}{15}$$

$$(22.4) \quad \|pHOTS\| = pHOTS.$$

При *строгий подход за оценяване*, точките първоначално се изобразяват в интервала [0, 6]. Оценки под 2,5 се закръглят на 2. Съответно, функциите на принадлежност (membership functions) за компонентите от второ ниво са:

$$(23.1.1) \quad \|tLOTS\| = \begin{cases} 2, & 6 \frac{tLOTS}{30} < 2.5 \\ 6 \frac{tLOTS}{30}, & 6 \frac{tLOTS}{30} \geq 2.5 \end{cases}$$

$$(23.2.1) \quad \|tHOTS\| = \begin{cases} 2, & 6 \frac{tHOTS}{15} < 2.5 \\ 6 \frac{tHOTS}{15}, & 6 \frac{tHOTS}{15} \geq 2.5 \end{cases}$$

$$(23.3.1) \quad \|pLOTS\| = \begin{cases} 2, & 6 \frac{pLOTS}{15} < 2.5 \\ 6 \frac{pLOTS}{15}, & 6 \frac{pLOTS}{15} \geq 2.5 \end{cases}$$

$$(23.4) \quad \|pHOTS\| = pHOTS.$$

Представени с IF-THEN логика, формулите изглеждат по следния начин:

$$(23.1.2) \quad \|tLOTS\|: \text{IF } 6 \frac{tLOTS}{30} < 2.5 \text{ THEN } \|tLOTS\| \text{ IS } 2 \\ \text{ELSE } \|tLOTS\| \text{ IS } 6 \frac{tLOTS}{30}$$

$$(23.2.2) \quad \|tHOTS\|: \text{IF } 6 \frac{tHOTS}{15} < 2.5 \text{ THEN } \|tHOTS\| \text{ IS } 2 \\ \text{ELSE } \|tHOTS\| \text{ IS } 6 \frac{tHOTS}{15}$$

$$(23.3.2) \quad \|pLOTS\|: \text{IF } 6 \frac{pLOTS}{15} < 2.5 \text{ THEN } \|pLOTS\| \text{ IS } 2 \\ \text{ELSE } \|pLOTS\| \text{ IS } 6 \frac{pLOTS}{15}$$

Средно-аритметичните стойности от под-компонентите не винаги са удачно решение и затова се стига до субективна преценка от оценителя. Има множество специфични случаи, които е необходимо да се формализират, за да се опише логиката на оценителя. Някои от използваните при нашето оценяване разсъждения за поставяне оценка на компонента *практика* са следните:

- Ако pHOTS или pLOTS е 2, оценката на компонент практика е най-много 3, дори и другия под-компонент да има оценка 6.
- Ако оценката pHOTS е висока, а pLOTS – ниска, то възникват съмнения за нечестни прояви на студента при решаването на задачата, формираща оценката pHOTS. Тогава, поставяме оценката да бъде по-ниска от средно-аритметичната стойност. Това се случва най-често, когато pHOTS е близо до 6, а pLOTS – до 3.
- Ако pLOTS е висока, а pHOTS – ниска, то показаните на теста знания гарантират, че студента има необходимата основа за бъдещо развитие по дисциплината. Въпреки, че не се е представил добре при решаването на задачата, то заслужава стимул в разумни граници. Те трябва да са предварително формализирани с точни числови стойности и правила.

Подобни, разсъждения са валидни и за теоретичните знания. При това, високите оценки tHOTS е добре да се толерират дори и при частичната липса на теоретични знания tLOTS.

Използваните в експеримента правила за поставяне на оценка на практика и теория при *Вариант 1 за теоретико-практическото йерархично многокомпонентно оценяване* са:

(24.1)  $\mathit{practice}(\|pLOTS\|, \|pHOTS\|)$ :

$$\begin{aligned} & \text{IF } \frac{\|pHOTS\| + \|pLOTS\|}{2} < 2.5 \text{ THEN } \mathit{practice} \text{ IS } 2 \\ & \text{ELSE IF } \|pLOTS\| = 2 \text{ OR } \|pHOTS\| = 2 \text{ THEN } \mathit{practice} \text{ IS } 3 \\ & \text{ELSE IF } (\|pHOTS\| - \|pLOTS\|) \\ & \quad \geq 2 \text{ THEN } \mathit{practice} \text{ IS } \left( \frac{\|pHOTS\| + \|pLOTS\|}{2} - 0.5 \right) \\ & \text{ELSE } \mathit{practice} \text{ IS } \left( \frac{\|pHOTS\| + \|pLOTS\|}{2} \right) \end{aligned}$$

(24.2)  $\mathit{theory}(\|tLOTS\|, \|tHOTS\|)$ :

$$\begin{aligned} & \text{IF } \frac{\|tHOTS\| + \|tLOTS\|}{2} < 2.5 \text{ THEN } \mathit{theory} \text{ IS } 2 \\ & \text{ELSE IF } (\|tHOTS\| - \|tLOTS\|) \\ & \quad \geq 2 \text{ THEN } \mathit{theory} \text{ IS } \left( \frac{\|tHOTS\| + \|tLOTS\|}{2} + 0.3 \right) \\ & \text{ELSE } \mathit{theory} \text{ IS } \left( \frac{\|tHOTS\| + \|tLOTS\|}{2} \right) \end{aligned}$$

За определяне на **крайната оценка** при представения **Вариант 1** е приложено теорията и практиката да са равнопоставени и правилата са формулирани като:

(25)  $\mathit{finalR}(\mathit{theory}, \mathit{practice})$ :

$$\begin{aligned} & \text{IF } \frac{\mathit{theory} + \mathit{practice}}{2} < 2.5 \text{ THEN } \mathit{finalR} \text{ IS } 2 \\ & \text{ELSE IF } \mathit{theory} = 2 \text{ OR } \mathit{practice} = 2 \text{ THEN } \mathit{finalR} \text{ IS } 3 \\ & \text{ELSE } \mathit{finalR} \text{ IS } \left( \frac{\mathit{theory} + \mathit{practice}}{2} \right) \end{aligned}$$

Крайната оценка  $\mathit{finalN}$  се определя като се закръгли  $\mathit{finalR}$  към най-близкото цяло число:

(26)  $\mathit{finalN}(\mathit{finalR}) = \text{ROUND}(\mathit{finalR})$

Определянето на оценките HOTS и LOTS при **Вариант 2 за HOTS-LOTS йерархично многокомпонентно оценяване** също са обект на подобни разсъждения и могат да бъдат персонализирани в зависимост от вижданията на конкретния оценител.

Формулите за оценка на компонентите **практика** и **теория** (24.1) и (24.2) са съобразени с **Модел 1**, затова участват компоненти единствено от предходното ниво. В определени гранични случаи е удачно използването на подкомпоненти от по-високо ниво и използване на **Модел 2**.

Предимство на размитата логика при оценяване е възможността за постигане на точно автоматизирано оценяване според субективната логика на оценителя. Основен недостатък е трудността за неспециалисти да формализират правилата за формиране на оценката.

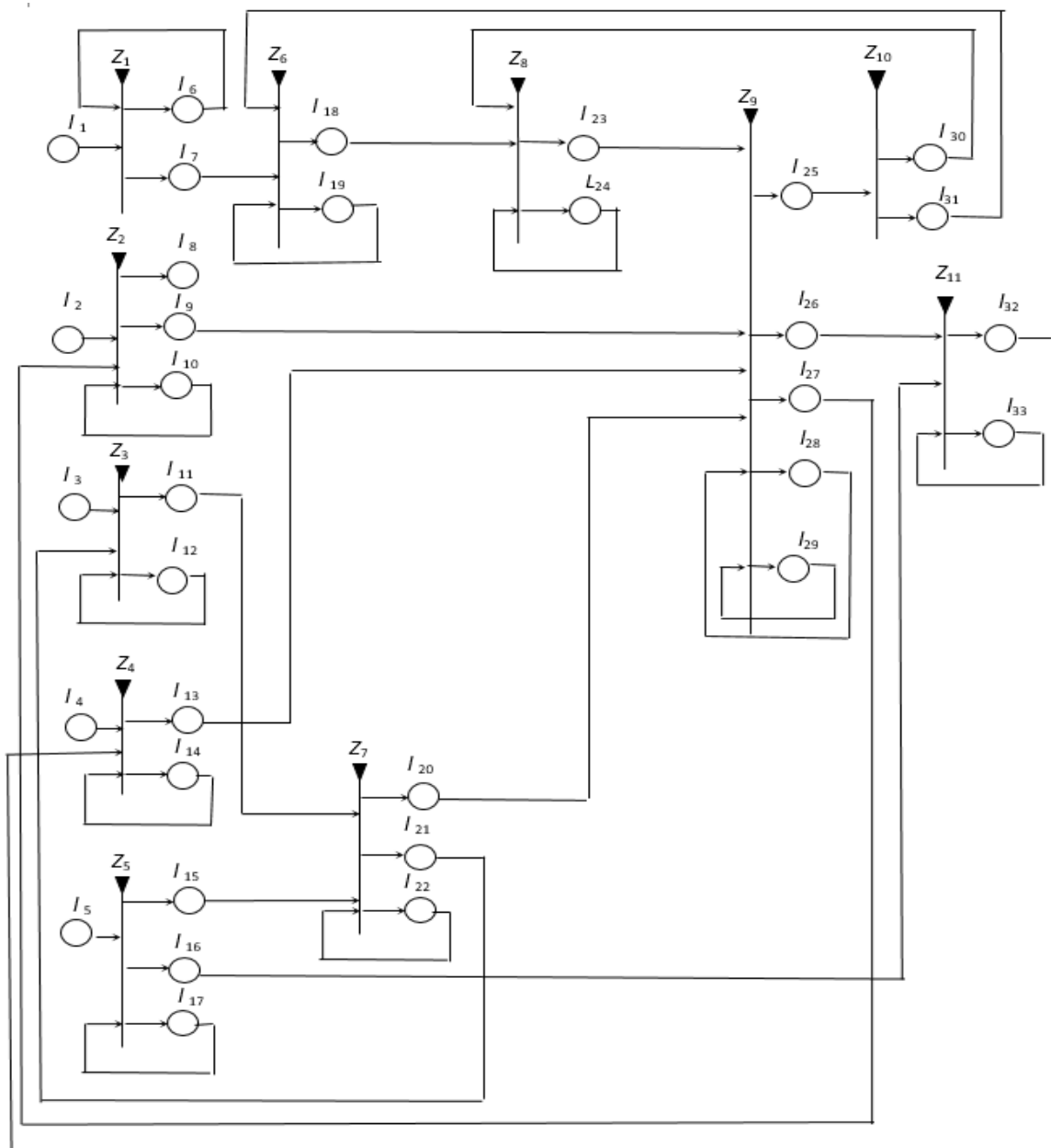
**ИЗВОД:** използването на размита логика гарантира получаването на коректни оценки, при коректно разписани правила, докато използването на методи на изкуствения интелект изисква достатъчно голям брой подходящи входно-изходни образци за обучение. Следва да отбележим, че всички автоматизирани подходи изискват наблюдение и контрол от страна на преподавателя на формираните оценки. Предложените модели за йерархично многокомпонентно оценяване имат за цел комплексно оценяване на различни мисловни умения от висок и нисък порядък, теоретични знания и практически умения и др. Те позволяват използването на размита логика и могат да бъдат адаптирани за употреба при оценяване на различни дисциплини, в различни възрастови групи.

## ГЛАВА 4. ОБОБЩЕНОМРЕЖОВИ МОДЕЛИ ВЪВ ВИРТУАЛНО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО

### 4.1. Модел на процесите на персонализиране и използване на среда за обучение

Целта на виртуалното образователно пространство е да интегрира по един „интелигентен“ начин реалния обучаващ процес с виртуално изградения свят. Образователните пространства трябва да предоставят възможности за: активно и интерактивно участие; работа в екип; търсене и споделяне на информация; дискутиране и представяне; произвеждане на нови знания; подпомагане дейностите на обучаемите и обучаващите; връзка с експерти и не на последно място персонализиране на обучението. Инфраструктурата на образователното пространство включва:

- *Изграждащи елементи* на пространството - могат да бъдат: обучаеми, обучаващи, администратори, персонални асистенти, дигитални библиотеки, електронни услуги и т.н.
- *Взаимовръзки* – съществуващите между изграждащите елементи взаимоотношения, осигуряващи съвместната им работа при опериране в пространството.



Фигура 24. OM модел на процесите на персонализиране и използване на среда за обучение

Създадена е серия от модели, свързани с наблюдение и анализ на дейностите на обучаемите в образователна среда, представени в публикациите [60, 70, 72, 97]. Обобщеномрежови модели за работа с големи данни във виртуално колаборативно пространство са създадени в [62, 63, 71].

Тук се представя обобщен модел на процесите на персонализиране и използване на среда за обучение (фигура 24). В този модел студентите (обучаваните обекти) се интерпретират чрез  $\sigma$ -ядра, обучаващите се интерпретират чрез  $\tau$ -ядра, обслужващият административен персонал (инспектори от учебен отдел) чрез  $\alpha$ -ядра, обучителните курсове (SCORM-пакети) в електронната библиотека чрез  $\lambda$ -ядра, електронните услуги предоставяни от средата за електронно обучение чрез  $\Upsilon$ -ядра и софтуерните агенти, в ролята на персонални асистенти на обучаемите чрез  $\beta$ -ядра.

Този модел е минимален редуциран обобщеномрежови модел [8, 9]. В дисертационния труд е дадено подробно описание на отделните преходи.

Обобщеномрежовият модел се състои от 11 прехода:

- $Z_1$  – постъпване на нов студент (обучаем) в електронната среда за обучение;
- $Z_2$  – постъпване на обучаващ преподавател в електронната среда;
- $Z_3$  – обслужване на обучението от административен учебен отдел;
- $Z_4$  – въвеждане на нови курсове за обучение в електронната библиотека;
- $Z_5$  – работа с електронни услуги, предлагани от обучаващата среда;
- $Z_6$  – избор на курс за обучение от студент в e-Learning средата;
- $Z_7$  – процес на администриране на нов курс за обучение;
- $Z_8$  – работен процес на агента (персонален асистент на студента);
- $Z_9$  – процес на обучение на студента по избрания електронен курс;
- $Z_{10}$  – процес на оценяване на обучението на студента по избрания курс;
- $Z_{11}$  – процес на изграждане на учебно съдържание на електронен курс.

ИЗВОД: Създаденият ОМ модел [68], дава възможност за проследяване на процесите на персонализация и използване на различни интелигентни инструменти за електронно обучение. Може да се извлече информация за предпочитанията на обучаемите и резултатите от тяхното обучение. На базата на модела и натрупана статистика от реални данни могат да се правят оценки и да се откриват тенденции за развитието на процесите, свързани с електронното обучение и неговото обслужване. Могат да бъдат въведени допълнителни параметри на модела и допълнителни характеристики на ядрата, като се вземат предвид факторите, влияещи върху процеса на обучение, с цел повишаване на интереса на обучаемите. Активното участие на обучаемите в процеса на усвояване на знания и придобиване на умения в голяма степен може да бъде повлияно от качеството на използваното образователно пространство.

#### **4.2. Модел на процеса на прилагане на средства за извличане на знания от данните в среди за обучение чрез апарата на обобщените мрежи**

Създаденият обобщеномрежов модел описва възможности за избор и прилагане на подходящи техники за извличане на знания от данните в средите за обучение [59].

Този модел е редуциран обобщеномрежови модел, но от по-висок клас спрямо модела, представен на фиг.24. Тук за всеки преход е дефинирано условие за активиране, като се задава чрез булев израз. Ако стойността му е “*true*”, съответният преход може да се активира, ако е “*false*” – не. Моделът, представен на фиг. 25, съдържа 5 прехода и 21 позиции, групирани в две групи и свързани с два типа ядра, които постъпват в съответните типове позиции:  $\alpha$ -ядра и  $l$ -позиции представят процеса на извличане на знания от данните,  $\beta$ -ядра и  $t$ -позиции представят

критериите за ограничаване на средствата и избора на подходящи техники за извличане на знания от данните. За краткост се използва означението  $\alpha$ - и  $\beta$ -ядра вместо  $\alpha_i$ - и  $\beta_j$ -ядра, където  $i, j$  са номерата на съответните ядра и само влизащите в обобщената мрежа ядра са номерирани с  $\alpha_0, \alpha_1$  и  $\beta_1, \beta_2$ . Първоначално едно  $\beta_0$ -ядро стои в позиция  $t_6$  с начална характеристика:

„налични средства за извличане на знания от данните“.

На следващия преход от функционирането на мрежата,  $\beta$ -ядрото се разделя на две ядра. Оригиналното  $\beta_0$ -ядро ще продължи да стои в позиция  $t_6$ , докато другото  $\beta$ -ядро ще се придвижи към прехода  $Z_5$ , преминавайки през прехода  $Z_3$ .

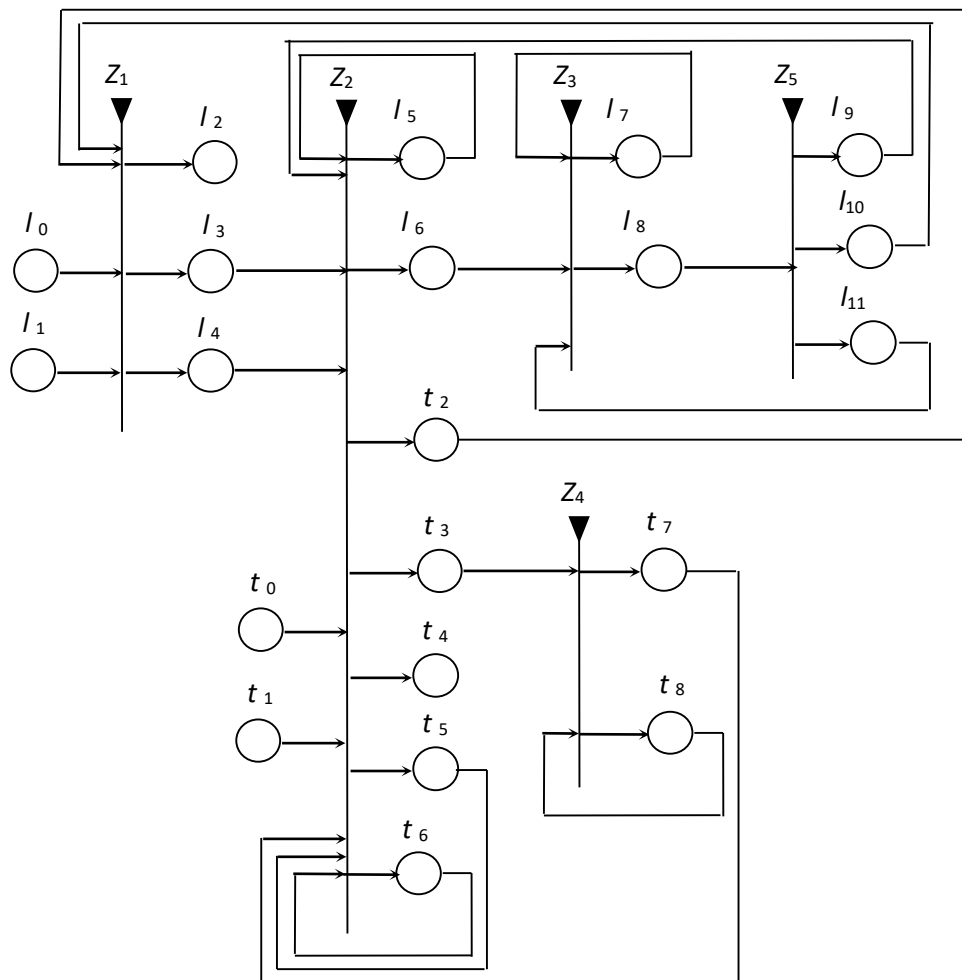
Ядрата  $\alpha_0$  и  $\alpha_1$ , постъпващи в мрежата през позиции  $l_0$  и  $l_1$  имат характеристики съответно:

„начални хипотези“;

„начални данни“.

Ядрата  $\beta_1$  и  $\beta_2$  постъпват в мрежата през позиции  $t_0$  и  $t_1$ . Тези ядра получават съответно начални характеристики: „нова техника за извличане на знания от данните“;

„критерии за избор на техника за извличане на знания от данните“.



Фигура 25. OM модел при прилагане на средства за извличане на знания в среда за обучение

**ИЗВОД:** Интегрирането на системите за обучение със средства за извличане на знания от данните е необходимо за процеса на персонализиране на курсове за електронно и дистанционно обучение. На базата на получените резултати могат да се въведат допълнителни мерки за анализ и промяна на обучаващите курсове и критериите за анализ. Това от своя страна е път към повишаване на качеството на обучението във висшето училище.

### 4.3. Обобщеномрежов модел на процес за многокомпонентно оценяване

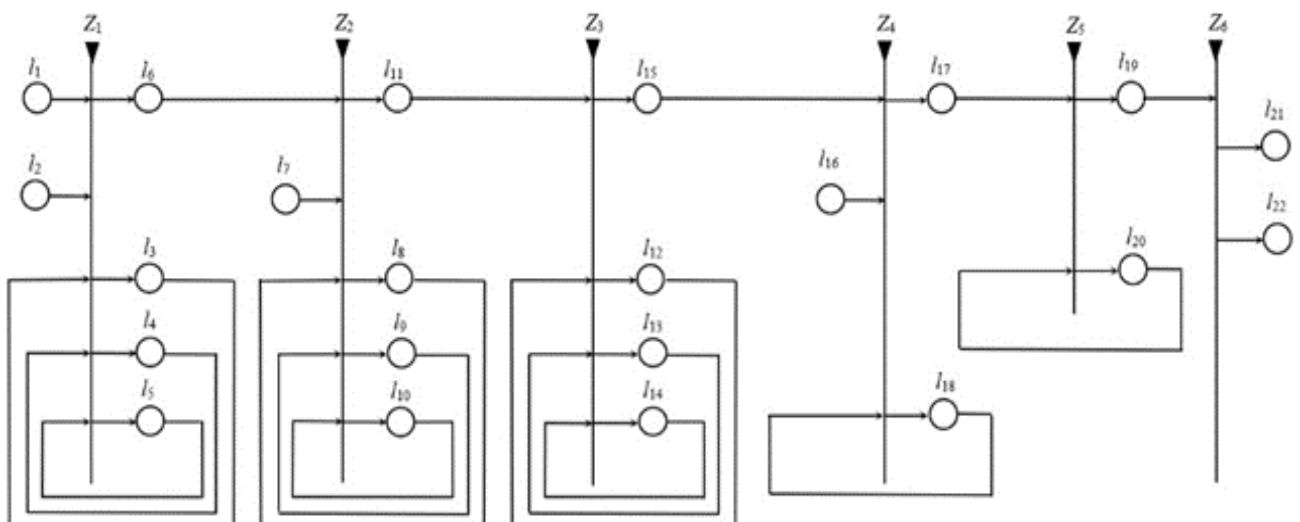
Обективното оценяване на знанията и уменията на обучаемите има множество характеристики: истински резултат (*true score*), надеждност (*reliability*) и валидност (*validity*), честност, диференцираност, всестранност на оценката и др. [91]. То предоставя на преподавателите по-добра възможност както за оценка на усвояването на знанията от обучаемите, така и за анализ на ефективността на преподаването. Създаден е обобщеномрежов **модел на процес за многокомпонентно оценяване**, който включва шест етапа:

1. **Създаване на мета-модел на тест.** Мета-моделът дефинира типове оценъчни компоненти, определящи какви знания и умения се оценяват според конкретна таксономия и моделите за оценяване.
2. **Създаване на модел на тест.** Моделът на тест включва характеристики за целевата група, предмета на обучение, началото и продължителността на теста, и др. На всеки тип оценъчни компоненти на мета-модела се задават съответен брой тестови единици.
3. **Конфигуриране на тест.** Задават се конкретни тестови единици, които се въвеждат или избират от база от данни с въпроси. За всяка конкретна тестова единица се указва към кой тип оценъчни елементи принадлежи. Задават и конкретните обучаеми.
4. **Провеждане на теста** е процес, при който се генерират конкретни тестове. Те включват определения в модела брой тестови единици от всеки тип. Обучаемите въвеждат своите отговори и решения.
5. **Процес на оценяване.** Включва автоматично оценяване на теста според избрания модел и ръчно оценяване на отворените въпроси и задачи, ако има такива.
6. **Анализ на резултатите и оценяване на теста.** Създателят на теста използва получените резултати, за да оцени теста и ефективността на подходите за оценяване.

ОМ моделът, описващ процеса на моделиране и конструиране на тестове за многокомпонентно оценяване на знанията и уменията на обучаемите [43] е представен на фиг.26. Този модел е минимален редуциран обобщеномрежов модел и съдържа множеството от преходи:

$$A = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6\}, \text{ където:}$$

- $Z_1$  – създаване на мета-модел на тест;
- $Z_2$  – създаване на модел на тест;
- $Z_3$  – конфигуриране на тест;
- $Z_4$  – провеждане на тестване;
- $Z_5$  – процес на оценяване;
- $Z_6$  – анализ на резултатите и оценяване на теста.



Фигура 26. Обобщеномрежови модел на процес за многокомпонентно оценяване

За описание на процесите в модела са използвани следните ядра:

- $\alpha$ - ядра интерпретират потребителите – администратори, създатели на тестове, автори на въпроси, обучаеми;
- $\beta$ - ядра – типове оценъчни компоненти, за оценяване на практически или теоретични умения, умения за мислене от нисък или висок порядък, умения, категоризирани според конкретна таксономия за обучение или др.;
- $\gamma$ - ядра интерпретират модели за оценяване – среднопретеглена оценка на оценъчните компоненти, оценка, базирана на теглови коефициенти, размита логика или др.;
- $\delta$ - ядро – мета-модел на тест;
- $\eta$ - ядро – модел на тест;
- $\lambda$ - ядро – банка от данни с тестови въпроси;
- $\mu$ - ядро – оценъчен компонент (например: тестов въпрос, задача, казус или др.).

Процесът се стартира от администратор на системата, представен чрез  $\alpha$ -ядро, който постъпва в мрежата през позиция  $l_1$  с начална характеристика:

*„уникален идентификатор и име на потребител на софтуерната система“.*

В позиция  $l_2$  се стартира заявка за създаване на нов мета-модел на тест, чрез  $\delta$ -ядро с начална характеристика:

*„уникален идентификатор и наименование на мета-модел на тест“.*

В позиция  $l_3$  цикли  $\alpha$ -ядро с характеристика:

*„списък с налични потребители на софтуерната система“.*

В позиция  $l_4$  и в позиция  $l_5$  циклят съответно  $\beta$ -ядро и  $\gamma$ -ядро с характеристики:

*„типове оценъчни компоненти“* и

*„налични модели за оценяване“.*

В дисертационния труд е дадено подробно описание на отделните преходи.

ИЗВОД: Предложен е обобщеномрежови модел, който има за цел от една страна да обобщи процеса на оценяване, а от друга – да даде възможност за персонализация на начина за формиране на тестове и оценяване по време на обучение [43]. Това е постигнато чрез дефиниране на мета-модели и модели за оценяване, които задават рамки за създаване на конкретни тестове и подходи за оценяването им. Разработването на софтуерна система, имплементираща описания процес, ще предостави на преподавателите гъвкава платформа за експериментиране с различни стандартни и авторски подходи за конструиране на тестове и оценяване на обучаемите.

#### **4.4. Обобщеномрежови модел на процесите в проектно-базирано обучение**

По своята същност проектно-базираното обучение е педагогически модел на междупредметни дейности, насочени към реално съществуваща проблематика [11]. Основни умения, които обучаемите формират са: да се научат да идентифицират етапите в разработването на проект, да планират дейността си, да спазват планираните срокове, да работят съвместно с други членове на екипа, да оценяват дейността на другите членове на екипа, да самооценяват действията си, да участват в дискусиите по темата на проекта, като формират и защитават аргументирано собствени идеи и умения.

##### **Процес на управление на работата по проекти в обучението**

Успешното управление зависи от ръководителя на дисциплината: той може да разпределя задачите за изпълнение, да създава график за работата, да контролира изпълнението на подзадачите и да следи цялостната работа. За целта могат да се правят срещи за представяне на текущите резултати и обсъждане на дейностите или да се работи в подходяща интерактивна среда като: Moodle, Acolad или среда на база Wiki технологията.

### **(1) Основните дейности при дефиниране на проекти**

Основните дейности по подготовката и планирането на проектите се формализират като:

- Спецификация на задачата за конкретен проект.
- Дефиниране на подзадачите, времето, за изпълнение и кой ще се заеме с тях.
- Дефиниране и разпределяне на ресурсите.
- Управление на изпълнението на задачите.
- Събиране на различни данни за статистика и за измерване на развитието на проекта.
- Описание на възможните рискове за проекта и план за управлението му.
- Генериране на различни справки, за работата по проекта.

### **(2) Планиране на екипи (групи)**

Основни дейности, свързани по планирането на групите, работещи по проектите са:

- Определяне на броя на участниците в екип по даден проект.
- Определяне на структурата на екипа.
- Определяне на ръководител (от преподавателя или избран от участниците).

### **(3) Поставяне на задачата пред студентите**

При дефинирането на проекта трябва да бъдат представени пред студентите последователно:

- предварителна информация;
- задачата, поставена интересно и мотивиращо;
- постъпково описание на процеса при изпълнение на задачата;
- набор от препоръчителни информационни източници;
- насоки за организиране и съхраняване на информацията;
- начини и критерии за оценяване на изпълнението на задачата.

### **(4) Дейности на обучаемите, изпълняващи проекти**

При изпълнението на проект, обучаемите се налага да: прецизират поставения въпрос или задача (например да го конкретизират или обобщават); събират и анализират данни от различни източници; споделят, генерират и дискутират различни идеи; правят свои аргументирани предположения, хипотези и предвиждания; провеждат и анализират собствени експерименти; създават артефакти (реферати, бази от данни, мултимедия, модели, прототипи); правят доказателства, обобщения и изводи; съобщават и презентират идеите и откритията си пред други хора; поставят за разглеждане нови въпроси и проблеми.

### **(5) Проверка и оценяване**

Проверяване и оценяване на резултатите от работата по проекти включва:

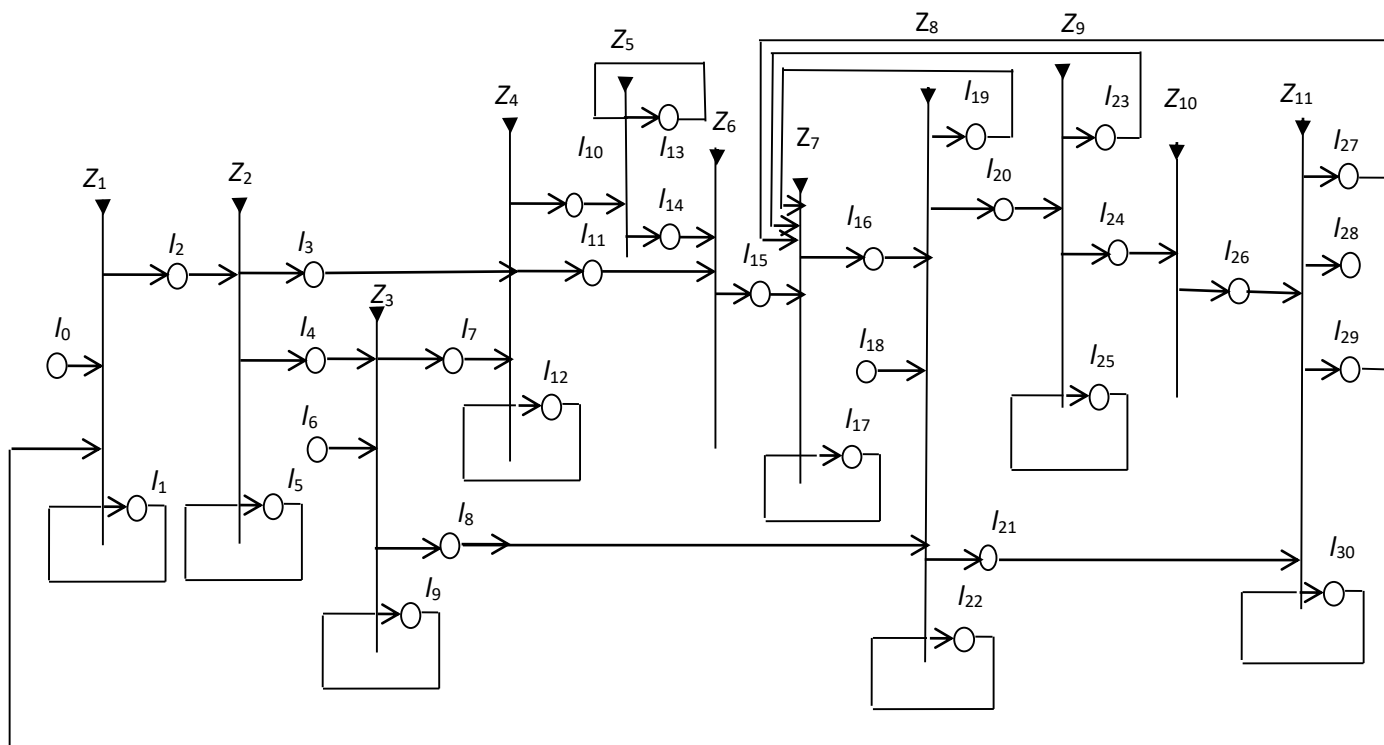
- определяне какво точно ще се проверява;
- определяне чрез какви средствата ще се извърши проверката;
- коректно цитирани източници;
- публичната изява като естествен завършек за разработвания проект.

Ако преподавателят използва хранилище с данни, в които е събрана информация за минали проекти, то след приключване на работата по текущите проекти, базата от данни трябва да бъде обновена. Натрупваната информация може да се използва за набиране на статистически данни за уменията на студентите и за обучаваната група като цяло.

### **Моделиране на процеса на обучение чрез работа по проекти**

В обобщеномрежовия модел, представен на фиг. 27, задачите (проектите) се интерпретират от  $\alpha$ -ядра, преподавателите са интерпретирани от  $\beta$ -ядра, обучаемите (обектите на обучение) от  $\gamma$ -ядра, критериите за оценка съответно чрез  $\delta$ -ядра [65]. Този модел е минимален редуциран обобщеномрежов модел.





Фигура 27. OM модел на процесите в проектно-базирано обучение

OM моделът се състои от 11 прехода със следното значение:

- $Z_1$  = Подготовка за проектно-базираното обучение
- $Z_2$  = Стартира конкретно проектно-базираното обучение
- $Z_3$  = Формиране на екипи от студенти
- $Z_4$  = Определяне на задачите за всеки екип
- $Z_5$  = Дефиниране на подзадачите на всеки проект
- $Z_6$  = Разпределение на подзадачите по участниците в екипа
- $Z_7$  = Работа по конкретна подзадача на проекта
- $Z_8$  = Процес на работа по цялостния проект
- $Z_9$  = Завършване на работата по проекта
- $Z_{10}$  = Представяне и защита на проекта
- $Z_{11}$  = Оценяване на проектите

В дисертационния труд е дадено подробно описание на отделните преходи.

**ИЗВОД:** Разгледани са характеристиките и възможностите на проектно-базираното обучение в допълнение на традиционните педагогически подходи. Изгражда се общомрежов модел на процеса на обучение, посредством разработване на проект [65]. Моделът служи за анализиране на възможностите за работа по адаптирани към знанията и уменията на студентите проекти, чието разработване води до интегриране на знания, придобивани в различни курсове от обучението. Интеграцията на проектно-базираното обучение с възможностите на електронното и уеб-базираното обучение правят процеса на учене по-привлекателен и по-разнообразен за студентите.

#### 4.5. Модел за игровизация на електронен курс за обучение

Игровизацията на обучението има за цел да интегрира игрови елементи и техники в процеса на е-обучение, за подпомагане усвояването на учебното съдържание. Тези игри имат образователни, тренировъчни или информационни цели. Игровизацията предлага учебният процес да бъде организиран в игрови нива (заклучени или отключени с входни изисквания), които представят различни секции с учебни ресурси и дейности за преминаване. Курсът на обучение се въвежда с интересен контекст (сюжет/история) на дейностите, които трябва да бъдат извършени. Правилата за обучение се въвеждат като игрови правила. Всяко ниво включва предизвикателства - учебни дейности, които трябва да бъдат осъществени за постигане на учебните цели на нивото. Някои от заданията са индивидуални, а други съвместни, където обучаемите работят в екипи.

За завършване на някои дейности обучаемите получават очаквани бонуси (например точки). За постигане на набор от изисквания обучаемите могат да получават значки: (Champion), Суперзвезда (Superstar), Авантюрист (Adventurer), Изследовател (Explorer) и т.н. За отлични резултати или за извършване на конкретна дейност, могат да получат неочаквани награди (допълнителна интересна информация, точки/оценка, материални награди (виртуални предмети) и т.н.). За специфична завършена дейност могат да се получават предимства като комбо (помощ, препоръки, по-подробни примери, удвояване на точки от дадена дейност и т.н.). Някои от учебните елементи могат да са заключени (скрити съкровища) и да бъдат отключени, когато обучаемите отговорят на определени изисквания.

Обучаемите могат да участват в игровизирания курс с конкретна игрова роля, която има и визуално представяне чрез някакъв образ (аватар). Въз основа на събраните точки и текущото игрово ниво обучаемите се подреждат в класацията, където могат да бъдат видени водещите участници в учебния процес. По време на целия процес обучаемите имат достъп до информация за своя напредък в обучението като прогрес в играта и във всеки текущ момент за своя статус (достигнато игрово ниво, спечелени точки, значки и други награди).

Разработени са редица модели за проследяване на процеса на обучение и резултатите на обучаемите [74, 78, 99]. Използването на методи на изкуствения интелект дава възможност за предвиждане на резултатите и частична автоматизация на процеса на оценяване. Обобщеномрежови модел за процесите по избор и изграждане на подходяща E-Test система е даден в [59]. Нов подход за формиране на текущи оценки за знанията и уменията на обучаемите, базирани на размити множества е представен в [39]. Настоящия модел обобщава тези модели и ги разширява с възможности за прилагане на модули за игровизация на обучението и проследяване на обучаемите при работата им с игровизиран курс.

Създаденият обобщен модел на процеса на игровизация на обучаващ курс е минимален редуциран обобщеномрежови модел [69] и съдържа множеството от преходи:

$$A = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6\}, \text{ където:}$$

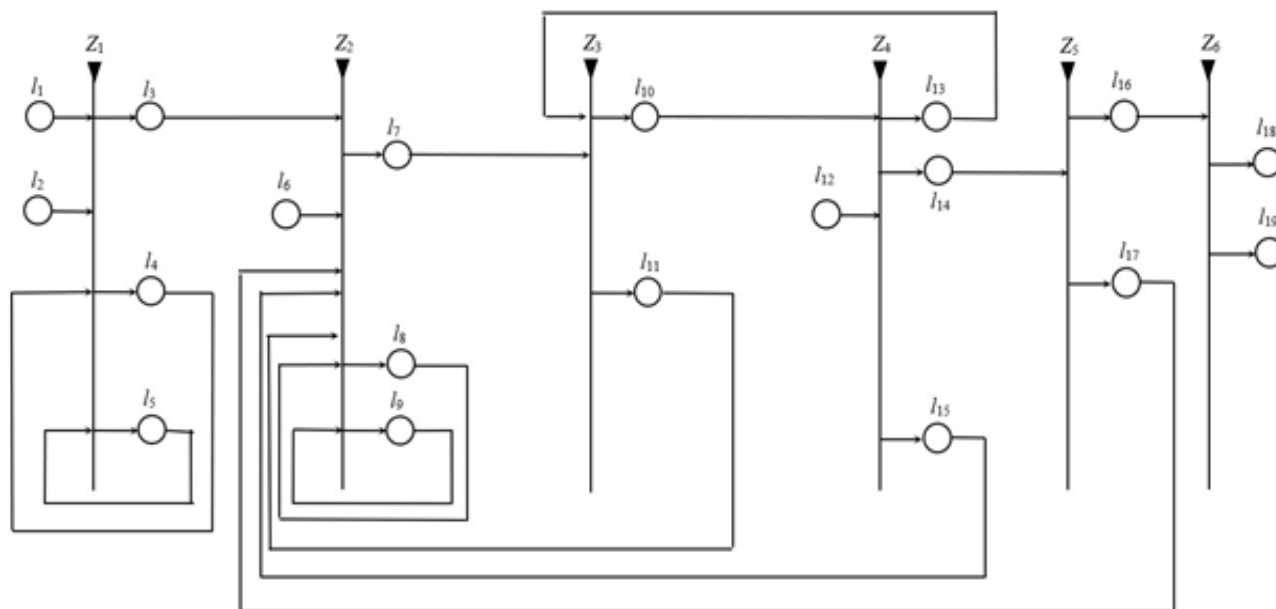
- $Z_1$  – Създаване на модел и игровизация на обучението;
- $Z_2$  – Дефиниране на правила за обучението в игровизиран курс;
- $Z_3$  – Провеждане на обучение с електронния курс;
- $Z_4$  – Оценяване на обучаемите и награждаване;
- $Z_5$  – Класиране и анализ на резултатите на обучаемите;
- $Z_6$  – Анализ на обучението и оценяване на курса.

Следните ядра се използват за описание на процесите:

- $\alpha$ -ядро – потребители на обучаващата среда (администратори, преподаватели, студенти);
- $\beta$ -ядро – курсове и инструменти за игровизация (плагини, софтуерни компоненти и др.);

- $\gamma$ -ядро – оценъчни модели;
- $\eta$ -ядро – данни за обучаваните обекти по време на обучението;
- $\kappa$ -ядро – специфични данни за игровизираните модули за обучаваните обекти;
- $\mu$ -ядро – правила и критерии за прилагане на игрови елементи и награждаване.

Обобщеномрежовият модел на процеса на игровизация на обучаващ курс е представен на фиг. 28. В дисертационния труд е дадено подробно описание на отделните преходи.



Фигура 28. Обобщеномрежови модел на процеса на игровизация на електронен обучаващ курс

Игровизацията на обучението, се разделя на два основни типа: структурна промяна на обучаващ курс и съдържателна промяна на обучаващ курс.

При *структурната промяна* обучаемите преминават през стандартните учебни ресурси, но към курса се включват различни игрови елементи като: точки, значки, класации на първенците, аватари, медали и награди и др.

При *съдържателна промяна* на обучаващ курс се използват игрови техники за представяне на учебното съдържание под формата на правила, нива и умения. Например:

- Правилата на учебния процес могат да бъдат разгледани като правила на игра;
- Мисия/ Предизвикателство/ Приключение – могат да бъдат всички учебни активности, които обучаемият трябва да осъществи в рамките на учебния курс и към тях може да бъде добавен игрови сюжет, който да описва целта на мисията;
- Скрито съкровище - скрити учебни ресурси, които могат да бъдат открити/отворени само при постигане на определени условия (например при завършване на мисия);
- Сюжет / История – учебния процес, може да описва история с мисии за изпълнение.

Игрите притежават силен мотивиращ механизъм за постигане на по-ефективно обучение. Примери за такива среди са: GENIE, The Knowledge Arcade, TalentLMS, Frog, Expertus One, Accord LMS, Axonify, etc. [6, 44]. Тези системи, не са свободно достъпни, а се предлагат като платени софтуери. Средата Мудъл предоставя възможности за създаване на персонализирана учебна среда и доставя набор от инструменти за поддръжка с отворен код, като модули за игровизация могат да се използват: LevelUp, Ranking block и Stash.

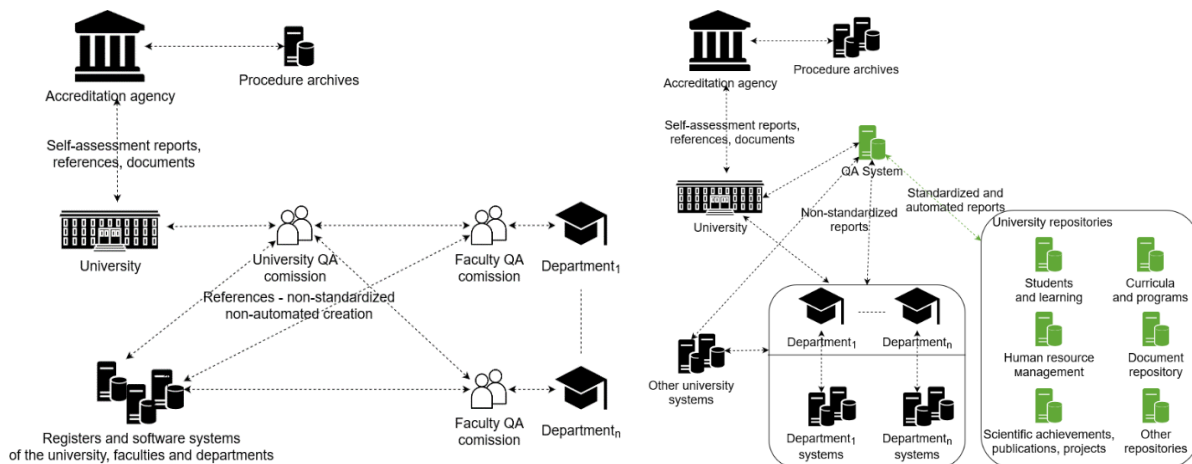
Създаденият обобщеномрежов модел на процеса на игровизация на обучаващ курс предлага решение на един от проблемите, формулиран в книгата на акад. Красимир Атанасов „Generalized Nets and Intuitionistic Fuzziness in Data Mining“ [34].

#### 4.6. Модели за осигуряване на качеството и акредитация на висшето образование

Осигуряването на качеството във висшето образование е непрекъснат процес, който изисква много усилия от страна на образователните институции и акредитиращите организации. Основните форми са две: **вътрешно и външно осигуряване на качеството**. Висшите училища изграждат **вътрешни системи за осигуряване на качеството**, като установяват собствени правила и критерии за оценяване на образователния процес. Те подлежат и на **външно оценяване – акредитация, което се извършва от лицензирани национални и международни агенции**.

Необходимостта от съизмеримост на качеството на висшето образование в Европейския съюз мотивира дефинирането на единни стандарти за качество. През 2005 година са разработени Standards and Guidelines for Quality Assurance in the European Higher Education Area (ENQA). Документа (ESG) [18] съдържа множество стандарти и насоки за прилагането им. България е една от страните, които прилагат стандарта ESG. Осигуряването на качеството на висшето образование в България се извършва от Националната агенция за оценяване и акредитация (НАОА)[2]. **Процедурите за осигуряване на качеството** протичат по обща схема и включват дейностите: самооценяване, посещение на експертна група, оценяване и следакредитационен контрол.

**Процедурите за оценяване и акредитация на висшето образование трябва да бъдат естествена и неразделна част от образователния процес и да се провеждат без много усилия, незабележимо за студенти и преподаватели.** За целта, е необходимо образователните институции да поддържат хранилища с актуална информация за дейността си, а акредитиращите – да осигурят софтуерни инструменти за автоматизирана оценка на качеството, съгласувани с информацията в националния център за информация и документация (НАЦИД). Модели на текущото състояние са създадени в [40, 43]. Обобщен модел на състоянието е даден на фиг. 29.



Фигура 29. Модел на процесите при изготвяне на доклад-самооценка от висшите училища

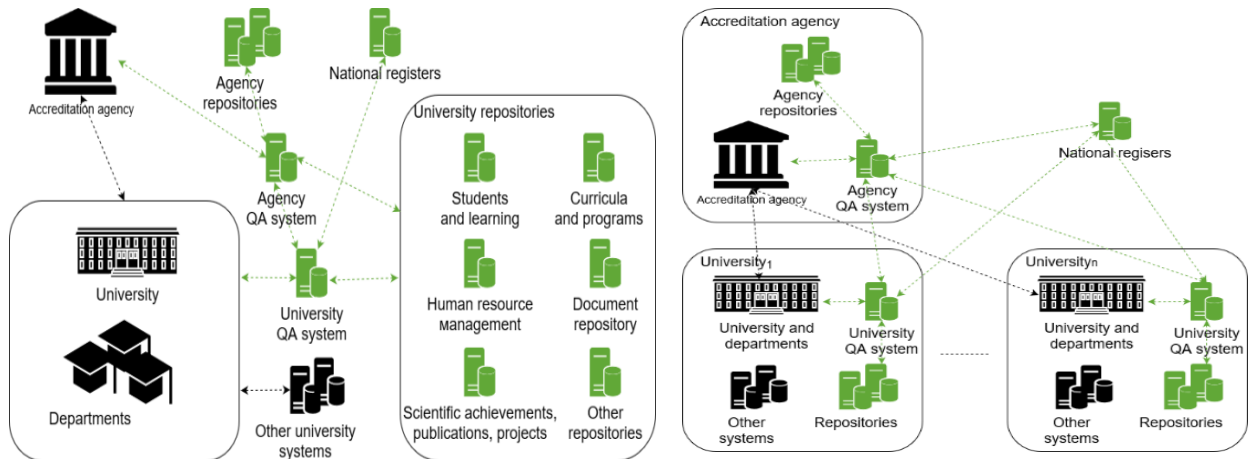
а) Модел без университетска софтуерна QA система за качество

б) Модел с университетска софтуерна QA система за качество

При множество висши училища липсва система, подпомагаща създаването на доклад-самооценка (фиг. 29а). Разпределянето на задачите по изготвянето му става йерархично: НАОА открива процедура за акредитация и осведомява университета; комисията по акредитация разпределя задачи по изготвяне на справки и подготовка на текстове за отделните стандарти; звената събират информация от различни регистри, университетски софтуерни системи, а също и не дигитализирана информация. Създаването на доклада-самооценка е трудно, поради необходимостта от обобщаване на предоставените в различни формат данни и привеждането на текстовете в общ документ. Наличието на инструменти, подпомагащи дейността (фиг. 29б) предоставя възможности за: контрол и проследяване на дейностите; автоматизация на достъпа до справки и др.

#### 4.6.1. Слабо-централизиран модел за акредитация

Един възможен подход за автоматизация на акредитационните процедури е чрез *изграждане на централизирана система на акредитационната агенция и НАЦИД и самостоятелни университетски системи за осигуряване на качеството* (фиг. 30). Всяка университетска QA система автоматизирано извлича необходимите акредитационни данни от университетските хранилища [45]. Друга част от информацията се въвежда от университетски отговорници. В определен момент, информацията от университетската QA система се експортира към QA системата на акредитиращата институция.



Фиг. 30. Слабо-централизиран модел за акредитация

а) Основни системи и комуникационни интерфейси в процеса на акредитация на едно висше училище

б) Основни системи и комуникационни интерфейси в процеса на акредитация в системата на висшето образование

Предимствата на подхода определят и неговите недостатъци. Създаването на собствено решение или поддръжката на стандартизирано приложение от страна на университетите изисква наличието на специализирани ИТ специалисти във всяко висше училище. Създаването на стандартизирана университетска QA система е трудоемка задача, която трябва да се съобрази с наличието на множество различни университетски хранилища и други софтуерни системи.

#### 4.6.2. Силно-централизиран модел за акредитация

Силно-централизираният модел предлага *използването на централизирана cloud система за акредитация*. Системата поддържа модул както за НАОА и НАЦИД, така и конфигурируеми модули, обслужващи акредитационните процеси за отделните висши училища. Използване на единна система предполага улеснена интеграция между основните модули, лесна за потребителите актуализация на софтуера, интеграция на софтуерни разширения, създаване на динамични анализи и справки върху данните, предоставени от всички институции [40].

Основна цел на описаните модели е предлагане на *платформа*, улесняваща работата на всички участници в процедурите по оценяване и акредитация – от страна на обучаващите, и от страна на акредитиращите институции. Използването на софтуерни системи, базирани на предложените модели има много предимства: централизиране на процесите по оценяване, което улеснява тяхното администриране и мониторинг; осигуряване на обща платформа за екипна асинхронна работа на потребители с различни права и отговорности; възможност за работа по всяко време и от всяко място; улеснена комуникация и прозрачност; автоматизирано генериране на документация; оптимизиране на времеви и материални ресурси; намаляване вероятността от допускане на грешки, и др.

#### 4.6.3. Обобщеномрежов модел на процесите на обработка на данни при самооценяване във висше училище

Модел на централизирана система за осигуряване на качеството на висшето образование е създаден в [42]. Обобщеномрежови модел на система за осигуряване на качеството във висше училище е представен в [41]. Тук се създава минимален редуциран обобщеномрежови модел, описващ процесите на обработка на данни при самооценяваща процедура във висшето учебно заведение, който съдържа следния набор от преходи:

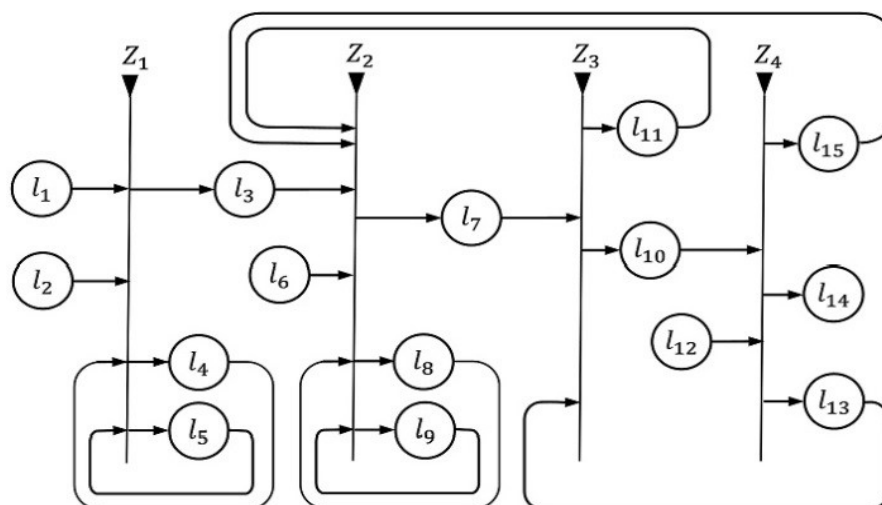
$$A = \{Z_1, Z_2, Z_3, Z_4\},$$

където:

- $Z_1$  – Събиране на данни от различни източници;
- $Z_2$  – Интегриране на данни;
- $Z_3$  – Обработка на данни за изискванията, свързани с критериалната система;
- $Z_4$  – Изготвяне на доклад за самооценка и съответните приложения.

Следните ядра се използват за описание на процесите:

- $\alpha$ -ядра – данни за дейността на висшето училище;
- $\beta$ -ядра – членове на административния и академичен персонал, членове на комисията по качеството и групата за изготвяне на доклад за самооценка;
- $\mu$ -ядра – критерии и критериална система на акредитиращата институция;
- $\eta$ -ядро – хранилище на данни с актуална информация за дейността на висшето училище, необходима за изготвяне на доклад за самооценка;
- $\gamma$ -ядра – инструменти за обработка на данни;
- $\sigma$ -ядра – заявки за данни от потребител/приложение.



Фигура 31. OM модел на процесите на обработка на данни при самооценяване във ВУЗ

**Обобщеномрежовият модел**, описващ процесите при обработка на данни при самооценяваща процедура във висшето учебно заведение е представен на фиг. 31. В дисертационния труд е дадено подробно описание на отделните преходи.

**ИЗВОД:** Предложен е формален модел, описващ процесите на събиране и обработката на данни по процедура за самооценка във висше учебно заведение. Този модел може да бъде развит и чрез използване на **йерархичен оператор** (НЗ от теорията на OM), който замества даден преход или позиция с подмрежа, която има същото, но описано по-подробно поведение на елементите. Въз основа на създадения обобщеномрежов модел и събраните данни от реални процеси могат да бъдат открити поведенчески модели и анализ на производителността на отделните компоненти.

## ГЛАВА 5. РЕШЕНИЯ НА НАУЧНОПРИЛОЖНИ ПРОБЛЕМИ В ОБУЧЕНИЕТО

Резултати от приложения на направените изследвания и анализи в образователна среда са представени в няколко авторски публикации. Методически техники за въвеждане на базови концепции в курсове за студенти, адаптирани към знанията и специалностите на обучаемите са предложени в [64, 81]. Как да следваме съвременните тенденции в курсовете за анализ на данни, чрез решаване на подходящо подбрани реални проблеми и задачи се обсъжда в [67]. Подход за представяне на концепции и стандарти за информационна сигурност във висшето образование се предлага в [73]. Програмиране с реактивни блокове се представя в [66] за курсове по Интернет на нещата. В [98] се анализират трудностите при обучението по дедуктивни методи за проверка и синтез на програми. Тенденции в обучението на софтуерни инженери се представят в [34, 35]. В тази глава от дисертационния труд се обобщават резултати от тези публикации и се предлагат основни елементи на учебни програми и подходи за въвеждане на базови модули и инструменти на науката за данните във Висшите училища.

### 5.1. Учебна програма за анализ на данните

Всяко направление и конкретна специалност се фокусират върху целите, които си поставят с обучението на студентите. Приложните специалности акцентират върху способността на обучаваните да могат да избират подходящ софтуер за анализ на данни. В други специалности акцентът е обучаемите да усвоят основни алгоритми за анализ на данни, машинно обучение и изкуствен интелект. Студентът трябва да може да конструира модели за анализ на съществуваща ситуация и бъдеща прогноза, да се научава как да използва различни техники на изкуствения интелект с цел откриване на аномалии и създаване на оптимални модели. В най-амбициозния план, след завършване на обучението студентът трябва да може да **управлява целия цикъл от данни и да извлича от тях знания с помощта на методи за събиране и анализ на данни и използване на подходящи алгоритми и системи за управление на данните.**

Отчитайки спецификата на обучението в направлението и специалността, учебната програма на дисциплината, свързана с анализ на данните може да акцентира върху различни аспекти. Следва предложение на **базови модули**, които могат да намерят различно място в обучението по дисциплината за анализ на данни (независимо от конкретното наименование).

**(1) Предварителна подготовка.** Студентите трябва да са преминали основните курсове по програмиране, математика, теория на вероятностите и статистически методи, бази от данни, изкуствен интелект, дискретна математика, структури от данни и комуникация, за да се даде възможност на обучаемия да свързва методологиите с въвежданите нови знания по дисциплината.

**(2) Въгъпителна част.** В този модул се представят основите етапи в цикъла на анализ и извличане на данни и да се направи преглед на различни инструменти, които поддържат всеки етап от цикъла. Целта на модула е да помогне на студентите в бъдеще успешно да се реализират като анализатори на данни или да провеждат научни изследвания и развиват академична кариера.

**(3) Основна част на курса.** Този модул трябва да осигури теоретичното обучение с базовите алгоритми и подходи за анализ на данни като: класификационен анализ, регресионен анализ, асоциативен анализ, клъстерния анализ, анализ на шумове и др. Основен акцент в процеса е прилагането на съответен Data Mining алгоритъм или алгоритми, позволяващи получаването на знания, описващи: връзка между свойства на данните, модели на данните, резултатите от класификацията и клъстеризацията на данните и др. Интересно е какви типове зависимости могат да бъдат открити в процеса на извличане на закономерности от данни. За целта се разглеждат шест типа задачи, решаването, на които води до получаване на интересуващите ни зависимости: *описание и обобщение на данни; сегментация (или клъстеризация); анализ на крайностите (outlier detection); описание на понятия (concept description); класификация и регресия; анализ на зависимостите.*

За извличане и представяне на моделите от данни се използват различни математически средства и методи на изкуствения интелект. Основни методи, използвани от техниките за извличане на закономерности от данни са: метод на поддържащите вектори (Support Vector Machines SVM), метод на наивния Бейс (Naïve Bayes), разлагане по метода на главните компоненти (Principal Component Analysis), метод на изкуствена невронна мрежа за анализ на данни (Artificial Neural Networks) и др.

**(4) Практическо обучение** – обучаемите експериментират с извличане на данни, управление на данни и анализ, подходи за визуализиране на данни, чрез прилагане на различни системи с отворен код. В курса следва да се представят популярни системи за обработване на данни и извличане на знания като: *WEKA* [103], *RapidMiner* [87], *KNIME* [52], *Orange* [56] и др.

**(5) Курсов проект.** Студентите с по-слаби умения по програмиране и недостатъчни математически знания изпитват трудности при разбирането на същността на анализа на данните и прилагането на избраната техника. За преодоляване на трудностите се предлага като последен модул включване на проектно-базирано обучение. Курсовите проекти - индивидуални или екипни трябва да са със специфична ориентация към определена бизнес област. Например:

- организационния и бизнес анализ на данните: финансови данни, застраховане, банкиране, управление на инвестиции, управление на риска, пазарни данни и др.;

- анализа на данните за общественото здраве: данни за качеството на живота на човека, изграждане на модели за качествен живот, предотвратяване на бъдещи заболявания и епидемии и формулиране на програми за управление на здравето в общността;

- управлението на публичните институции, болници, държавни институции и др.;

- анализ на данни в интелигентни градове и кибер-сигурност.

Независимо дали целта е да се открият интересни взаимовръзки, да се категоризират обекти в групи, да се оптимизира планирането на ресурси или да се определят тарифи за таксуване, работата по проектите ще помогне на студентите да получат основно разбиране на техниките за анализ на данни и да усвоят умения за извличане на знания с конкретни задачи.

**ИЗВОД:** Развитие на такива академични програми е от решаващо значение за успеха на обучаемите като бъдещи изследователи в дигиталния свят. Интересът на обучаемите и активното им участие в процеса на обучение могат да бъдат повлияни от качеството на обучаващия процес и използваната учебна програма.

## **5.2. Инструменти за извличане на знания чрез примери**

Обучението в курсовете по Базис на данни на студентите от направление *Информатика и компютърни науки* в повечето български университети се провежда през втори или трети семестър. За целта последователно се въвеждат: основни модели на данните; процес на проектиране на релационни бази от данни, анализ и нормализация на релационни схеми, езици за описание и обработка на данни. Студентите се запознават с проблемите при разпределени бази от данни и обектно-ориентирани бази от данни и новите перспективи за развитие на големите данни (BigData). След като реализират редица приложения под формата на курсови проекти, студентите преминават към анализ на хранилища от данни и Data mining техники. Тези техники предлагат богати възможности за изследване на данните, но изучаването им предполага силна математическа основа. Имайки предвид ограниченото време, в курсовете тук се предлага разкриване на възможностите на тези техники, като се предлагат на студентите конкретни примери върху предварително подготвени данни. Студентите могат да анализират данните, следвайки методологията на CRISP-DM модела, като в последователност от шест стъпки се следват шестте основни фази в процеса на извличане на закономерности от данни. За целта се използват подходящо подбрани реални задачи за анализ на данни, които са адаптирани с цел по-лесно прилагане на съответен алгоритъм, подробно представени в дисертационния труд.



ИЗВОД: Анализът на резултатите от проведеното обучение показва, повишаване степента на усвоеност на учебното съдържание и повишаване на интереса към дисциплината [67]. Обучението по тематика не предизвиква сериозни трудности сред студентите, които притежават необходимата предварителна подготовка. Студентите с по-слаби умения по програмиране и не достатъчни математически знания изпитват редица трудности при разбирането на анализа и прилагането на избраната техника на извличане на закономерности от данни.

### 5.3. Въвеждане на онтологии в обучението на студенти

Предлага се метод за въвеждане на онтологии в курсовете на специалностите в направление “Информатика и компютърни науки” чрез реализация на три основни задачи, които се решават в процеса на обучение [64]. Първата задача е свързана с описание на природата на онтологията. Представя се произхода и значението на онтологията и различните видове онтологии. Студентите се запознават с основните компоненти на структурата на онтологията. Втората задача е свързана с представяне на схеми, които реализират онтологии. Тази част от обучението се базира на примери, сравнение между различни примерни онтологии, както и анализ на всяка от тях. За да се направи това, се въвеждат средствата на езика OWL. Третата задача е разработването на пример за OWL онтология, с помощта на софтуера Protégé.

Въвеждането на понятието „онтология“ започва със запознаване на неговия произход от философията и отразява природата на нещата, които реално съществуват. В съвременната литература този термин се използва за **означаване на определена система от категории, които са следствие от определена система от възгледи за света**. Представя се тяхна класификация в зависимост от различни класификационни признаци. В [38] се разглежда класификация, според която онтологията се разделят на: общи онтологии; онтологии, ориентирани към конкретна област; онтологии, ориентирани към конкретна задача и приложни онтологии.

По отношение на структурата на онтологията явно се отделят два компонента: имена на съществуващи концепции и връзки в домейна. Редица ограничения могат да бъдат наложени над домейните. Онтологията заедно с множество конкретни инстанции на класовете съставя база знания (knowledge base). Така разработването на една онтология преминава през следните фази [36]: дефиниране на класовете в онтологията; изграждане на йерархия от класовете; дефиниране на характеристики и описание на възможните стойности за тях; попълване на конкретни стойности за характеристиките, за да се получат конкретни инстанции.

Разработването на онтология е итеративен процес. Основно правило е концепциите в онтологията да са близки до обектите и връзките в конкретния домейн. При построяване на дадена онтология предварително трябва да се определи – с каква цел се създава онтологията, за какви типове въпроси се предполага да бъдат получавани отговори (с нейна помощ), как ще се използва и поддържа. Обучението продължава с изучаване на: методологията за проектиране на онтология; изучаване на Web Ontology Language (OWL) и изграждане на примерна онтология; използване на Reasoner за работа с онтологии. За целите на обучението се разработва пример за OWL онтология със софтуера Protégé.

В резултат се постига въвеждане на онтологията, чрез разясняване на базовите понятия и основните принципи в областта, базирайки се на примери. Разработва се съвместно със студентите примерна онтология, преминавайки през всички стъпки в процеса на създаването. Предоставят се възможности за допълване на изградената онтологичната система – с нови понятия, отношения или атрибути.

## ЗАКЛЮЧЕНИЕ

Извличането на данни в образователни среди (Educational Data Mining) и анализът на обучението (Learning Analytics) са сравнително нови области, които имат за цел да подобрят образователния опит, да помагат на заинтересованите страни (преподаватели, студенти, администратори, изследователи), да вземат по-добри решения, използвайки натрупаните данни. Въпреки големите очаквания и нарасналия обем на публикациите в областта на извличането на данни в образователни среди, все още стоят бариери и предизвикателства пред изследователите в областта, като **липсата на изчерпателни и лесни за използване и разбиране инструменти, които могат да бъдат интегрирани в най-популярните среди за управление на обучението.**

Съвременното обучение използва софтуерни платформи за управление на учебното съдържание, които осигуряват технологичен апарат, паралелно с публикуването на учебното съдържание да се събира и съхранява информация и за активността на потребителите – преподаватели и обучаеми. Насока за бъдещо развитие на работата е създаване на лесни за ползване софтуерни инструменти, с възможности за интегриране в средите за управление на обучението за ранното откриване на обучаеми в риск и своевременно уведомяване на преподавателите кои обучаеми се нуждаят от допълнителна помощ и кои практики на преподаване оказват най-голямо въздействие. Освен това трябва да продължат изследванията относно възможностите за персонализиране на учебния процес и съдържание, така че всеки обучаем да получава ресурси според текущите си познания и нагласа за обучаващия процес.

Предизвикателство е разработване на адаптивни курсове, които се персонализират автоматично според профилите на обучаемите (нужди, цели, произход, държава, стил на учене и т.н.). В дисертационния труд фокусът е върху наблюдение и идентифициране на обучаеми в риск, за които има вероятност да отпаднат или да се провалят по време на обучението. Но проведените анализи на данни са върху сравнително малки групи обучаеми. По-добри резултати могат да бъдат получени при анализиране на голям брой студенти, курсове и институции. Например при MOOCs (Massive Open Online Courses) могат да бъдат използвани данни от хиляди студенти. Участниците имат различни произход, зрялост, опит, нива на образование, езикови умения, цели, нужди и стилове на учене. Но значителна бариера, [37] е свързана с **етиката при използване на личните данни, която трябва да се взема предвид на всички етапи - от събирането на данни до тълкуването на резултатите и вземането на решения.** Например данни свързани с пол, социален статус, раса, религиозни вярвания, идеология или увреждане биха могли да доведат до дискриминационно отношение. От друга страна трябва да се вземе предвид и поредицата мерки, свързани със защитата на личните данни.

Следваща стъпка на изследванията в областта е създаване и прилагане на **методологии за използването на Big Data Analytics в образователни среди.** В ерата на големите данни, възможностите за съхраняване, управление и обработка на данни от онлайн среди за обучение дават възможност да се изследват по-добре процесите на обучение и да се търсят ефективни начини за тяхното подобряване. **Комбинацията от Big Data и анализът на обучението (Learning Analytics) е обещаваща област на изследване.**

## Приноси на дисертационния труд

### А. Научни приноси (в класа обогатяване на съществуващите знания).

Предложени са методи и модели, като резултат от теоретични обобщения на процесите на наблюдение и анализ на дейностите на обучаемите:

1. Петстъпков метод за оценяване и прогнозиране на знанията, уменията и компетентностите на обучаемите във виртуалното образователно пространство, с възможност да се прилагат изчислителни модели и критерии за динамично оценяване, а окончателната оценка се формира от множество оценъчни компоненти. Проведените емпирични изследвания потвърждават, че предложеният метод е перспективен за разработване на система за ранно предупреждение за различни заинтересовани страни от обучаващия процес.
2. Модели чрез размита логика на йерархично многокомпонентно оценяване на различни мисловни умения от висок и нисък порядък, теоретични знания и практически умения: *Модел 1* на йерархична дървовидна организация на оценъчните компоненти и *Модел 2* на йерархична графова (недървовидна) организация на оценъчните компоненти. Представените модели са приложими при различни подходи за оценяване.
3. Метод с уеб метрики и индуктивна размита класификация за оценка на степента на използване на уеб-сайтове от обучаемите, при анализ на поведението им в средата и уеб пространството.
4. Метод за анализ на типа на документ, на базата на класификационни алгоритми, като целевия атрибут е дали документа е свързан с областта на обучение. Анализът на данните дава възможност да се търсят трудностите, които обучаемите срещат при работа с литературни източници и да се проектират модули, които да отговарят на индивидуалните им нужди при търсене и използване на документи.
5. Трифакторен модел на обучаемия, който включва: фактори на компетентност, емоционални фактори, фактори на въздействие на социалната среда. Чрез моделът на обучаемия може да се проследи промяната на знанията, уменията и компетентностите и да се прави предвиждане, според натрупаните данни от работата на среди и системи за електронно обучение.
6. Обобщеномрежови модели: с възможност за проследяване на процесите на използване на различни инструменти в среди за електронно обучение; на процеса на прилагане на Data Mining средства в среди за обучение; на процеса на многокомпонентно оценяване в шест етапа с възможност за персонализация на начина за формиране на тестове чрез дефиниране на мета-модели, задаващи рамки за създаване на тестове и подходи за оценяването им; на процесите в проектното обучение и възможностите за интеграция с електронното и уеб-базираното обучение; за игровизация на E-learning курс, за анализиране на възможностите и проблемите при обучение с игрови ситуации. Създадените модели предлагат частично решение на проблемите, формулиран в книгата на акад. Кр. Атанасов „Generalized Nets and Intuitionistic Fuzziness in Data Mining“ [9].
7. Слабо-централизиран и силно-централизиран модел за осигуряване на качеството и акредитацията във висшето образование. Базов обобщеномрежов модел на информационните потоци при обработката на данни при самооценяване във висшето училище.

### Б. Научно-приложни приноси (в класа приложение на научните постижения в практиката)

1. Многостъпкови методи за създаване на модели в образователен контекст чрез инструментите за машинно обучение на системата *Orange Data Mining*.
2. Софтуер на инструменти за анализиране и сравняване на поведението на два агента за машинно обучение: *Rule-Based System* и *Reinforcement Learning*.
3. Софтуерни инструменти за анализ на звукови честоти и преобразуването им в цветове в RGB модела. Модел на система, която реализира преобразуване на звуковите честоти в цвят.
4. Основни модули на учебна програма на науката за данните в дисциплини на Висшето училище.
5. Методически техники за обучение на студенти с инструменти за извличане на знания чрез примери от реални проблеми и задачи за проектиране на онтологии в обучението на студенти.

**Списък на научните публикации по темата на дисертационния труд  
в издания, които са реферирани и индексирани в световноизвестни бази данни с научна  
информация (Web of Science и Scopus)**

1. **Orozova D.**, Appropriate E-Test System Selection Model, Comptes rendus de l'Academie bulgare des Sciences, 2019, Vol 72, No 6, pp. 811-820, ISSN 1310-1331, DOI: 10.7546/CRABS.2019.06.14, **IF=0.343** (2019), **WoS Q4**, SJR=0.218, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000477784300014>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85070988936&partnerID=MN8TOARS>

2. **Orozova D.**, K. Atanassov, Generalized Net Model of Processes Related to Big Data, Comptes rendus de l'Académie bulgare des Sciences, book No 12, vol. 71, 2018, pp. 1679 -1686 ISSN 13101331, DOI: 10.7546/CRABS.2018.12.13, **IF=0.321** (2018), **WoS Q4**, Scopus SJR=0.205, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000456750000013>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85063227040&partnerID=MN8TOARS>

3. **Orozova D.**, K. Atanassov, Model of Big Data Map/Reduce Processing, Comptes rendus de l'Academie bulgare des Sciences, 2019, Vol 72, No11, pp.1537-1545, ISSN 1310-1331, DOI: 10.7546/CRABS.2019.11.11, **IF=0.343** (2019), **WoS Q4**, SJR=0.218, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000502809100011>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85109014572&partnerID=MN8TOARS>

4. **D. Orozova**, K. Atanassov, M. Todorova, Generalized Net Model of the Process of Personalization and Usage of an e-Learning Environment, Proceedings of the Jangjeon Mathematical Society 19 (2016), No. 4, pp. 615 – 624. ISSN (print):1598-7264, ISSN (online):2508-7916, ISBN:89-87809-15-3, SJR=0.508, (Scopus Q2)

<http://www.scopus.com/inward/record.url?eid=2-s2.0-84996565258&partnerID=MN8TOARS>

5. **Orozova D.**, M. Todorova, Ontology concept in courses on students, TEM Journal. Volume 7, Issue 3, Pages 693-697, ISSN 2217-8309, DOI:10.18421/TEM73-29, 2018. SJR=0.148. (Scopus Q4)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000442699500029>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85052218825&partnerID=MN8TOARS>

6. Popchev I., **D. Orozova**, Towards Big Data Analytics in the E-learning Space. Cybernetics and Information Technologies, Vol. 19(3), 2019, pp.16-25, ISSN: 1311-9702, DOI: 10.2478/cait-2019-0023, SJR=0.31, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000486349500002>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85073873211&partnerID=MN8TOARS>

7. **Orozova D.**, K. Kaloyanova, M. Todorova, Introducing Information Security Concepts and Standards in Higher Education, TEM Journal. Volume 8, Issue 3, Pages 1017-1024, ISSN 2217-8309, DOI: 10.18421/TEM83-46, 2019, SJR=0.167, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000483978200046>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85072752036&partnerID=MN8TOARS>

8. Popchev I., **D. Orozova**, Towards a multistep method for assessment in e-learning of emerging technologies, Cybernetics and Information Technologies, Volume 20, No 3, Sofia, 2020, pp.116-129, ISSN: 1311-9702, DOI: 10.2478/cait-2020-0032, SJR=0.272, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000572532500009>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85093080750&partnerID=MN8TOARS>

9. Hadzhikolev, E., S. Hadzhikoleva, K. Yotov, **D. Orozova**, Models for Multicomponent Fuzzy Evaluation, with a Focus on the Assessment of Higher-Order Thinking Skills, TEM Journal, Vol.9, No.4, pp.22-28, 2020, ISSN: 2217-8309, DOI: 10.18421/TEM94-43, SJR=0.199, (Scopus Q4)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000596297300043>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85092629596&partnerID=MN8TOARS>

10. Todorov, J., V. Valkanov, S. Stoyanov, B. Daskalov, I. Popchev, **D. Orozova**. Chapter 6: Personal Assistants in a Virtual Education Space, Sgurev V., Jotsov V., Kacprzyk J. (eds) Practical Issues of Intelligent Innovations. Studies in Systems, Decision and Control, 140, Springer International Publishing AG part of Springer Nature, 2018, Library of Congress Control Number 2011893667, ISSN 2198-4182, DOI: 10.1007/978-3-319-78437-3\_6, SJR=0.131, (Scopus Q4)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000466552900007>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85049662711&partnerID=MN8TOARS>

11. Todorov J., E. Doychev, **D. Orozova**, A. Stoyanova-Doycheva, IoT Multiagent Assistant in Virtual Educational Space, 2018, AIP Conference Proceedings Volume 2048, pp. 020031; ISSN: 0094243X, ISBN: 978-073541774-8, doi: 10.1063/1.5082049, <https://doi.org/10.1063/1.5082049>, SJR=0.182.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000468108800031>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85058786246&partnerID=MN8TOARS>

12. Stoyanova-Doycheva A., T. Glushkova, S. Stoyanov, **D. Orozova**, Lifelong Learning Supported by an Intelligent Tourist Guide, 2018, AIP Conference Proceedings, Volume 2048, pp. 020038; ISSN: 0094243X, ISBN: 978-073541774-8, SJR=0.182, DOI: 10.1063/1.5082056

<https://www.webofscience.com/wos/woscc/full-record/WOS:000468108800038>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85058819893&partnerID=MN8TOARS>

13. Hadzhikolev E., S. Hadzhikoleva, **D. Orozova**, K. Yotov, A Comprehensive Approach to Assessing Higher and Lower Order Thinking Skills, Uncertainty and Imprecision in Decision Making and Decision Support: New Advances, Challenges, and Perspectives, Lecture Notes in Networks and Systems, Volume 338, pp. 164–177, Lecture Notes in Networks and Systems, Volume 338, pp. 164–177, ISSN 2367-3370, 2022, DOI: 10.1007/978-3-030-95929-6\_13, SJR=0.151, (Scopus Q4)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000775291100013>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85126177886&partnerID=MN8TOARS>

14. Hadzhikoleva S., **D. Orozova**, N. Andonov, E. Hadzhikolev, Generalized Net Model of a System for Quality Assurance in Higher Education, 45th International Conference Applications of Mathematics in Engineering and Economics (AMEE'19), 2019, pp. 040005-1 -040005-8, ISBN: 978-0-7354-1919-3. AIP Conference Proceedings, 2019, DOI: 10.1063/1.5133515, Part of ISSN: 15517616 0094243X, SJR=0.19.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000521744400033>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85075796767&partnerID=MN8TOARS>

15. Orozov B., **D. Orozova**, An AI application in E-Learning, CEUR Workshop Proceedings 3061, pp. 76-80, September 27–28, 2021, Plovdiv, Bulgaria, ISSN 1613-0073, SJR=0.228 (3a 2021).

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85122281151&partnerID=MN8TOARS>

16. Ivan Popchev, **Daniela Orozova**, Algorithms for Machine Learning with Orange System, International Journal of Online and Biomedical Engineering (iJOE), Vol. 19, No. 04, 2023, pp. 109 -123, ISSN: 2626-8493, SJR=0.28, (Scopus Q3), DOI: <https://doi.org/10.3991/ijoe.v19i04.36897>.  
<https://www.webofscience.com/wos/woscc/full-record/WOS:000967834800007>  
<https://www.scopus.com/record/display.uri?eid=2-s2.0-85152704594&origin=resultslist&sort=plf-f>
17. **D. Orozova**, N. Hristova, An application of analytical data research in e-learning system, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering Volume 1031 (2021) 012058, IOP Publishing, Online ISSN: 1757-899X, DOI: 10.1088/1757-899X/1031/1/012058.  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85101662489&partnerID=MN8TOARS>
18. Popchev I., **D. Orozova**, DataScience: Experience and Trends, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering, Volume 1031 (2021) 012057, IOP Publishing, Online ISSN: 1757-899X, Print ISSN: 1757-8981, doi:10.1088/1757-899X/1031/1/012057.  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85101718405&partnerID=MN8TOARS>
19. M. Todorova, **D. Orozova**, Training Difficulties in Deductive Methods of Verification and Synthesis of Program, International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 9, No. 7, 2018, pp. 18-22, ISSN ISSN: 2158107X ISSN (online):2156-5570, doi:10.14569/IJACSA.2018.090703.  
<https://www.webofscience.com/wos/woscc/full-record/WOS:000441330600003>  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85054031652&partnerID=MN8TOARS>
20. **D. Orozova**, SQL User Queries Execution Model, Advances in Intelligent Systems and Computing book series (AISC, volume 1081), Uncertainty and Imprecision in Decision Making and Decision Support: New Challenges, Solutions and Perspectives, Volume 1081 AISC, 2021, Pages 184-189, ISSN 2194-5357, DOI: 10.1007/978-3-030-47024-1\_20  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85087754542&partnerID=MN8TOARS>
21. Orozov B., **D. Orozova**, Rule Based System Against Reinforcement Learning, CompSysTech'21: ACM International Conference Proceeding Series on Computer Systems and Technologies '21, June 2021, Pages 67–70, ISBN: 978-145038982-2, DOI: 10.1145/3472410.3472437.  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85117607164&partnerID=MN8TOARS>
22. **Orozova D.**, M. Georgieva, A model of the process of Big Data with generalized net, 2016, IEEE 8th International Conference on Intelligent Systems, IS 2016 – Proceedings, 7 2016, Article number 7737487, pp.599-603, ISBN: 978-150901353-1, DOI: 10.1109/SIELA.2016.7543030  
<https://www.scopus.com/record/display.uri?eid=2-s2.0-85006035368&origin=resultslist&sort=plf-f>
23. Hadzhikolev E., S. Hadzhikoleva, **D. Orozova**, Digital Model of a Document in a University Document Repository, IEEE XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018, 3-6 June 2018, Bulgaria, ISBN 978-1-5386-3418-9, DOI: 10.1109/SIELA.2018.8447089.  
<https://www.webofscience.com/wos/woscc/full-record/WOS:000454837100037>  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85053840788&partnerID=MN8TOARS>

24. **Orozova D.**, A. Ivanov, Generalized Net Model of Virtual Collaboration Space, IEEE XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018, 3-6 June 2018, Bourgas, Bulgaria, ISBN 978-1-5386-3418-9, DOI: 10.1109/SIELA.2018.8447090.  
<https://www.webofscience.com/wos/woscc/full-record/WOS:000454837100038>  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85053825957&partnerID=MN8TOARS>
25. **Orozova D.**, E. Sotirova, Modeling of a Learning Management System, IEEE XVIII-th International Symposium on Electrical Apparatus and Technologies SIELA 2014, 2014, Bulgaria, pp. 95-98, ISBN: 978-147995817-7, DOI: 10.1109/SIELA.2014.6871879.  
<https://www.webofscience.com/wos/woscc/full-record/WOS:000345744800038>  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-84906751001&partnerID=MN8TOARS>
26. **Orozova D.**, M. Georgieva, A model of the process of Big Data with Generalized Net, Electrical Apparatus and Technologies (SIELA), 2016, IEEE 19th International Symposium, Bourgas, 238-241, ISSN: 978-1-4673-9521-2, ISBN 978-619-160-648-1, DOI: 10.1109/SIELA.2016.7543030.  
<https://www.webofscience.com/wos/woscc/full-record/WOS:000382936800059>  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-84985945018&partnerID=MN8TOARS>
27. I. Popchev, **D. Orozova**, S. Stoyanov, IoT and Big Data Analytics in E-Learning, 2019 Big Data, Knowledge and Control Systems Engineering (BdKCSE), IEEE 2019, pp. 1-5, ISBN 978-172816481-6, DOI: 10.1109/BdKCSE48644.2019.9010666.  
<https://www.webofscience.com/wos/woscc/full-record/WOS:000563467600027>  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85082305981&partnerID=MN8TOARS>
28. **Orozova D.**, I. Popchev, Cyber-Physical-Social Systems for Big Data, IEEE XXI International Symposium on Electrical Apparatus and Technologies SIELA 2020, 3-6 June 2020, Bourgas, Bulgaria, pp. 334 – 337, ISBN 978-1-7281-4345-3, DOI: 10.1109/SIELA49118.2020.9167161.  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85091338714&partnerID=MN8TOARS>
29. **Orozova D.**, N. Hristova, Generalized Net Model for dynamic decision making and prognoses, IEEE XXI International Symposium on Electrical Apparatus and Technologies SIELA 2020, Bulgaria, 2020, pp.330 – 333, ISBN 978-1-7281-4345-3, DOI: 10.1109/SIELA49118.2020.9167077.  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85091339727&partnerID=MN8TOARS>
30. Popchev I., **D. Orozova**, Text Mining in the Domain of Plant Genetic Resources, Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems, 2020, pp. 596-600, ISBN 978-172815456-5, DOI: 10.1109/IS48319.2020.9200174.  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85092747530&partnerID=MN8TOARS>
31. Hadzhikoleva S., **D. Orozova**, E. Hadzhikolev, N. Andonov, Model of a Centralized System for Quality Assurance in Higher Education, Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems 2020, pp. 87-92, ISBN 978-172815456-5, DOI: 10.1109/IS48319.2020.9199951.  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85092696472&partnerID=MN8TOARS>
32. Georgieva P., E. Nikolova, **D. Orozova**, Data Cleaning Techniques in Detecting Tendencies in Software Engineering, MIPRO 2020, IEEE 43rd International Convention, 2020, Croatia, Opatija, pp.1272-1277, ISSN: 2623-8764, DOI: 10.23919/MIPRO48935.2020.9245416.  
<https://www.webofscience.com/wos/woscc/full-record/WOS:000790326400188>  
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85097229443&partnerID=MN8TOARS>



33. **D. Orozova**, S. Hadzhikoleva, E. Hadzhikolev, A Course Gamification Model for the Development of Higher Order Thinking Skills, MIPRO 2021, IEEE 44th International Convention, September 27, 2021 – October 1, 202, Opatija, Croatia, pp. 1772-1777, ISSN 1847-3946, DOI: 10.23919/MIPRO52101.2021.9597151.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85123052448&partnerID=MN8TOARS>

34. I. Popchev, **D. Orozova**, Data Analytics Approach in Virtual Educational Space, 2021 IEEE Big Data, Knowledge and Control Systems Engineering (BdKCSE), 2021, pp. 1-5, doi: 10.1109/BdKCSE53180.2021.9627247, ISBN 978-166541042-7.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85123994089&partnerID=MN8TOARS>

35. **D. Orozova**, M. Todorova, Applying Reactive Blocks in Courses of Internet of Things, Proceedings of the 10th Annual International Conference of Education, Research and Innovation, Seville, 16-18 November, 2017, стр. 6960-6967, ISSN (print):2340-1095.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000429975307011>

36. **D. Orozova**, M. Todorova, How to Follow Modern Trends in Courses in "Databases" - Introduction of Data Mining Techniques by Example, Proceedings of the 11th Annual International Technology, Education and Development Conference, Valencia, 2017, 8186-8194, ISSN:2340-1079.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000427401303026>

37. Georgieva P., E. Nikolova, **D. Orozova**, Tendencies in Software Engineering Education, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 3601- 3609, ISBN: 978-84-09-14755-7.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000530109203108>

38. Georgieva P., E. Nikolova, **D. Orozova**, V. Jecheva, Certification In Teaching Information Technologies For Engineering Graduates, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 4564-4573.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000530109204100>

39. D. Tuparova, G. Tuparov, **D. Orozova**, Educational Computer Games and Gamification in the Higher Education – Students’ Points of View, 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, Croatia, Opatija, pp.1879-1884 Electronic ISSN: 2623-8764, DOI: 10.23919/MIPRO48935.2020.9245251.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000790326400284>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85097241244&partnerID=MN8TOARS>

40. Magdalena Todorova, **Daniela Orozova**, Generalized Net Model of Sequential Programs, In proc of the 20th International Symposium on Electrical Apparatus and Technologies (SIELA 2018), 3-6 June 2018, Bourgas, Bulgaria, DOI: 10.1109/SIELA.2018.8447068, ISBN: 978-1-5386-3420-2.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000454837100016>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85053816562&partnerID=MN8TOARS>

41. Zlatev Z., J. Ilieva, D. Orozova, G. Shivacheva, N. Angelova, Design and Research of a Sound-to-RGB smart acoustic device, Multimodal Technologies and Interaction, 2023, 7(8), 79; WoS, Scopus Q2 SJR=0.504 (за 2022), <https://doi.org/10.3390/mti7080079>

42. **Учебник:** Иван Попчев, **Даниела Орозова**, Представяне на знания в системи с изкуствен интелект, издателство Божич, Бургас, 2018, ISBN 978-619-7181-59-3.



## БИБЛИОГРАФИЯ

1. Вълканова, В., М. Сандалски, И. Попчев, С. Стоянов. Образователен портал за подготовка на ученици за зрелостни изпити. - Взаимодействието теория - практика: Ключови проблеми и решения. Международна научна конференция, 2011, 287-294. ISBN: 978-954-9370-80-5.
2. НАОА: Националната агенция за оценяване и акредитация <https://www.neaa.government.bg/>
3. Стоянов С., Д. Орозова, И. Попчев, Е. Дойчев Виртуално пространство за продължаващо обучение, БСУ Научна конференция с международно участие “Хоризонти в развитието на човешките ресурси и знанието” том 2, 2015, стр. 419-425.
4. Годорова М., Подходи, програмни среди и езици за проверка на коректността на програми и прилагането им при подготовката на софтуерни специалисти, София, 2013.
5. Най-добри 12 инструмента и софтуер за уеб анализ за стартиращи фирми <https://squeezegrowth.com/bg/best-web-analytics-tools-for-startups/>
6. Ahlers, R., Garris, R., & Driskell, J. (2014). Games, Motivation, and Learning: A Research and Practice Model. Computer Science Applications. Florida Maxima Corporation.
7. Anderson L. W., D. R. Krathwohl (Eds). A taxonomy for learning, teaching and assessing: A revision of Bloom’s Taxonomy of educational objectives: Complete edition, New York, Longman, 2001.
8. Atanassov K., On Generalized Nets Theory “Prof. Marin Drinov” Publishing House of Bulgarian Academy of Sciences, Sofia, 2007.
9. Atanassov, K., Generalized Nets and Intuitionistic Fuzziness in Data Mining, “Prof. M. Drinov” Publishing House of Bulgarian Academy of Sciences, Sofia, 2020, ISBN 978-619-245-022-9.
10. Barr A., Feigenbaum E., The Handbook of Artificial Intelligence, vol.2, 1984.
11. Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. Retrieved from <http://tech.ed.gov/wp-content/uploads/2014/03/edm-ia-brief.pdf>, последен достъп: 01.02.2023
12. Big Data Definition, Teradata, <https://www.teradata.com/Glossary/What-is-Big-Data>: 01.03.2023
13. Bloom B., M. Engelhart, E. Furst, et. al., “Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain”. New York: David McKay Company, 1956.
14. Bradley, J., Atkins, E. Coupled Cyber-Physical System Modeling and Coregulation of a CubeSat, IEEE Transactions on Robotics, 2015, Vol. 31, No. 2, 443-456.
15. Brookhart C., “How to Assess Higher-Order Thinking Skills in Your Classroom”, publ. ASCD, 2010, ISBN: 978-1-4166-1048-9.
16. Buckland, B., F. Schreier & TH. Winkler. Democratic Governance Challenges of Cyber Security. DCAF Horizon 2015 Working No. 1, Geneva Centre for the Democratic Control of Armed Forces.
17. Egger, N. etc., Themes in Science & Technology Education, 2014, Vol. 7 Issue 1, p.3-17, 15p.
18. Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG). (2015). Brussels, Belgium. ISBN: 978-9-08-168672-3.
19. Bower, J., C. Christensen. Disruptive Technologies: Catching the Wave. Harvard Business Review, January-February 1995, pp. 43-53.
20. Brynjolfsson E., A. McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies, W.W. Norton & Company, 2014.
21. Cao L. Data science: a comprehensive overview. ACM Comput Surv (CSUR). 2017;50 (3) pp.1-42.
22. Chapman A., “Principles and Methods of Data Cleaning - Primary Species and Species-Occurrence Data.”, Report for the Global Biodiversity Information Facility, Copenhagen, 2005.
23. Chatti, M, A, Dychhoff, A, L, Schroeder, U., and Thus, H. A reference model for learning analytics, International journal of Technology Enhanced learning, 2012.
24. Chemchem, A., Drias, H. From data mining to knowledge mining: Application to intelligent agents. International Joint Conference on Artificial Intelligence: Expert Systems with Applications, (2014).
25. Clifton, B., 2008: Advanced Web Metrics with Google Analytics, Wiley, New York, USA.
26. Dietrich D., B. Heller, B. Yang, Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Published John Wiley & Sons, Inc., 2015, ISBN: 978-1-118-87613-8

27. Ekman P., W.V. Friesen, Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 1971, 17(2), 124–129.
28. Escobar-Jeria, V. H., Martín-Bautista, M. J., Sánchez, D., Vila, M., 2007: Web Usage Mining Via Fuzzy Logic Techniques. In: Melin, P., Castillo, O., Aguilar, I. J., Kacprzyk, J., Pedrycz, W. (Eds.), 2007: *Lecture Notes In Artificial Intelligence*, Vol. 4529, Springer, New York, USA, pp. 243-252.
29. Estivill – Castro, V. & Yang, J. A. (2000). Fast and robust general purpose clustering algorithm. *Pacific Rim International Conference on Artificial Intelligence* (pp. 208-218).
30. Etheridge, D., *Excel Programming: Your Visual Blueprint for Creating Interactive Spreadsheets*, John Wiley & Sons, Inc. 2010.
31. Ethics Guidelines for Trustworthy AI [Online]. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> последен достъп: 01.03.2023
32. Fraley C., Raftery A. E. (1998). How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. Technical Report No. 329. Department of Statistics University of Washington.
33. Georgieva P., E. Nikolova, D. Orozova, Data Cleaning Techniques in Detecting Tendencies in Software Engineering, MIPRO 2020, 43rd International Convention, 2020, Croatia, Opatija, pp.1272-1277, Electronic ISSN: 2623-8764
34. Georgieva P., E. Nikolova, D. Orozova, Tendencies in Software Engineering Education, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 3601- 3609, ISBN: 978-84-09-14755-7
35. Georgieva P., E. Nikolova, D. Orozova, V. Jecheva, Certification In Teaching Information Technologies For Engineering Graduates, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 4564-4573
36. Gomez-Perez A., M. Fernando-Lopez, O. Corcho. “Ontological Engineering with examples from the areas of Knowledge management,” *e-Commerce and the Semantic Web 2nd. Ed.* London, Springer-Verlag, 2004.
37. Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics, *Educational Technology & Society*, 15(3), 42-57.
38. Gruber R., Translation Approach to Portable Ontologies, *Knowledge Acquisition*, 1993 V.5. 199-220.
39. Hadzhikolev, E., S. Hadzhikoleva, K. Yotov, D. Orozova, Models for Multicomponent Fuzzy Evaluation, with a Focus on the Assessment of Higher-Order Thinking Skills, *TEM Journal*, Vol.9, No.4, pp.22-28, 2020.
40. Hadzhikolev E., S. Hadzhikoleva, D. Orozova, Digital Model of a Document in a University Document Repository, XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018.
41. Hadzhikoleva S., D. Orozova, N. Andonov, E. Hadzhikolev, Generalized Net Model of a System for Quality Assurance in Higher Education, 45th International Conference Applications of Mathematics in Engineering and Economics (AMEE'19), Sozopol, June, 2019, pp. 040005-1 -040005-8, ISBN: 978-0-7354-1919-3. AIP Conference Proceedings.
42. Hadzhikoleva S., D. Orozova, E. Hadzhikolev, N. Andonov, Conceptual Model of a Centralized System for Quality Assurance in Higher Education, *Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems 2020*, pp. 87-92, ISBN 978-172815456-5, DOI 10.1109/IS48319.2020.9200174.
43. Hadzhikolev E., S. Hadzhikoleva, D. Orozova, K. Yotov, A Comprehensive Approach to Assessing Higher and Lower Order Thinking Skills, Uncertainty and Imprecision in Decision Making and Decision Support: New Advances, Challenges, and Perspectives, *Lecture Notes in Networks and Systems*, Volume 338, pp. 164–177, *Lecture Notes in Networks and Systems*, Vol. 338, pp. 164–177.
44. Hayashi, C. (1998). What is Data Science ? Fundamental Concepts and a Heuristic Example. In: Hayashi, C., Yajima, K., Bock, HH., Ohsumi, N., Tanaka, Y., Baba, Y. (eds) *Data Science*,

- Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Tokyo.
45. Han J., Kamber M., Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2001.
  46. Hornick M., E. Marcade, S. Venkayala, Java data mining: strategy, standard and practice, A practical Guide for Architecture, Design, and Implementation, 2006.
  47. International Educational Data Mining Society. 2022, <https://educationaldatamining.org>. последен достъп: 01.03.2023
  48. Jecheva V., D. Orozova, Ontology-Based Electronic Test Result Evaluation, Advances in Intelligent and Soft Computing, Third International Conference of Software, Services and Semantic Technologies S3T, Springer, 213-214, 2011. ISSN: 1867-5662.
  49. Kapp, K. (2012). The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education. Pfeiffer.
  50. Kaufmann, M., Meier, A., 2009: An Inductive Fuzzy Classification Approach applied to Individual Marketing, In: Proceedings of the 28th North American Fuzzy Information Processing Society Annual Conference, USA.
  51. Kaushik, Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity, Wiley, 2010
  52. KNIME – Open for Innovation: <https://www.knime.com/> последен достъп: 01.03.2023.
  53. Marr B., “Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance”. John Wiley & Sons Ltd, 2015.
  54. Millington, Ian, 2006. Artificial Intelligence for Games. Elsevier Inc. Ch.7.6 Reinforcement Learning, 612-628.
  55. Mobasher, B., 2007: Web Usage Mining, In: (Liu 2007, pp. 449-483).
  56. Orange system: <https://orange.biolab.si/training/introduction-to-data-mining/>, достъп: 01.03.2023
  57. Orozov B., D. Orozova, An AI application in E-Learning, CEUR Workshop Proceedings 3061, pp. 76-80, September 27–28, 2021, Plovdiv, Bulgaria, ISSN 1613-0073.
  58. Orozov B., D. Orozova, Rule Based System Against Reinforcement Learning, CompSysTech'21: International Conference on Computer Systems and Technologies '21, June 2021, Pages 67–70, ISBN: 978-145038982-2
  59. Orozova D., Appropriate E-Test System Selection Model, Comptes rendus de l'Acadé'mie bulgare des Sciences, 2019, Vol 72, No 6, pp. 811-820, ISSN 1310-1331.
  60. Orozova D., SQL User Queries Execution Model, Advances in Intelligent Systems and Computing book series (AISC, volume 1081), Uncertainty and Imprecision in Decision Making and Decision Support: New Challenges, Solutions and Perspectives, pp.184-189, 2020, ISSN 2194-5357, DOI: 10.1007/978-3-030-47024-1\_20
  61. Orozova D., I. Popchev, Cyber-Physical-Social Systems for Big Data, XXI International Symposium on Electrical Apparatus and Technologies SIELA 2020, pp. 334 – 337, ISBN 978-1-7281-4345-3
  62. Orozova D., K. Atanassov, Generalized Net Model of Processes Related to Big Data, Comptes rendus de l'Académie bulgare des Sciences, book No 12, vol. 71, 2018, pp. 1679 -1686.
  63. Orozova D., K. Atanassov, Model of Big Data Map/Reduce Processing, Comptes rendus de l'Acadé'mie bulgare des Sciences, 2019, Vol 72, No11, pp.1537-1545.
  64. Orozova D., M. Todorova, Ontology concept in courses on students, TEM Journal. Volume 7, Issue 3, Pages 693-697, ISSN 2217-8309, DOI: 10.18421/TEM73-29, 2018. ISSN (print):2217-8309.
  65. Orozova D., M. Krawczak, Project-based Learning Modelled with a Generalized Net, Issues in Intuitionistic Fuzzy Sets and Generalized Nets, Warsaw, Poland, Vol.12, 2015/2016, pp. 68–79, ISBN 978-83-61551-13-3.
  66. Orozova D., M. Todorova, Applying Reactive Blocks in Courses of Internet of Things, Proceedings of the 10th Annual International Conference of Education, Research and Innovation, Seville, 16-18 November, 2017, стр.6960-6967, ISSN (print):2340-1095
  67. Orozova D., M. Todorova, How to Follow Modern Trends in Courses in "Databases" - Introduction of

- Data Mining Techniques by Example, Proceedings of the 11th Annual International Technology, Education and Development Conference (INTED), Valencia, 2017, стр.8186-8194, ISSN:2340-1079.
68. Orozova D., K. Atanassov, M. Todorova, Generalized Net Model of the Process of Personalization and Usage of an e-Learning Environment, Proceedings of the Jangjeon Mathematical Society 19 (2016), No. 4, 615 – 624. ISSN (print):1598-7264, ISSN (online):2508-7916, ISBN:89-87809-15-3.
  69. Orozova, S. Hadzhikoleva, E. Hadzhikolev, A Course Gamification Model for the Development of Higher Order Thinking Skills, MIPRO 2021, 44th International Convention, September 27, 2021 – October 1, 202, Opatija, Croatia, pp. 1772-1777, ISSN 1847-3946.
  70. Orozova D., E. Sotirova, Modeling of a Learning Management System, XVIII-th International Symposium on Electrical Apparatus and Technologies, Proceedings of digests, SIELA 2014, 29-31 May 2014, Bulgaria, pp. 95-98, ISBN: 978-147995817-7
  71. Orozova D., M. Georgieva, A model of the process of Big Data with Generalized Net, Electrical Apparatus and Technologies (SIELA), 2016 19th International Symposium, Bourgas, 238-241, ISSN: 978-1-4673-9521-2, ISBN 978-619-160-648-1
  72. Orozova D., A. Ivanov, Generalized Net Model of Virtual Collaboration Space, XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018, 3-6 June 2018, Bourgas, Bulgaria, ISBN 978-1-5386-3418-9
  73. Orozova D., K. Kaloyanova, M. Todorova, Introducing Information Security Concepts and Standards in Higher Education, TEM Journal. Volume 8, Issue 3, Pages 1017-1024, ISSN 2217-8309, DOI: 10.18421/TEM83-46, August 2019.
  74. Orozova D., N. Hristova, Generalized Net Model for dynamic decision making and prognoses, XXI International Symposium on Electrical Apparatus and Technologies SIELA 2020, 3-6 June 2020, Bourgas, Bulgaria, pp. 330 – 333, ISBN ISBN 978-1-7281-4345-3.
  75. Orozova D., N. Hristova, An application of analytical data research in e-learning system, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering Volume 1031 (2021) 012058, IOP Publishing, Online ISSN: 1757-899X
  76. Petri C.A., “Kommunikation mit Automaten”. Bonn:Institut fuer Instrumentelle Mathematik, Schriften des IIM Nr.3, 1962. Also, English translation: “Communication with Automata”, New York: Griffiss Air Force Base. Tech. Rep. RADC-TR-65-377, vol.1, Suppl. 1, 1966.
  77. Popchev I., D. Orozova, Towards a multistep method for assessment in e-learning of emerging technologies, Cybernetics and Information Technologies, Vol. 20, 3, 2020, 116-129, ISSN: 1311-9702
  78. Popchev I., D. Orozova. Towards Big Data Analytics in the E-learning Space. Cybernetics and Information Technologies, Vol. 19(3), 2019, pp. 16-25. ISSN: 1311-9702.
  79. Popchev, D. Orozova, S. Stoyanov, IoT and Big Data Analytics in E-Learning, 2019 Big Data, Knowledge and Control Systems Engineering (BdKCSE), 2019, pp. 1-5, ISBN 978-172816481-6
  80. Popchev I., D. Orozova, Text Mining in the Domain of Plant Genetic Resources, Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems, 2020, pp. 596-600.
  81. Popchev I., D. Orozova, DataScience: Experience and Trends, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering, Volume 1031 (2021) 012057, IOP Publishing, Online ISSN: 1757-899X, Print ISSN: 1757-8981 doi:10.1088/1757-899X/1031/1/012057.
  82. Popchev I., D. Orozova, Data Analytics Approach in Virtual Educational Space, Big Data, Knowledge and Control Systems Engineering (BdKCSE), 2021, pp. 1-5, ISBN 978-166541042-7.
  83. Popchev I., D. Orozova, Algorithms for Machine Learning with Orange System, International Journal of Online and Biomedical Engineering (iJOE), Vol. 19, No. 04, 2023, pp. 109 -123, ISSN: 2626-8493, DOI: <https://doi.org/10.3991/ijoe.v19i04.36897>
  84. Popchev, I., I. Radeva, Risk Analysis – an Instrument for Technology Selection. Engineering Sciences (2019) LVI (4) 5-20, ISSN 1312-5702, DOI: 10:7546/ Eng.Sci.LVI.19.04.01.
  85. Policy and Investment Recommendations for Trustworthy Artificial Intelligence [Online]

- <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence> последен достъп: 01.06.2023
86. Popper, Karl Raimund (1972). Objective knowledge. Oxford, Clarendon Press
  87. RapidMiner - Depth for Data Scientists: <https://rapidminer.com/>
  88. Reinforcement Learning, Searching for optimal policies: <https://www.cs.upc.edu/~mmartin/Ag4-4x.pdf>
  89. Russell, S and P. Norvig. Artificial Intelligence: A Modern Approach, Prentice Hall, Upper Saddle River, New Jersey, 1995.
  90. Siemens, G, and Long, P, Penetrating the fog-analytics in learning and education. Asynchronous Learning Networks, 2011.
  91. Schaughency E., J. Smith, J. van der Meer, and D. Berg, Classical Test Theory and Higher Education: Five Questions, In: Handbook on Measurement, Assessment, and Evaluation in Higher Education, Edited by Charles Secolsky and D. Brian Denison, Publ. Routledge, 2017, ISBN: 1138892157.
  92. Social Analytics 2022, <https://www.gartner.com/en/information-technology/glossary/social-analytics>
  93. Stoyanov, S., I. Popchev, I., E. Doychev, D. Mitev, V. Valkanov, A. Stoyanova-Doycheva, V. Valkanova, I. Minov. DeLC Educational Portal. – Cybernetics and Information Technologies, Vol. 10, 2010, No. 3, 49-69, ISSN: 1311-9702
  94. Stoyanova-Doycheva A., T. Glushkova, S. Stoyanov, D. Orozova, Lifelong Learning Supported by an Intelligent Tourist Guide, 2018, AIP Conference Proceedings, Volume 2048, pp. 020038; ISSN: 0094243X, ISBN: 978-073541774-8, DOI: 10.1063/1.5082056
  95. Todorov, J., V. Valkanov, S. Stoyanov, B. Daskalov, I. Popchev, D. Orozova. Chapter 6: Personal Assistants in a Virtual Education Space, Sgurev V., Jotsov V., Kacprzyk J. (eds) Practical Issues of Intelligent Innovations. Studies in Systems, Decision and Control, 140, Springer International Publishing AG part of Springer Nature, 2018, ISSN 2198-4182, ISSN 2198-4190 (electronic), DOI:10.1007/978-3-319-78437-3, 22, 131-153
  96. Todorov J., E. Doychev, D. Orozova, A. Stoyanova-Doycheva, IoT Multiagent Assistant in Virtual Educational Space, AIP Conference Proceedings, Vol. 2048, 020031; ISSN: 0094243X,
  97. Todorova M., D. Orozova, Generalized Net Model of Sequential Programs, In proc of the 20th International Symposium on Electrical Apparatus and Technologies (SIELA 2018), 2018, Bulgaria, DOI: 10.1109/SIELA.2018.8447068, ISBN: 978-1-5386-3420-2.
  98. Todorova M., D. Orozova, Training Difficulties in Deductive Methods of Verification and Synthesis of Program, International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 9, No. 7, 2018, pp. 18-22, doi:10.14569/IJACSA.2018.090703.
  99. Tuparova D., G. Tuparov, D. Orozova, Educational Computer Games and Gamification in the Higher Education – Students’ Points of View, 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, Croatia, pp.1879-1884 ISSN: 2623-8764.
  100. Unsupervised Learning and Data Clustering: <https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a>, последен достъп: 01.03.2023
  101. Venkatram K., Geetha M. A., Review on Big Data & Analytics– Concepts, Philosophy, Process and Applications, Cybernetics and Information Technologies, Vol. 17(2), 2017, 3-27, ISSN: 1311-9702;
  102. Wang F-Y., Driving into the future with ITS, IEEE Intelligent System, Vol.21, No.3, 2006, 94–95.
  103. WEKA Waikato Environment for Knowledge Analysis: <http://www.cs.waikato.ac.nz/ml/weka>
  104. Wordclouds - free online generator: [www.wordclouds.com/](http://www.wordclouds.com/)
  105. Zadeh, L., The concept of a linguistic variable and its application to approximate reasoning, Published 1975, Information Sciences, v. 8, pp. 199–249.
  106. Zlatev Z., J. Ilieva, D. Orozova, G. Shivacheva, N. Angelova, Design and Research of a Sound-to-RGB smart acoustic device, Multimodal Technol. Interact. 2023, 7(8), 79.
  107. Zumstein D., M. Kaufmann, A Fuzzy Web Analytics Model for Web Mining, IADIS European Conference on Data Mining, 2009, pp.59-66, ISBN: 978-972-8924-88-1.