

**Коляда Михайло Георгійович,**  
доктор педагогічних наук, завідувач  
кафедри інженерної та комп'ютаційної  
педагогіки Інституту післядипломної освіти  
інженерно-педагогічних працівників  
(м. Донецьк) ДВНЗ  
«Університет менеджменту освіти»

## **ВИКОРИСТАННЯ СИСТЕМ ВИДОБУТКУ ЗНАНЬ DATA MINING ДЛЯ ПРОГНОЗУВАННЯ ПЕДАГОГІЧНИХ ПРОЦЕСІВ ТА ЯВИЩ**

*Анотація.* У статті висвітлено надзвичайно актуальну проблему освітньої діяльності — використання систем видобутку знань для прогнозування педагогічних процесів та явищ. Проаналізовано предметно-орієнтовані аналітичні системи, суто статистичні пакети та клас нейромережних систем видобутку знань із масивів педагогічної інформації. Обґрунтовано, що тільки системи видобутку знань Data Mining мають можливість «нащупувати» педагогічні закономірності та закони, які іншими способами взагалі виявити неможливо.

**Ключові слова:** педагогічне прогнозування, системи видобутку знань, Data Mining, приховані знання.

**Коляда Михаил Георгиевич**

## **ИСПОЛЬЗОВАНИЕ СИСТЕМ ДОБЫЧИ ЗНАНИЙ DATA MINING ДЛЯ ПРОГНОЗИРОВАНИЯ ПЕДАГОГИЧЕСКИХ ПРОЦЕССОВ И ЯВЛЕНИЙ**

*Аннотация.* В статье освещена чрезвычайно актуальная проблема образовательной деятельности — использование систем добычи знаний для прогнозирования педагогических процессов и явлений. Проанализированы предметно-ориентированные аналитические системы, чисто статистические пакеты и класс нейросетевых систем добычи знаний из массивов педагогической информации. Обосновано, что только системы добычи знаний Data Mining имеют возможность «нащупывать» педагогические закономерности и законы, которые другими способами вообще обнаружить невозможно.

**Ключевые слова:** педагогическое прогнозирование, системы добычи знаний, Data Mining, скрытые знания.

**Koliada Mykhailo Georgijovith**

## **USE OF SYSTEMS OF EXTRACTION OF KNOWLEDGE DATA MINING FOR FORECASTING OF PEDAGOGICAL PROCESSES AND THE PHENOMENA**

**Abstract.** In article it is covered extremely actual problem of educational activity — use of systems of extraction of knowledge for forecasting of pedagogical processes and the phenomena. Subject-oriented analytical systems, purely statistical packages and class neural networks systems of extraction of knowledge from files of the pedagogical information are analyzed. It is proved that only systems of extraction of knowledge Data Mining have

*possibility «to grope» pedagogical laws and laws which in general cannot be found out in other ways.*

**Key words:** *pedagogical forecasting, systems of extraction of knowledge, Data Mining, the latent knowledge.*

### **Постановка проблеми**

**Актуальність проблеми дослідження.** У освітній, як і будь-якій іншій галузі людської діяльності, існує величезна затребуваність у прогнозах, які були б зорієнтовані на різноманітні цілі освітньої діяльності та на різну часову перспективу. В умовах спільної діяльності тих, хто навчає і виховує та тих, кого навчають, виховують і розвивають, роль педагогічного прогнозування в регуляції їхньої взаємодії виявляється особливо гостро, оскільки без об'єктивного результату цього процесу погодженість обох сторін неможлива.

У сьогоденні на допомогу дослідникам, які займаються прогнозуванням педагогічних процесів та явищ, приходять системи комп'ютерного моделювання та прогнозування. На нашу думку, нині існує лише чотири групи моделей, що могли б реалізувати цю складну педагогічну проблему прогностики.

*Перша група* — це комп'ютерні моделі, засновані на якій-небудь класичній теорії навчання або виховання; *друга група* — комп'ютерні моделі, що засновані на попередньому емпіричному аналізі зібраних педагогічних даних, унаслідок проведеного педагогічного дослідження (експерименту), без готових зв'язків із якою-небудь теорією; *третья група* — комп'ютерні моделі, що запозичені з інших сфер знання, наприклад, соціології, екології, економіки, синергетики і т. ін.; *четверта група* — комп'ютерні моделі, які самі отримані в комп'ютерних моделюючих середовищах, виявлених на основі алгоритмів штучного інтелекту (нейронних мереж, теорії нечіткої логіки тощо).

**Актуальність проблеми використання систем видобутку знань у великих масивах освітньої інформації для прогнозування педагогічних процесів та явищ обумовлена такими протиріччями:**

- між соціальною обумовленістю необхідності гарантованого забезпечення результату навчання, виховання та розвитку, і відсутністю методологічної і методичної бази, що орієнтувала б дослідників на соціально значущий, особистісно орієнтований та прогнозований кінцевий результат;
- між доцільністю автоматизованого вибору керуючих освітніх впливів, адекватних динамічним реаліям педагогічних процесів та явищ, і недостатньою науковою обґрунтованістю використання систем видобутку знань для аналізу і побудови педагогічного прогнозу;
- між необхідністю використання комп'ютерних систем штучного інтелекту для виявлення освітніх закономірностей та законів, як основи для прогнозування та їх утіленням у реальну педагогічну практику.

### **Аналіз останніх досліджень та публікацій**

Серед науковців, що займалися проблемами прогнозування соціальних систем можна виділити І. Бестужева-Ладу, О. Карманчікова, О. Кірик, Л. Кудринську, Г. На-

местнікову, В. Сафронова, К. Ставицьку. Питання теорії педагогічного прогнозування розроблялися такими дослідниками, як Г. Олександровим, Б. Гершунським<sup>2</sup>, В. Кутєвим, В. Ледневим, В. Сластьоніним, Т. Шаховою та ін.

Проблемам вирішення завдань педагогічного прогнозування приділяли свою увагу такі дослідники як О. Белкін, Б. Вульфсон, О. Гендін, Л. Грибова, І. Журавльов, Е. Костяшкін, І. Лернер, В. Сітаров, М. Скаткін; окремі аспекти застосування прогнозування в освітньому процесі вивчали С. Бакулєв, В. Баландін, Т. Димова, Є. Єфремов, В. Іванов та ін.

Проте обґрунтованого розгляду систем видобутку знань для прогнозування педагогічних процесів та явищ ще не зроблено, — складність обговорення цих інформаційних технологій полягає в тому, що вони практично не попали у поле уваги фахівців-педагогів; їх застосування для прогнозування педагогічних об'єктів є дискусійним; ще не сформувалася єдина точка зору професіоналів-користувачів на доцільність їх використання у педагогічній практиці. У вищеназваних наукових працях, питання розгляду таких програмних засобів штучного інтелекту не зроблено, проблему ефективного використання систем Data Mining для прогнозування педагогічних процесів та явищ, їх узагальненого, оптимального, обґрунтованого застосування ще не вирішено. Тому, з урахуванням означеної проблематики, та керуючись основними положеннями зазначених протиріч, визначимо *мету* статті: *розгляд систем видобутку знань для прогнозування педагогічних процесів та явищ; виявлення можливостей використання систем Data Mining для прогнозування різних аспектів розвитку педагогічних систем.*

Серед *завдань*, що виникають при цьому, було виділено такі: 1) виявити основні типи стандартних закономірностей систем Data Mining; 2) розглянути сутнісну характеристику нейромережних систем видобутку знань у масивах педагогічної інформації; 3) охарактеризувати найбільш затребувані алгоритми та ідеї систем знаходження знань.

### Виклад основного матеріалу

Інформаційна база даних будь-якої сучасної освітньої установи або навчального закладу зазвичай містить багатий фактографічний матеріал.

Це звіти про успішність, якість знань учнів (студентів) за семестрами (за чвертями), це контрольні зрізи, що проводилися у вигляді поточних, модульних, семестрових (підсумкових), ректорських (директорських) контрольних робіт, це підсумки інспекторських перевірок усіх рівнів (починаючи від місцевих відділів освіти і науки та закінчуючи Міністерством освіти і науки України). Величезний потенціал міститься в матеріалах досліджень різних наукових психолого-педагогічних тем, починаючи від кафедральних і закінчуючи цільовими програмами на рівні державних дослідницьких центрів та інститутів. Дуже цінний матеріал містять дані самих зацікавлених осіб, а саме: вчителів-дослідників, аспірантів, докторантів, інших категорій наукових співробітників, які збирають, накопичують, систематизують освітній матеріал не заради суто прагматичної мети, а заради суто наукового інтересу,

---

<sup>2</sup> Гершунский Б. С. Философия образования: учеб. пособие для студ. высших и сред. пед. учеб. заведений / Б. С. Гершунский. — М.: Москов. психолого-социальный ин-т, 1998. — 432 с.

творчого покликання, для вдосконалення і поліпшення системи навчання та виховання тих категорій осіб, з якими вони безпосередньо контактують у процесі освітньої взаємодії.

Як правило, кожен запис у подібній таблиці описує якийсь конкретний освітній об'єкт чи факт. Наприклад, записи у класних журналах середніх навчальних закладів відбивають той факт, що якийсь учень мав конкретні бали за навчальними предметами, мав визначені рівні вихованості, підтверджені розгорнутими характеристиками класного керівника. У цьому переліку можна виділити більш вузькі специфічні дані, що стосуються окремих категорій дітей (наприклад, тих, у кого не було батьків, чи тих, хто займався у спортивних секціях тощо). Дуже важливим є інформація про динаміку навчання і виховання окремих особистостей із чверті у чверть, з одного навчального року — в іншій (із семестру — в семестр, із курсу — на курс), або ґрунтовне обстеження психологічних властивостей особистості, її мотивів, типології, рівнів розвитку окремих сторін індивідуальності та їх рушійних сил. Здавалося б, що за великим рахунком нічого, крім цих зведень, накопичений матеріал не містить. Однак сукупність великої кількості таких записів, зібраних за кілька років, може стати важливим джерелом додаткової, більш важливої інформації, яку не можна одержати на основі лише одного конкретного запису, а саме — зведень про закономірності, тенденції або взаємозалежності між цими даними. Прикладами такої інформації є звіти про те, як залежить успішність від вікових чи інших характеристик, від відвідування предметних кружків або спортивних секцій учнями; від особистісних і професійних якостей педагогів. Або — яка категорія учнів має більш високі показники в навчанні; що впливає на рівень вихованості; які заходи мають більш ефективний вплив на рівень креативності; який рівень логічного мислення учнів, і багато інших залежностей.

Подібного роду інформація зазвичай і використовується у стратегічному плануванні, аналізі та прогнозуванні освітніх і виховних процесів та заходів їх удосконалення, і цінність її постійно зростає не тільки для педагогів-дослідників та педагогів-новаторів, але й для рядових учителів, освітян-управлінців та викладачів-практиків. Тому процес пошуку закономірностей, зв'язків та залежностей і одержав назву *Data Mining*, що в перекладі означає — *система «видобутку»* (elicitation) чи *«розкопки даних»*. Поряд із словосполученням *Data Mining* (видобуток даних) також вживається термін *Data Miner* (добувач даних).

В основу сучасної технології *Data Mining* (discovery-driven data mining — керування відкритим збором даних, або «відкриття, що робиться через збір даних») покладена концепція шаблонів (так званих паттернів), що відбивають фрагменти багатоаспектних взаємин у даних, що аналізуються. Ці шаблони являють собою закономірності, властиві підвибіркам даних, що можуть бути компактно виражені в зрозумілій для людини формі. Пошук шаблонів виконується за допомогою методів, не обмежених рамками апріорних припущень про структуру вибірки і типи розподілів значень аналізованих показників.

Важливе положення систем *Data Mining* — це нетривіальність розшукуваних шаблонів, яка означає, що знайдені стереотипи повинні відбивати неочевидні, несподівані (unexpected) регулярності в аналізованих даних, які становлять так звані «сховані знання» (hidden knowledge), нерозкриті зв'язки, невиявлені кореляції та

закономірності. Дослідники і розробники комп'ютерних програм такого напрямку дійшли висновку, що «сирі дані» (raw data — *вихідні нерозкриті, необроблені дані*) містять глибинний шар знань, за грамотної розкопки якого можуть бути виявлені дійсні самородки педагогічних рішень.

Як основне визначення словосполучення Data Mining можна прийняти таке формулювання: *Data Mining — це процес виявлення в «сирих» інформаційних даних раніше невідомих, нетривіальних, практично корисних і доступних до інтерпретації знань, необхідних для прийняття рішень у різних сферах людської діяльності*<sup>7</sup>.

З іншого боку, *Data Mining — це клас комп'ютерних програмних засобів, що працюють на ідеях штучного інтелекту і являють собою моделюючі інформаційні системи аналізу, систематизації, інженерії знань та виявлення прихованих закономірностей для прогнозування поведінки системи, яка вивчається, та прийняття остаточних рішень в умовах невизначеності.*

Оскільки педагогічна галузь відрізняється труднощами, складністю системної організації, ймовірнісним характером навчально-виховних явищ і множинністю причин та чинників, що одночасно діють на об'єкти прогнозування, то вона належить до надкібернетичного рівня організації освітніх систем<sup>6</sup>, закономірності якого не можуть бути досить точно описані мовою статистичних або інших аналітичних математичних моделей<sup>1</sup>. При цьому педагогічні дані мають неоднорідний, гетерогенний, нестационарний, різноплановий характер і вони часто відрізняються високою розмірністю.

Система Data Mining є мультидисциплінарною сферою, що виникла і розвивається завдяки комп'ютерно-комунікаційній техніці на базі досягнень *прикладної статистики, теорії розпізнавання образів та візуалізації даних, теорії сховищ даних і баз даних, теорії оперативної аналітичної обробки інформації, теорії інформаційного пошуку, теорії ефективних обчислень, теорії експертних систем, теорії нейромереж та інших методів штучного інтелекту*. Звідси множина методів і алгоритмів, реалізованих у різних діючих прогнозуючих системах Data Mining, тому багато з таких систем інтегрують у собі відразу кілька підходів. Проте, як правило, у кожній такій системі наявний якийсь ключовий компонент, на який робиться головна ставка. Нижче наводиться огляд зазначених ключових складників на основі робіт<sup>3,4</sup>.

Існує багато різних класів систем Data Mining, а саме: *предметно-орієнтовані аналітичні системи, суто статистичні пакети*, до яких можна віднести SAS (компанія SAS Institute), SPSS (фірми SPSS), STATGRAPICS (Manugistics), STATISTICA, STADI, але з погляду комп'ютаційного прогнозування, найбільший інтерес становить *клас нейромережних систем*.

---

<sup>7</sup> Piatetsky-Shapiro G. Bayesian Networks for Knowledge Discovery. In Advances in Knowledge Discovery and Data Mining, eds. / U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, D. Heckerman. — Menlo Park, Calif. : AAAI Press, 1996. — P. 273–306.

<sup>6</sup> Boulding K. E. General Systems Theory / K. E. Boulding // The Skeleton of Science. — Management Science. — 1956. — № 2. — P. 197–208.

<sup>1</sup> Ван Гиг Дж. Прикладная общая теория систем: в 2-х кн. / Дж. Гиг Ван / под ред. Б. Г. Сушкова, В. С. Тюхнина ; пер. с англ. — М. : Мир, 1981. — Кн. 1. — 336 с.

<sup>3</sup> Дюк В. А. Data Mining : учеб. курс / В. А. Дюк, А. П. Самойленко. — СПб : Питер, 2001. — 368 с.

<sup>4</sup> Кирик О. Б. Социальное прогнозирование : учеб. пособие / О. Б. Кирик. — Вологда : ВоГТУ, 2010. — 105 с.

*Штучні нейронні мережі (ШНС)*. Вони будуються за принципами організації і функціонування аналогів біологічних об'єктів, зокрема за подобою роботи головного мозку людини. Крім *розпізнавання образів, ідентифікації, оптимізації, управління складними об'єктами* такі системи здатні вирішувати і *завдання прогнозування*.

Прикладами нейромережних систем є програмні продукти: *Brain Maker* (фірми CSS), *NeuroShell* (Ward Systems Group), *OWL* (HyperLogic), *Neuro Builder* (РосБізнес-Консалтинг, Росія).

Найбільш відомим серед таких засобів є пакет *STATISTICA Neural Networks* (виробник StatSoft). Істотною перевагою цього пакету є те, що він природним чином вбудований у могутній арсенал аналітичних засобів програми STATISTICA. Саме сполучення класичних і нейромережних методів дає потрібний ефект у прогнозуванні<sup>5</sup>.

Крім перерахованих систем Data Mining коротко охарактеризуємо деякі інші популярні ідеї й алгоритми, що найбільш часто застосовуються в системах видобутку даних та знаходженні знань.

*Системи міркувань на основі аналогічних випадків (case based reasoning — CBR)*. Для здійснення прогнозу на майбутнє в системах такого типу використовується підхід, заснований на ідеї знаходження в даних минулих періодів близьких аналогій наявної ситуації і при відборі рішення (відповіді) обирають таке, яке вже було знайдено на основі вирішення аналогічної ситуації, за умови, що це рішення в попередніх задачах визначалося як найкраще і правильне.

Цей метод ще називають *методом «найближчого сусіда» (nearest neighbour)* або методом, який заснований на алгоритмі всієї сукупності інформації, що накопичена в пам'яті (у базі даних) — *memory based reasoning*.

Головним недоліком таких підходів є те, що системи цих класів взагалі не створюють моделей чи правил, що узагальнюють попередній досвід. Прикладами систем, що використовують ідеї CBR, можна назвати такі: *KATE tools* (Acknosoft, Франція), *Pattern Recognition Workbench* (Unica, США).

Системи Data Mining, що використовують *дерева рішень (decision trees)*. Програмні продукти такого типу створюють ієрархічну деревоподібну структуру правил типу, що класифікують за ознаками типу «ЯКЩО... ТО...» (if – then).

Правила зв'язку типу: *ЯКЩО Причина, ТО Наслідок, де Причина і Наслідок* представлені кодовими чи текстовими значеннями (об'єктами) або комбінацією кодових чи текстових значень (об'єктів). Для ухвалення рішення утворюється система розгалужень, яка визначає, до якого класу треба віднести певний об'єкт чи ситуацію. Цей механізм реалізується різними способами, наприклад, потрібно відповісти на запитання, що стоять у вузлах цього дерева, починаючи з його кореня; найчастіше задаються питання типу «Об'єкт *x* більше об'єкта *y*? (чи менше або дорівнює йому). І так процес повторюється, утворюючи при цьому розгалужену деревоподібну систему.

Найвідомішими програмними продуктами цього класу є: *See5/35.0* (RuleQuest, Австралія), *Clementine* (Integral Solutions, Великобританія), *SIPINA* (University of Lyon,

---

<sup>5</sup> Нейронные сети. STATISTICA Neural Networks : методология и технологии современного анализа данных / под ред. В. П. Боровикова. – 2-е изд., перераб. и доп. — М. : Горячая линия-Телеком, 2008. — 392 с.

Франція), *IDIS* (Information Discovery, США), *KnowledgeSeeker* (ANGOSS, Канада), *AnswerTree* (SPSS).

*Програми еволюційного програмування.* У цьому класі систем Data Mining гіпотези про тип залежності цільової змінної від інших змінних, формулюються у вигляді програм на спеціальній внутрішній мові програмування. Процес побудови програм будується як еволюція у світі програм (у цьому підхід схожий на генетичний алгоритм – див. нижче). Коли система знаходить програму, що більш-менш задовольняє залежність, яку шукають, вона починає вносити в неї невеликі модифікації і відбирає серед побудованих дочірніх програм ті, які підвищують її точність. У такий спосіб система нібито «вирощує» лінії генетичних напрямків реалізації програм, які конкурують між собою в точності вираження залежності, яку відшуковують. Потім, спеціальний модуль системи переводить знайдені залежності з внутрішньої мови системи на зрозумілу користувачеві мову (наприклад, у вигляді математичних формул, таблиць, графіків тощо). Є у цих класах систем ще й інший напрямок еволюційного програмування, пов'язаний із пошуком залежностей цільових функцій від інших змінних, що проявляються у формі якогось визначеного типу (наприклад, у формі поліномів).

Яскравим представником програм такого типу можна назвати систему *PolyAnalyst* (Росія); багато ідей цієї системи реалізовано в іншій більш відомій системі *NeuroShell* (компанія Ward Systems Group).

Дуже популярним напрямом систем Data Mining є ідеї так званого *генетичного алгоритму*. Методологія роботи таких систем запозичена з біологічних теорій популяцій хромосом при розмноженні тварин та людини, а саме, використовуються закономірності популяцій хромосом. Для реалізації концепції вибору найкращого рішення вводиться спосіб зіставлення різних хромосом. Популяція обробляється за допомогою процедур репродукції, мінливості (тобто – мутацій) та генетичної композиції. Ці процедури імітують біологічні процеси, серед яких найбільш важливими є такі: *випадкові мутації* даних в індивідуальних хромосомах, *переходи* (кроссинговер) і *рекомбінація* генетичного матеріалу, що міститься в індивідуальних батьківських хромосомах (аналогічно гетеросексуальній репродукції) та міграції генів. У ході роботи таких процедур на кожній стадії еволюції з'являються популяції з усе більш досконалішими індивідуальними особливостями.

Головною перевагою генетичних алгоритмів є те, що їх легко розпаралелити. Наприклад, можна розбити покоління на кілька підгруп і працювати з кожною з них незалежно, обмінюючись час від часу декількома хромосомами. До недоліків таких систем відносять критерій добору хромосом, при цьому використовувані процедури є евристичними і далеко не гарантують знаходження найкращого прогнозного рішення. Як і в реальному житті, еволюцію може «заклинити» на якій-небудь непродуктивній гілці біологічного розвитку, або, навпаки, цей алгоритм може привести в процесі еволюції генетичного алгоритму до дуже ефективних кінцевих результатів (у біології це називають високоефективною селекцією в доборі генетичного матеріалу для майбутнього потомства). Це особливо проявляється при вирішенні багаторозмірних задач зі складними внутрішніми зв'язками. Прикладом такої системи може бути програма *GeneHunter* (фірма Ward Systems Group).

Всесвітньо відома компанія Oracle (США) у фірми Thinking Machines Corporation у червні 1999 року придбала програмну систему Data Mining під назвою *Darwin*, а в наступні роки на її основі почала випускати новий програмний продукт під назвою *Darwin Oracle Data Mining Suite*, а потім *Oracle9i Data Mining* (скорочено — ODM).

Щоб ще раз підкреслити значущість програм такого типу, вкажемо ще на такий факт. Компанія Microsoft теж офіційно оголосила про створення своєї системи Data Mining. Спеціальна дослідницька група фірми Microsoft, яку очолює Усам Файяд, із представниками п'яти запрошених фірм-партнерів (Angoss, Datasage, E.piphany, SAS, Silicon Graphics, SPSS) готують спільний проект не тільки з розробки власної версії такої могутньої системи видобутку знань, але і з розробки стандарту обміну даними і засобів для інтеграції інструментів Data Mining з базами і сховищами даних. Це вказує на те, що програми такого класу завойовують усе більший простір у системах комп'ютерного знаходження закономірностей і прогнозування результатів розвитку об'єктів великої розмірності.

### **Висновки**

Розглядаючи системи видобутку знань Data Mining, дійшли висновку, що більшість із них можна використовувати для прогнозування педагогічних процесів та явищ. Мало того, на сьогоднішній день немає більш придатних для цієї мети програмних засобів, що могли б так ефективно вирішувати завдання педагогічної прогностики. Перевагою використання у педагогічному прогнозуванні інформаційних комп'ютерних систем видобутку знань є той факт, що тільки такі системи мають можливість «нащупати» педагогічні закономірності та закони, які іншими способами взагалі виявити неможливо.

### **Перспективи подальших розвідок у цьому напрямі**

Комп'ютерні системи видобутку знань Data Mining мають величезні потенційні можливості, ми знаходимося тільки на початку шляху у визначенні нових педагогічних відкриттів. Перспективи їх використання дуже привабливі для дослідників; їх використання залежить не тільки від ступеня підготовленості пошукувачів, але й від «інтелектуального» рівня спроможності програмного забезпечення систем Data Mining.

### **Список використаних джерел**

1. Ван Гиг Дж. Прикладная общая теория систем: в 2-х кн. / Дж. Гиг Ван / под ред. Б. Г. Сушкова, В. С. Тяхнина ; пер. с англ. — М. : Мир, 1981. — Кн. 1. — 336 с.
2. Гершунский Б. С. Философия образования : учеб. пособие для студ. высших и сред. пед. учеб. заведений / Б. С. Гершунский. — М. : Москов. психолого-социальный ин-т, 1998. — 432 с.
3. Дюк В. А. Data Mining : учеб. курс / В. А. Дюк, А. П. Самойленко. — СПб : Питер, 2001. — 368 с.
4. Кирик О. Б. Социальное прогнозирование : учеб. пособие / О. Б. Кирик. — Вологда : ВоГТУ, 2010. — 105 с.



5. Нейронные сети. STATISTICA Neural Networks : методология и технологии современного анализа данных / под ред. В. П. Боровикова. – 2-е изд., перераб. и доп. — М. : Горячая линия-Телеком, 2008. — 392 с.

6. Boulding K. E. General Systems Theory / K. E. Boulding // The Skeleton of Science. — Management Science. — 1956. — № 2. — P. 197–208.

7. Piatetsky-Shapiro G. Bayesian Networks for Knowledge Discovery. In Advances in Knowledge Discovery and Data Mining, eds. / U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, D. Heckerman. — Menlo Park, Calif. : AAAI Press, 1996. — P. 273–306.

## **USE OF SYSTEMS OF EXTRACTION OF KNOWLEDGE DATA MINING FOR FORECASTING OF PEDAGOGICAL PROCESSES AND THE PHENOMENA**

In article it is covered extremely actual problem of educational activity – use of systems of extraction of knowledge for forecasting of pedagogical processes and the phenomena. Four groups of models which could realize pedagogical problems prognostics are characterized. The methodology of model live neuron for the decision of problems of artificial intelligence is considered. The main idea of modern technologies Data Mining which is shown through opening which is carried out by means of data gathering is defined.

The basic ideas of technologies Data Mining and types of standard laws which in them are used are resulted: associations, sequences, classifications, clasterizations and forecasting. Subject-oriented analytical systems, purely statistical packages and class neural networks systems of extraction of knowledge from files of the pedagogical information are analyzed.

The most requested algorithms and ideas of systems of a finding of knowledge are characterized: «reasons» on the basis of similar cases, evolutionary programming, genetic algorithm and algorithm of the limited search. It is defined that the most suitable systems of artificial intelligence for forecasting of pedagogical objects are neural networks systems. The conventional that neural networks are natural addition of classical systems of the analysis and forecasting and are applied there where standard procedures do not give necessary effect. It is proved that only systems of extraction of knowledge Data Mining have possibility «to grope» pedagogical laws and laws which in general cannot be found out in other ways.