

УДК 159.923:343.83

DOI 10.33463/2687-1238.2022.30(1-4).1.86-95

**СВЕТЛАНА НИКОЛАЕВНА СОРОКОУМОВА,**

доктор психологических наук, профессор,  
Российская академия образования, г. Москва, Российская Федерация;  
профессор факультета психологии,  
Российский государственный социальный университет,  
г. Москва, Российская Федерация;  
профессор кафедры социальных и гуманитарных дисциплин,  
Новосибирский военный ордена Жукова институт  
имени генерала армии И. К. Яковлева войск Национальной гвардии,  
г. Новосибирск, Российская Федерация;  
профессор кафедры юридической психологии и педагогики,  
Академия ФСИН России, г. Рязань, Российская Федерация,  
ORCID 0000-0001-8339-6597,  
e-mail: [4013@bk.ru](mailto:4013@bk.ru);

**ДЕНИС АЛЕКСЕЕВИЧ КУРДИН,**

адъюнкт факультета подготовки научно-педагогических кадров,  
Академия ФСИН России, г. Рязань, Российская Федерация,  
ORCID 0000-0002-9292-7342,  
e-mail: [dnskurdin@gmail.com](mailto:dnskurdin@gmail.com)

## **ИСПОЛЬЗОВАНИЕ DATA MINING В ИЗУЧЕНИИ ДИНАМИКИ ЛИЧНОСТНОГО РОСТА КУРСАНТОВ ВЕДОМСТВЕННОГО ВУЗА ФЕДЕРАЛЬНОЙ СЛУЖБЫ ИСПОЛНЕНИЯ НАКАЗАНИЙ**

### **Для цитирования**

Сорокоумова, С. Н. Использование data mining в изучении динамики личностного роста курсантов ведомственного вуза Федеральной службы исполнения наказаний / С. Н. Сорокоумова, Д. А. Курдин // Человек: преступление и наказание. – 2022. – Т. 30(1–4), № 1. – С. 86–95. – DOI : 10.33463/2687-1238.2022.30(1-4).1.86-95.

**Аннотация.** Исследования в области компьютерных технологий и создание новых программных продуктов сегодня крайне актуальны. Данные сферы развивают смежные научные области. Data mining – одно из ключевых понятий современной науки, метод, способный внести новые веяния в устоявшиеся инструменты добычи новых знаний. Данный метод имеет под собой философский фундамент в виде ответвления механицизма – коннекционизма. Одно из самых перспективных направлений data mining – это нейронные сети. Искусственная нейронная сеть представляет собой алгоритм, позволяющий принимать решения в отношении испытуемых посредством распределения их по группам. В статье отра-

© Сорокоумова С. Н., Курдин Д. А., 2022



Статья лицензируется в соответствии с лицензией [Creative Commons Attribution-NonCommercial-ShareAlike 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)

жены результаты работы по созданию искусственной нейронной сети на базе Академии ФСИН России как эффективного инструмента психологического сопровождения курсантов ведомственных вузов Федеральной службы исполнения наказаний. Целью исследования являлось рассмотрение теоретических основ понятия «data mining» в психологии; изучение возможности использования некоторых методов data mining в психологии; описание практических примеров применения методов data mining. В исследовании приняли участие 548 курсантов Академии ФСИН России, при создании искусственной нейронной сети использовались обследования августа 2019 г., для проверки гипотезы и создания программы профилактики – обследования 2021 г.

В качестве методик тестирования были выбраны: Опросник самооценки (В. В. Столин, 1985); Диагностика эмоционального интеллекта (Н. Холл); Опросник «Структура индивидуального правосознания» (Д. В. Сочивко, Т. А. Симакова); Самоактуализационный тест (САТ); Опросник психодинамической типологии просоциального (дистантного) поведения (Д. В. Сочивко). Эти данные послужили основой для разработки и создания искусственной нейронной сети, которая впоследствии анализировала данные зачисленных на первый курс будущих сотрудников.

**Ключевые слова:** data mining, курсанты, искусственная нейронная сеть, психологические исследования, профилактика отчислений, психологическое сопровождение.

## Введение

Использование новых методов обработки информации открывает позитивные перспективы поиска новых знаний, в том числе психологических. Математические знания в симбиозе с вычислительной мощностью компьютеров (ЭВМ) могут дать психологической науке мощный толчок развития. Использование инноваций имеет в своем фундаменте философские знания, метод data mining не стал исключением. Предпосылки метода лежат в философии механицизма, а конкретнее – в его ответвлении – коннекционизме. Коннекционизм (от англ. *connectionism* от *connection* – связь): 1) психологическое направление, заявленное в работах Э. Л. Торндайка, посвященных исследованию законов научения как формирования связей между стимулом (ситуацией) и реакциями организма; 2) направление когнитивной науки, обозначаемое как нейронные сети [1]. По нашему мнению, определение данного понятия заужено, необходимо включить в него не только нейронные сети, но и все смежные методы, которые возможно объединить общим термином «data mining», то есть добыча данных.

Data mining – это собирательное значение для методов, дающих возможность ориентироваться в потоке информации, необходимых для принятия решений в различных сферах человеческой деятельности. Данный термин был впервые введен Г. И. Пятецким-Шапиро. Огромный прогресс виден и в text mining (большинство программных комплексов для data mining теперь включают в себя компоненты для text mining), а также в multimedia mining. И то и другое – прекрасные области для исследований [2]. Data mining – это процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей (шаблонов информации) [3].

Использование в практической области data mining достаточно распространено. Существуют научные исследования, связанные с этим методом, в том числе

отечественные. А. Б. Мосягин использовал методологию data mining при решении задач обработки социальных данных [4]. Т. А. Нестик, А. Л. Журавлев исследовали результаты сбора и учета больших данных, характеризующих реальное поведение людей, а также последствия влияния больших данных на общество, социальные науки и психологическое знание [5]. Р. Б. Куприянов, Д. Л. Агранат, Р. С. Сулейманов не только применили data mining в психологическом знании, но и выстроили систему индивидуальных образовательных траекторий [6]. На сегодняшний день есть примеры законченных исследований в психологической науке с использованием искусственной нейронной сети (ИНС) (А. А. Арзамасцев, П. А. Азарова, Н. А. Зенкова [7], Е. В. Славутская, В. С. Аbruков, Л. А. Славутский [8], М. Г. Доррер [9] и др.).

### Методы

Data mining можно условно разделить на три уровня [10]:

- 1) поверхностный – по аналогии с психодиагностикой это простая первичная обработка данных. Цифровое представление ответов на стимулы в шкалах методики;
- 2) неглубокий – по той же аналогии. Это статистическая обработка данных, то есть применение различных критериев (t-Стюдента) для подтверждения гипотез;
- 3) скрытый – раскрывается понятие коннекционизма. Обработка данных происходит с элементами искусственного интеллекта, нейросетевого анализа и др.

Рассмотрим применение data mining в психологических исследованиях более подробно. Следует обозначить методы, которые, по нашему мнению, могут быть использованы в психологических исследованиях: муравьиный алгоритм, искусственная нейронная сеть, клеточный автомат, генетический алгоритм (табл. 1). Именно в такой последовательности от менее требовательных к вычислительным мощностям ЭВМ к более требовательным.

Таблица 1

Методы data mining в психологии

Метод data mining	Область применения в психологии	Ресурсозатратность
1	2	3
Муравьиный алгоритм	Данный подход к поиску оптимального пути представлен в вычислениях, производимых компьютерами. Если за оптимальный путь принять психологические параметры, подходящие, например, под определенный род деятельности (по примеру профессионального психологического отбора), то алгоритм в вычислениях подберет те значения из имеющихся, которые будут максимально подходить под заявленные психологические требования (например, кандидатов на должность)	Использование мощности ЭВМ невелико, но пропорционально количеству точек оптимизации пути
ИНС	Использование ИНС дает существенные преимущества по сравнению с муравьиным алгоритмом. ИНС не только анализирует оптимальный путь, но и прогнозирует успешность его нахождения, вычисляет и анализирует связи имеющихся значений	Использование мощности ЭВМ достаточно велико, но пропорционально входящим данным и количеству скрытых слоев

Окончание таблицы 1

1	2	3
Клеточный автомат	Использование клеточного автомата – это уровень взаимодействия между данными психологического характера. Каждый представленный населяющий клеточный автомат бот включает в себя ИНС, а она соответственно муравьиный алгоритм. Таким образом клеточный автомат создает условия, вычисляет и анализирует взаимодействие связей существующих значений	Использование мощности ЭВМ велико и не хватит для полноценного моделирования в полном объеме человека. На сегодняшний день представляется возможным использование в анализе простейших микроорганизмов, вирусов, бактерий
Генетический алгоритм	Использование генетического алгоритма в авторском представлении именно как обобщающего метода, применимого в связке с клеточным автоматом, ИНС, муравьиным алгоритмом, позволит создавать новые значения	Использование мощности ЭВМ велико. На сегодняшний день не представляется возможным

Муравьиный алгоритм – один из методов искусственного интеллекта, предложенный М. Дориго. Основная идея алгоритма подсмотрена в природе и имитирует движение колонии муравьев [11].

Искусственная нейронная сеть – математическая модель, а также ее программное или аппаратное воплощение, построена по принципу организации и функционирования биологических нейронных сетей – сетей нервных клеток живого организма [12].

Клеточный автомат – среда обитания, закольцованная внутри себя, созданная модель программной реальности с моделируемыми законами, ресурсами, наполняемая искусственными программируемыми организмами (ботами), по аналогии с природой – вода и одноклеточные организмы. Клеточные автоматы – это «дискретные динамические системы, поведение которых полностью определяются в терминах локальных зависимостей... в информатике они являются аналогом физического понятия «„поле”» [13].

Генетический алгоритм – это алгоритм поиска, используемый для решения задач оптимизации и моделирования путем случайного подбора. Создатель этого метода Н. А. Барричелли был одним из пионеров эволюционных вычислений [14].

В практическом исследовании, проведенном нами на базе Академии ФСИН России, применялась ИНС по следующим причинам: во-первых, ИНС – один из самых развитых и наиболее исследованных в практическом аспекте методов, о чем уже упоминалось ранее; во-вторых, относительно небольшое использование ресурсов в построении ИНС, развитые программные комплексы для обработки данных ИНС. ИНС весьма многочисленны. Классификация ИНС выглядит следующим образом:

1) по типу организации слоев ИНС:

– полносвязные нейронные сети – структуры, в которых каждый нейрон сети имеет прямую связь с другими нейронами;

– многослойные нейронные сети (перцептроны) – нейроны объединяются в слои, содержащие совокупность нейронов с единными входными сигналами. Могут содержать входной, выходной и N промежуточных слоев;

2) типу входной информации:

– аналоговые нейронные сети – используют информацию в форме действительных чисел;

– двоичные нейронные сети – оперируют с информацией, представленной в двоичном виде;

– образные нейронные сети – оперируют с информацией, представленной в виде образов: знаков, иероглифов, символов;

3) характеру настройки синапсов:

– сети с фиксированными связями – весовые коэффициенты нейронной сети выбираются сразу, исходя из условий задачи;

– сети с динамическими связями – для них в процессе обучения происходит настройка синаптических связей.

4) характеру обучения:

– нейронные сети, использующие обучение с учителем;

– нейронные сети, использующие обучение без учителя;

– смешанная парадигма обучения.

Рассмотрим структуру ИНС. Основой нейросети служит нейрон, который состоит из входных данных ( $X$ ), весов связей ( $W$ ), сумматора, функции активации, выходных данных ( $Y$ ). Входные данные – это данные числового вида, которые могут поступать из любых источников, например данные о набранных баллах испытуемых при проведении психодиагностики. Веса связей – результат обучения ИНС, то есть изменяемые числа, благодаря которым ИНС обретает «опыт» предыдущих итераций. Сумматор и функции активации – внутренние инструменты ИНС, необходимы для ее функционирования (рис. 1).

В нашем исследовании ИНС создавалась с использованием следующих программных продуктов: IBM SPSS 22.0, Neural Excel, STATISTICA 13.3 Automated Neural Networks. Данные программы работали независимо друг от друга, что также позволяло увеличить прогноз и валидность результатов. Критериями успешно созданной ИНС послужили: высокий прогноз (минимум 85 %), стремление к минимальной вычислительной мощности, что уже обсуждалось в научном сообществе как оптимальное значение [15]. В качестве прототипа ИНС был выбран многослойный перцептрон. Для «обучения»

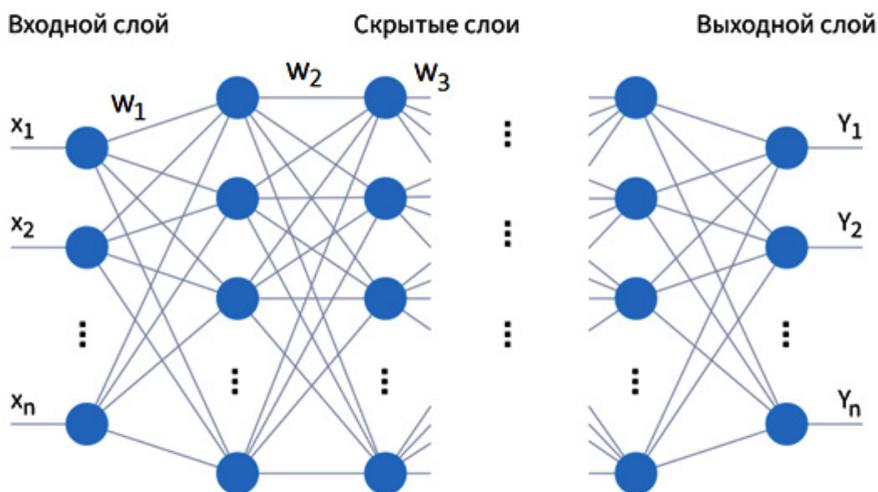


Рис. 1. Многослойная ИНС:

$X_1, X_2, X_3$  – данные психодиагностического обследования клиента (ответ на вопросы или баллы шкал);  $W_1, W_2, W_3$  – веса ИНС (изменяемые коэффициенты с целью оптимального результата);  $Y_1, Y_2, Y_3$  – результат прогноза или псевдовероятность

ИНС использовались «сырые» баллы методик: Опросник самооотношения (В. В. Столин, 1985); Диагностика эмоционального интеллекта (Н. Холл); Опросник «Структура индивидуального правосознания» (Д. В. Сочивко, Т. А. Симакова); Самоактуализационный тест (САТ); Опросник психодинамической типологии просоциального (дистантного) поведения (Д. В. Сочивко).

Ход создания ИНС выглядел следующим образом. Данные эмпирического обследования курсантов были переведены в матрицу данных Excel, где курсанты, отчисленные по отрицательным мотивам, составляли отдельную группу. В качестве входящих нейронов выступали набранные баллы по шкалам и ответы испытуемых на стимулы перечисленных методик, в качестве выходных нейронов – результаты классификации по параметру «отчислен» – «не отчислен», то есть вероятность принадлежности к группе, представленная в процентах. Всего нейронов в ИНС 1118, входящих нейронов – 10, скрытых нейронов – 10, выходов – 2, функция активации – экспонента, Softmax.

### Результаты

Обязательным условием стабильной работы ИНС является процесс обучения – порядок подбора коэффициентов ( $W$ ) с помощью метода обратного распространения ошибки. Заключается он в зависимости коэффициента ( $W$ ) от результата анализа ИНС (в нашем случае принадлежность к группе отчисленных курсантов). При первой итерации коэффициенты ( $W$ ) подбираются случайно, а далее корректируются разностью промежуточного результата вывода ИНС. Основная цель данного метода состоит в обучении ИНС определению максимально схожего значения в группе отчисленных курсантов. На практике это выглядело следующим образом. В августе 2019 г. было проведено обследование 273 курсантов Академии ФСИН России по методикам, упоминавшимся ранее. Принадлежность к факультетам, демографические данные не учитывались. На момент августа 2021 г. 25 курсантов было отчислено из Академии ФСИН России по следующим причинам: слабая успеваемость, низкая дисциплина и по собственному желанию (отсутствие мотивации к учебе, плохая адаптация и т. п.). 248 курсантов из обследованных продолжали обучение в Академии ФСИН России. Группу отчисленных курсантов разделили пропорционально: 70 и 30 %. Большую часть использовали для обучения ИНС, меньшую – в качестве проверки результатов обучения. Обучение ИНС происходило с помощью программы IBM SPSS Statistics 22.0 и STATISTICA 13.3 Automated Neural Networks.

В результате анализа было выявлено, что ИНС вполне справляется с поставленными задачами и может с высокой вероятностью предсказывать результат отчисления. В результате обученная ИНС смогла предсказать 88 % отчисленных курсантов (табл. 2).

Таблица 2

Результаты работы созданной ИНС

Группы	Количество курсантов	Отчислено по мнению ИНС	Обучаются по мнению ИНС	Точность прогноза ИНС, %
Отчисленные	25	22	3	88
Обучающиеся	248	29	219	89

Работа обученной ИНС заключалась в следующем. В августе 2021 г. проведено психологическое обследование курсантов первого курса набора 2021 г. по методикам, упомянутым выше. Всего обследовано 275 чел. Данные психодиагностики загрузились в



Рис. 2. Нормализованная важность параметра

обученную ИНС, которые она анализировала. В результате анализа из 275 курсантов 19 были помечены как отчисленные. Тем самым ИНС обозначила группу курсантов, которые могут быть отчислены.

В упомянутых программных комплексах существует функция, благодаря которой возможно наблюдать, на какие данные ИНС опиралась при составлении прогноза, и выстроить их в иерархию. Программная платформа IBM SPSS Statistics 22.0 создает иерархию важности компонентов, данный параметр имеет название «Нормализованная важность». Программный комплекс STATISTICA 13.3 Automated Neural Networks имеет отличное название этой функции – «Сенсорная чувствительность ИНС». Такая функция позволяет выявить степень значимости того или иного показателя на выходные решения. Она оценивает прогнозную силу каждого отдельного параметра и на его основе ранжирует факторы. На рисунке 2 представлены наиболее значимые переменные, которые имелись в нашей сети при ранжировании с помощью показателя нормализованной важности.

На левой стороне рисунка 2 – название шкал перечисленных методик, то есть факторы, качества, умения – то, что описывают эти шкалы, на правой – количественное выражение, вклад данных шкал в структуру ИНС. Другими словами, баллы шкалы «Самоуважение» методики «Самоактуализационный тест» на уровне 3,5 % влияют на результата прогноза ИНС. Значит, для профилактики отчисления из вуза необходимо корректировать уровень самоуважения курсанта.

На основе описанных исследований была создана программа профилактики. Данная программа включала в себя блок психодиагностики и сессии индивидуальных консультативных встреч с курсантами, которых ИНС отметила как отчисленные. Уникальность программы заключается в ее универсальности. Программу возможно модифицировать таким образом, что в поле ее зрения будут попадать любые респонденты с заданными характеристиками. Цель программы – это диагностика и профилактика негативных психологических особенностей курсантов ведомственных вузов ФСИН России, которые способствуют их отчислению. Задача программы – диагностировать респондентов,



Рис. 3. Схема работы психодиагностики с помощью ИНС

выявлять из их числа нуждающихся в психологической помощи, указывать психологу психологические особенности, необходимые для коррекции.

Первая часть программы – диагностика с помощью ИНС, поиск курсантов, которые нуждаются в профилактике отчисления, а также поиск психологических качеств для коррекции (рис. 3).

Блок 1 – ввод данных осуществляется испытуемыми при проведении психодиагностического обследования, ответы заносятся в базу данных, например на платформе Excel.

Блок 2 – на основе базы данных результатов обследований испытуемых создается ИНС. С Excel осуществляется перенос данных в программные комплексы IBM SPSS 22.0, STATISTICA 13.3 Automated Neural Networks, посредством которых формируется и обучается ИНС.

Блок 3 – использование прогноза ИНС в работе – это результат работы ИНС. В прогноз входит количественное представление соответствия с «внутренним представлением» ИНС. Психолог получает конкретный список испытуемых, с которыми необходимо провести психологические мероприятия (консультация или коррекция). В прогноз также входит иерархия факторов, которые влияют на построения ИНС, они же психологические свойства, которые следует корректировать. После прогноза у психолога есть конкретные курсанты и их психологические свойства и качества, что позволяет сосредоточить усилия именно там, где необходима психологическая помощь.

Блок 4 – данные, получаемые при прогнозе, могут отличаться от реальной обстановки в силу объективных причин. Стандартное эффективное использование ИНС строится на уровне 80–90 % достоверности. Если учесть, что и окружающие обстоятельства испытуемых также могут меняться, то ИНС необходимо корректировать, внося новые данные, и переобучать. Этот процесс сделает систему динамичной и гибкой к изменениям окружающей действительности.

Организация работы пенитенциарного психолога по указанной схеме позволит эффективно очерчивать круг курсантов, с которыми необходимо проводить работу по профилактике отчисления. Коррекция ИНС в зависимости от новых данных поможет не потерять актуальность.

Вторая часть программы связана с развитием свойств, направленных на профилактику отчисления. Нормализованная важность или сенсорная чувствительность программ, создающих ИНС, позволяют выдать те качества, умения, свойства психики, которые необходимо развивать. Эта информация была перенесена в план индивидуальных консультаций с курсантами, которых ИНС отметила как отчисленных. Консультативные сессии проводились еженедельно на протяжении 2 месяцев, каждый курсант в среднем посетил психолога 4 раза. При повторном тестировании группы отчисленных были получены следующие результаты (табл. 3).

Таблица 3

### Результаты по отдельным шкалам

Группа	Шкалы методик					
	Социальная тревожность	Самоуважение	Самотрансцендентность	Ответственность	Интегральное чувство за или против	Управление своими эмоциями
	В сырых средних баллах по выбранной группе					
До коррекции	5,90	12,09	61,86	43,13	20,05	6,68
Контрольная группа	7,45	11,4	61,94	41,04	20,79	6,86
После коррекции	6,68	11,13	61,22	41,86	20,54	6,45

Результаты повторного психодиагностического обследования загружались в ИНС для анализа. Целью служила оценка нейросетью проведенной работы и повторный прогноз по группе. Из 19 курсантов, отмеченных ранее как отчисленные, ИНС при повторном прогнозе оставила 7 курсантов.

### Выводы

1. Использование ИНС как одного из методов data mining является эффективным. Применение в качестве материалов для обучения ИНС данных психологических обследований отчисленных курсантов позволит объективизировать принятие решений в выборе объектов для усиленного психологического сопровождения. Обученная ИНС в состоянии эффективно справляться с задачами прогноза психологического сопровождения.

2. Структурные составляющие ИНС позволяют очертить психологические качества, свойства в качестве объектов в контексте профилактики отчисления курсантов из Академии ФСИН России. Данные ИНС позволяют создать эффективные планы консультаций с курсантами.

3. В результате проведенного исследования стало возможным создать и полностью апробировать систему профилактики отчисления курсантов Академии ФСИН России. Эта система была внедрена в деятельность психологических служб Академии ФСИН России и Самарского юридического института ФСИН России.

### Библиографический список

1. Психологическая энциклопедия. URL : [https://gufo.me/dict/psychology\\_encyclopedia](https://gufo.me/dict/psychology_encyclopedia) (дата обращения: 01.03.2020).

2. Левкович-Маслюк Л. И. Великие раскопки и великие вызовы // Компьютерра. 2007. № 11. С. 48–51.
3. Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям. М. : Питер, 2013. 701 с.
4. Мосягин А. Б. Использование методологии data mining при решении задач обработки социальных данных // Большая социология: расширение пространства данных : материалы V Междунар. социол. Грушинской конф. М., 2015. С. 143–145.
5. Нестик Т. А., Журавлев А. Л. Анализ больших данных в психологии и социогуманитарных науках: перспективные направления исследований // Психологический журнал. 2019. Т. 40. № 6. С. 5–17.
6. Куприянов Р. Б., Агранат Д. Л., Сулейманов Р. С. Использование технологий искусственного интеллекта для выстраивания индивидуальных образовательных траекторий обучающихся // Вестник Российского университета дружбы народов. Сер. Информатизация образования. 2021. Т. 18, № 1. С. 27–35.
7. Арзамасцев А. А., Азарова П. А., Зенкова Н. А. Модель профессиональных и личностных качеств студентов университета на основе искусственной нейронной сети с адаптивной структурой // Вестник тамбовского университета. Сер. Естественные и технические. 2007. № 5. С. 623–632.
8. Славутская Е. В., Аbruков В. С., Славутский Л. А. Простые нейросетевые алгоритмы для оценки латентных связей психологических характеристик младших подростков // Экспериментальная психология. 2019. № 2. С. 131–144.
9. Доррер М. Г. Психологическая интуиция искусственных нейронных сетей : дис. ... канд. тех. наук. Красноярск, 1998. 127 с.
10. Дюк В., Самойленко А. Data mining : учеб. курс. СПб. : Питер, 2001. 368 с.
11. Кирсанов М. Н. Графы в Maple. М. : Физматлит, 2007. 168 с.
12. Нейронная сеть // Большая российская энциклопедия : в 35 т. / гл. ред. Ю. С. Осипов. М. : Большая российская энциклопедия, 2004–2017.
13. Астафьев Г. Б. Клеточные автоматы : учеб.-метод. пособие. Саратов, 2003. 24 с.
14. Fogel, D. B. 2006, 'Historic perspective – Nils Barricelli-artificial life, coevolution, self-adaptation', *IEEE Computational Intelligence Magazine*, iss. 1, pp. 41–45.
15. Курдин Д. А. Прогноз успешности обучения курсантов с помощью искусственной нейронной сети на основе результатов исследования динамики личностного роста // Психология XXI века: вызовы, поиски, векторы развития : сб. материалов Всерос. симпозиума психологов с международным участием. Рязань : Академия ФСИН России, 2020. С. 799–805.