



Кафедра информатики
и информационных технологий

Б1.О.13.01 Основы информационных технологий

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

**к самостоятельной работе по теме
«Интеллектуальные информационные системы»**

Направление подготовки

35.03.04 Агрономия

Профиль подготовки

Цифровые технологии в растениеводстве

Квалификация (степень) выпускника

бакалавр

Уфа 2022

Рекомендовано к изданию методической комиссией экономического факультета (протокол № 6 от 24.03.2022 г.)

Составитель: доцент, к.ф.-м.н. Шамсутдинова Т.М.

Рецензент: ст. преподаватель Прокофьева С.В.

Ответственный за выпуск: зав. кафедрой ИИТ, д.т.н. Беляева А.С.

Г.Уфа, БГАУ, Кафедра информатики и информационных технологий

Интеллектуальные информационные системы

Цель самостоятельной работы: формирование у обучающихся общепрофессиональных компетенций, необходимых для профессиональной деятельности с использованием интеллектуальных информационных технологий.

1 Общие положения

1.1 Содержание раздела «Интеллектуальные информационные системы»

Введение в интеллектуальные информационные системы.

Понятие «Интеллект». Интеллектуальные задачи. Искусственный интеллект. Свойства интеллектуальных систем. База знаний. Признаки и классы интеллектуальных систем. Системы с интеллектуальным интерфейсом. Экспертные системы. Самообучающиеся системы. Адаптивные информационные системы.

Представление знаний в интеллектуальных системах.

Понятие «Знание». Методы классификации знаний. Примеры форм и моделей представления знаний. Примеры моделей представления знаний. Продукционная модель. Фреймовая модель. Семантическая модель. Логическая модель. Нечеткие знания. Пример нечеткой лингвистической переменной.

Интеллектуальный анализ данных. Data Mining.

Интеллектуальный анализ данных. Извлечение знаний из баз данных. Data Mining: модели. Кластерный анализ данных. Анализ текста. Text Mining. Большие данные. Big Data. Эволюция аналитики. Большие данные: Big Data Analytics. Технологии MapReduce.

Понятие нейросетевых технологий.

Биологический и искусственный нейрон. Модель нейрона Мак-Каллока и Питтса (McCulloch - Pitts). Многослойная нейронная сеть. Как работает нейронная сеть. Области применения нейронных сетей. Этапы решения нейросетевых задач. Обучение нейронных сетей. Пример алгоритма обучения нейронной сети (правило Хебба). Распознавание образов. Проектирование нейронных сетей.

Направления и перспективы развития интеллектуальных систем.

Развитие искусственного интеллекта. Инженерия знаний. Машинное обучение. Нейронные сети. Распознавание образов. Распознавание и синтез речи. Эволюционные и генетические алгоритмы. Робототехника. Новые архитектуры и технологии. Применение интеллектуальных систем в сельском хозяйстве.

Пример прогнозирования на основе временных рядов (видеопрактикум).

Прогнозирования временных рядов в программном комплексе Loginom. Применения регрессионных технологий для задачи прогнозирования изменения объема сезонных продаж по ежемесячным данным за 4 года.

Пример анализа данных с использованием нейронных сетей

(видеопрактикум).

Применения нейронных сетей для оценки урожайности с/х культур. Построение нейросетевой модели оценки урожайности озимой пшеницы в программном комплексе *Loginom*

1.2 Организация самостоятельной работы по теме «Интеллектуальные информационные системы»

В рамках изучения темы «Интеллектуальные информационные системы» работа обучающихся организуется в следующих формах:

- изучение лекционного материала (видеолекций, презентаций к лекциям);
- получение практических навыков по работе с программным комплексом *Loginom*;
- подготовка к лабораторной работе «Прогнозирование на основе временных рядов с использованием *Loginom*»;
- изучение вопросов, выносимых за рамки лекционных занятий, в рамках самостоятельного изучения теоретического материала (СИТМ);
- решение дополнительных задач (включая задачи повышенного уровня сложности);
- подготовка к тестированию.

Формами контроля самостоятельной работы по теме могут являться следующие технологии:

- самооценка обучающегося;
- оценивание преподавателем;
- взаимная оценка обучающихся;
- компьютерное тестирование и др.

2 Методические указания к самостоятельной работе

2.1 Подготовка к лабораторной работе и получение практических навыков по работе с программным комплексом *Loginom*

Лабораторная работа по теме «Прогнозирование на основе временных рядов с использованием *Loginom*» должна проводиться в компьютерном классе с установленной программой *Loginom*.

Программный комплекс *Loginom* относится к типу программ «Системы сбора, хранения, обработки, анализа, моделирования и визуализации массивов данных» и внесен в Реестр российского программного обеспечения.

Бесплатная версия для некоммерческого использования доступна на сайте разработчика по ссылке

<https://loginom.ru/download>

(где необходимо заполнить анкету, после чего высылается ссылка для скачивания на указанный в анкете e-mail).

При изучении принципов работы в системе *Loginom* можно использовать руководство пользователя <https://help.loginom.ru/userguide/>

В качестве контроля знаний по основам работы в программном комплексе *Loginom* могут быть использованы тестовые вопросы из приложения А.

2.2 Кейс-задание для самостоятельной работы

Рассмотрим возможность применения нейронных сетей для оценки урожайности сельскохозяйственных культур (на примере урожайности озимой пшеницы).

Среди внешних факторов, способных повлиять на урожай, можно выделить важные моменты:

- природно-климатические факторы – сюда входит как состояние почвы и ее состав, так и количество осадков и т.д.;
- обеспечение посадок удобрениями, включая различные виды органических и минеральных удобрений, и др.

Задача:

Подготовить обучающий набор данных для обучения нейронной сети оценке урожайности с/х культур.

Содержание кейса:

Используя видеопрактикум «Пример анализа данных с использованием нейронных сетей (видеопрактикум)», изучить этапы применения нейронных сетей для оценки урожайности сельскохозяйственных культур.

Проанализировать последовательность построения нейросетевой модели оценки урожайности озимой пшеницы на основе следующих данных:

Входные данные:

- Качество почв, баллы от 1 до 100 (экспертная оценка);
- Осадки, мм;
- Внесение минеральных удобрений на 1 га, кг;

Выходные данные:

- Урожайность озимой пшеницы с 1 га, ц.

	А	В	С	Д
1	Качество почв, баллы	Осадки, мм	Внесение минеральных удобрений на 1 га, кг	Урожайность озимой пшеницы с 1 га, ц
2	45	1,3	84	20,6
3	45	1,3	84	20,6
4	45	1,3	105	21,15
5	64	0,7	97	21,25
6	67	0,7	101	22,75
7	48	0,9	164	22,3
8	58	1,4	75	22,85
9	65	0,8	115	23,05

Рисунок 1 – Фрагмент таблицы исходных данных для обучения нейросети

Сформулировать, какие еще дополнительные входные факторы можно учесть при построении модели оценки урожайности?

С использованием открытых источников сети Интернет подготовить электронную таблицу, содержащую данные для возможного обучения нейронной сети оценке урожайности с/х культур. Количество строк в обучающем наборе данных - не менее 20-30.

Результат самостоятельной работы представить в виде таблицы *MS Excel*.

2.3 Дополнительное задание повышенного уровня сложности*
(выполняется по желанию, оценивается дополнительными поощрительными баллами).

Используя разработанную в пункте 2.2 (кейс-задание) таблицу с обучающим набором данных, самостоятельно реализовать нейросетевую модель оценки урожайности сельскохозяйственных культур в программном комплексе *Loginom*.

Методические указания к выполнению дополнительного задания

Построим нейросетевую модель оценки урожайности озимой пшеницы на основе обучения по следующим эмпирическим данным:

- Качество почв, баллы от 1 до 100 (экспертная оценка);
- Осадки, мм;
- Внесение минеральных удобрений, кг на 1 га;
- Урожайность озимой пшеницы с 1 га, ц.

Количество заполненных строк в обучающем наборе данных - не менее 30.

Этапы построения нейросетевой модели в программном комплексе *Loginom* включают в себя:

- импорт исходных данных в систему;
- настройку и обучение модели «Нейросеть (регрессия)»;
- расчет ошибки аппроксимации, анализ данных;
- экспорт полученных численных результатов в *MS Excel*.

В сценарии обработки данных будем использовать модель «Нейросеть (регрессия)». Данная нейросеть решает задачу регрессионного анализа, связанного с выявлением зависимости прогнозируемого значения от набора входных параметров. Перед построением прогноза модель предварительно обучается на обучающей выборке.

Рассмотрим настройку нашего сценария в системе *Loginom*. Сценарий состоит из следующих узлов:

- 1) импортируемый файл *MS Excel* с исходными данными;
- 2) обработчик «Нейросеть (регрессия)» для построения модели прогноза;
- 3) обработчик «Калькулятор» для расчета ошибки аппроксимации;
- 4) экспортируемый файл *MS Excel* с результатами работы.

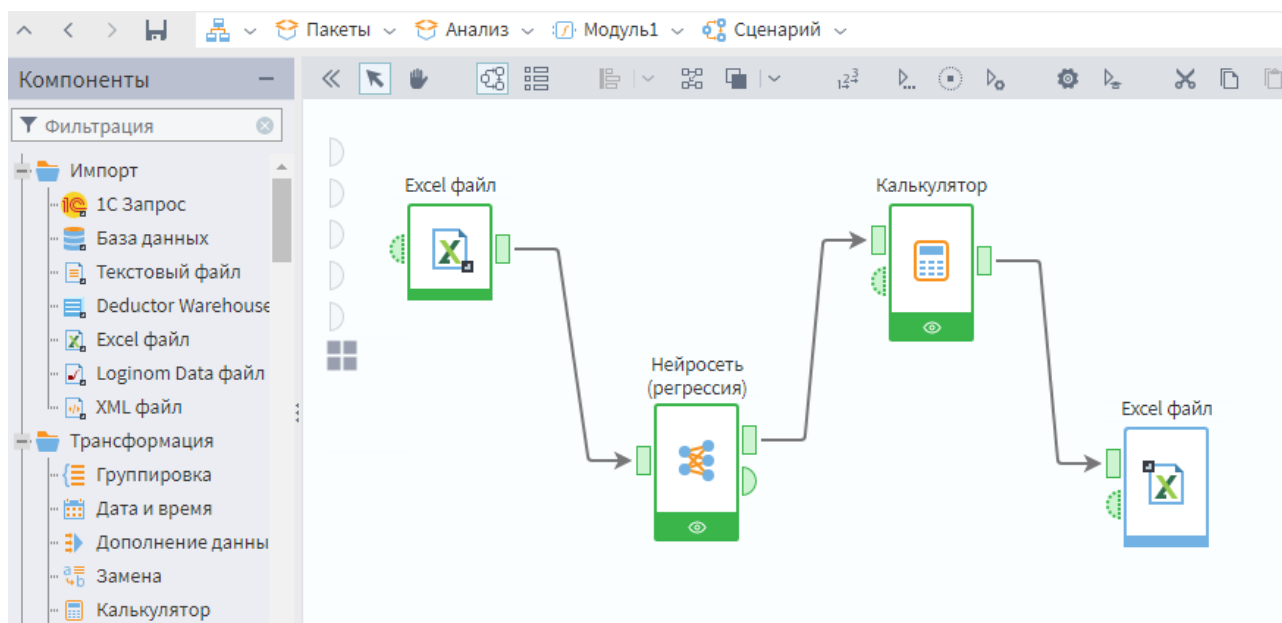


Рисунок 2 – Сценарий обработки данных в системе Loginom

Настроим узел Импорта данных из файла MS Excel. Вызовем Мастер настройки узла, укажем имя файла MS Excel с исходными данными. Оставим остальные настройки без изменения и завершим работу с узлом, выполнив узел (кнопка «Выполнить узел»).

Настроим узел «Нейросеть (регрессия)». При этом проведем настройку и обучение нейронной сети.

Данная настройка проводится по шагам.

Шаг 1. Настройка входных столбцов. На данном шаге необходимо задать назначение полей входного множества данных.

В нашем примере на выходе должна быть прогнозируемая урожайность, соответственно входом модели должны являться остальные данные. Установим:

- «Качество почв, баллы» – входное поле;
- «Осадки, мм» – входное поле;
- «Внесение минеральных удобрений, кг на 1 га» – входное поле;
- «Урожайность озимой пшеницы с 1 га, ц» – выходное поле.

Шаг 2. Настройка нормализации. В машинном обучении нормализацией называют метод предобработки данных с целью их приведения к некоторой общей шкале. При нормализации входные данные преобразуются к определенным заданным диапазонам в соответствии с выбранным алгоритмом. Установим для всех данных **Масштабирование [0;1]** – приведение данных линейным преобразованием к диапазону от 0 до 1.

Шаг 3. Разбиение на множества. На данном шаге входные данные можно разделить на обучающее и тестовое множества, а также настроить метод валидации. *Валидация* – это проверка правильности работы аналитической модели, построенной на основе машинного обучения. Установим разбиение на обучающее и тестовое множество как 95 к 5. Выберем метод «K-fold кросс-валидация».

Следующие шаги:

Шаг 4. *Настройка параметров нейросети.*

Шаг 5 *Настройка автоматического подбора параметров Нейросети.*

Оставим настройки данных шагов параметров нейросети по умолчанию без изменения.

Для созданной нейронной сети запустим команду ее обучения (вызывается в системе команд меню узла как *Другие действия/ Переобучить узел*).

Для наглядного представления результатов работы нейронной сети можно при желании настроить *Визуализатор Диаграмма* в узле *Нейросеть*. При этом можно построить диаграмму разброса данных, показывающую качество обучения нейросети.

В Диаграмме нужно перетащить в область построения исходную и прогнозируемую урожайность, тип графиков в обоих случаях установить как «Разброс» (в окне *Добавить серию* установить Тип: *Разброс*). Осью абсцисс установить исходную урожайность (для этого перетащить данное поле из левой панели в нижнюю часть диаграммы).

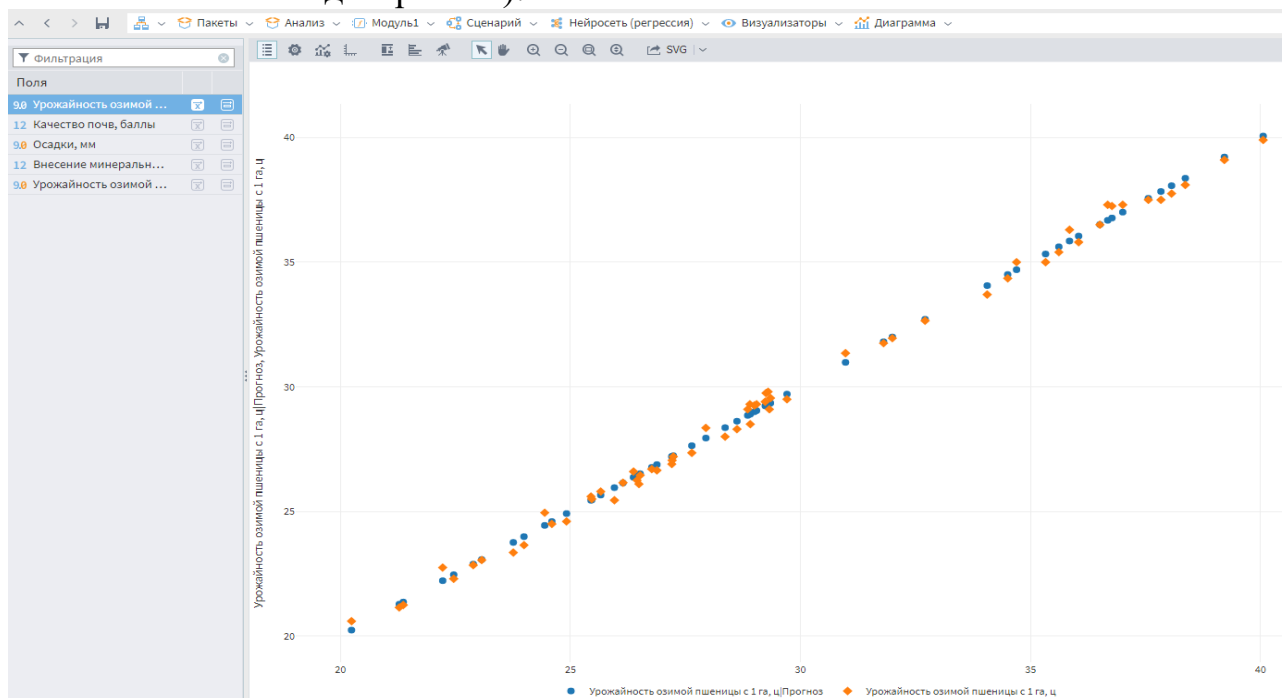


Рисунок 3 – Диаграмма разброса данных

Далее рассчитаем ошибку аппроксимации нашего прогноза с помощью обработчика «Калькулятор». С помощью встроенной в калькулятор функции *RelErr* рассчитаем ошибку аппроксимации для каждого объекта как

$\text{RelErr}(\text{Урожайность озимой пшеницы с 1 га, ц_Прогноз}, \text{Урожайность озимой пшеницы с 1 га, ц}) * 100$

Примечание. При создании формулы следует использовать не русскоязычные *метки* полей, а их *Имена* в системе (при этом достаточно перетащить мышью данные поля из левой нижней части окна в формулу) (рис. 4).

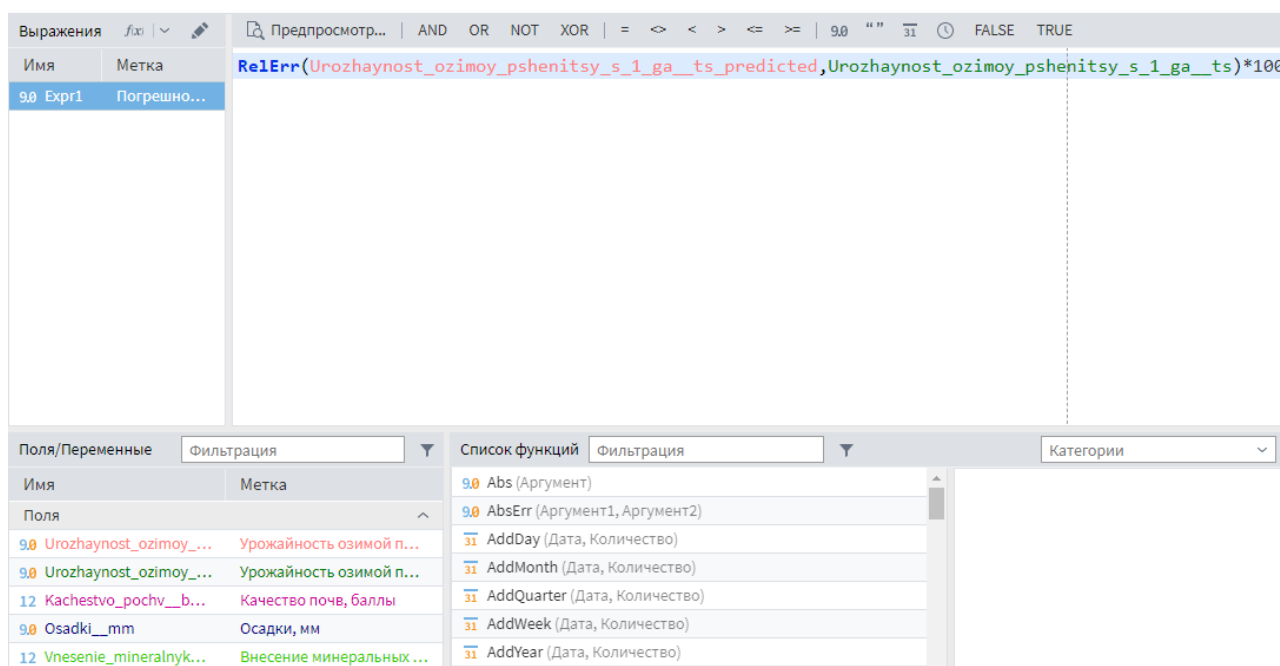


Рисунок 4 – Расчет ошибки аппроксимации прогноза

Данное значение показывает, насколько расчетные значения отличаются от фактических значений, т.е. дается представление о качестве модели.

Настроим для калькулятора визуализатор *Статистика* для расчета статистических значений. Данный визуализатор предназначен для просмотра различных статистических показателей по каждому полю набора данных. В общем случае, средняя ошибка аппроксимации обучения нейронной сети до 10% считается приемлемой.

Далее завершим работу с калькулятором и проведем экспорт результатов в файл MS Excel.

В настройках узла укажем имя файла для экспорта данных.

Открыв данный файл MS Excel, мы увидим в нем столбец прогноза.

Далее мы можем реализовать работу нейронной сети на новых данных, т.е. на данных, которые не использовались при ее обучении. Создадим в Excel файл с данными для анализа, причем укажем данные, которые не использовались при обучении сети. Затем перезапустим узел импорта Excel. Далее запустим всю сеть без переобучения остальных узлов.

В результирующем файле экспорта Excel появится новое значение прогноза. Тем самым, мы выполнили цель нашей работы – рассмотрели возможность применения нейронных сетей для оценки урожайности сельскохозяйственных культур.

2.4 Самостоятельное изучение теоретического материала

При самостоятельном изучении теоретического материала (СИТМ) рекомендуется ориентироваться на содержание раздела «Интеллектуальные информационные системы», представленное в пункте 1.1.

В качестве контроля изучения теоретического материала могут быть использованы тестовые вопросы проверки знаний (приложения Б, В) и

контрольные вопросы для самостоятельного изучения.

Контрольные вопросы для самостоятельного изучения теоретического материала:

- 1) История возникновения систем искусственного интеллекта. Тест Тьюринга
- 2) Экспертные системы: история появления, классификация, этапы проектирования
- 3) Технологии анализа данных KDD (Knowledge Discovery in Databases - извлечение знаний из баз данных)
- 4) Кластерный анализ данных: примеры прикладных задач. Методы кластерного анализа
- 5) Биологический нейрон как прототип искусственных нейронных сетей
- 6) Персептрон Розенблатта
- 7) Нейронные сети: классификация, модели.
- 8) Эволюционные и генетические алгоритмы: основная терминология, история развития
- 9) Робототехника и ее прикладное значение
- 10) Примеры применения интеллектуальных информационных систем в сельском хозяйстве

Библиографический список

- 1) Андрейчиков А.В. Интеллектуальные информационные системы и методы искусственного интеллекта / А. В. Андрейчиков, О. Н. Андрейчикова. – М. : НИЦ ИНФРА-М, 2022. – 530 с.
- 2) Замятин А.В. Интеллектуальный анализ данных. Томск: Издательство Томский государственный университет, 2020. – 196 с.
- 3) Ярушкина Н.Г., Перфильева И.Г. Интеллектуальный анализ временных рядов. М.: Издательство Издательский Дом ФОРУМ, 2022. – 160 с.
- 4) ARIMAX. Описание [Электронный ресурс]: Loginom Company, 2022. URL: <https://help.loginom.ru/userguide/processors/datamining/arimax.html> (дата обращения: 26.04.2022).

Приложение А

Тест для проверки знаний по основам работы в программном комплексе *LOGINOM*

1. В состав программного комплекса *Loginom* НЕ входит следующий раздел компонентов
 - A) Импорт
 - B) Экспорт
 - C) Data Mining
 - D) Процессинг
2. Программный комплекс *Loginom* допускает импорт данных из текстового файла
 - A) Верно
 - B) Неверно
3. В системе *Loginom* обработчик *Сглаживание* относится к разделу компонентов
 - A) Импорт
 - B) Трансформация
 - C) Предобработка
 - D) Data Mining
4. В системе *Loginom* обработчик *ARIMAX* относится к разделу компонентов
 - A) Импорт
 - B) Трансформация
 - C) Предобработка
 - D) Data Mining
5. В системе *Loginom* обработчик *ARIMAX* реализует модель
 - A) авторегрессионного анализа данных
 - B) автокорреляционного анализа данных
 - C) линейного программирования
 - D) массового обслуживания
6. В системе *Loginom* в настройке обработчика *ARIMAX* имеется возможность настройки нормализации данных
 - A) Верно
 - B) Неверно
7. В системе *Loginom* в настройке обработчика *Сглаживание* допускается возможность Метода обработки... (укажите 2 варианта ответа):
 - A) фильтр Ходрика-Прескотта
 - B) Вейвлет-сглаживание

- C) Монте-Карло
- D) K-fold кросс-валидация

8. В системе *Loginot* обработчик *Ассоциативные правила* относится к разделу компонентов

- A) Импорт
- B) Трансформация
- C) Предобработка
- D) Data Mining

9. В системе *Loginot* в настройке обработчика *Нейросеть (регрессия)* при разбиении на множества допускается возможность выбора метода валидации... (укажите 2 варианта ответа):

- A) фильтр Ходрика-Прескотта
- B) Вейвлет-сглаживание
- C) Монте-Карло
- D) K-fold кросс-валидация

10. В системе *Loginot* обработчик *Excel файл* относится к разделам компонентов... (укажите 2 варианта ответа)

- A) Импорт
- B) Трансформация
- C) Экспорт
- D) Data Mining

Приложение Б

Тест для проверки знаний лекционного материала по теме «Интеллектуальные информационные системы»

1. Способность решать задачи путем приобретения, запоминания и целенаправленного преобразования знаний в процессе обучения называют
 - А) интеллектом
 - В) логикой
 - С) памятью
2. Знания типа «ЕСЛИ А, ТО В» называют
 - А) семантическими правилами
 - В) продукционными правилами
 - С) гносеологическими правилами
 - Д) фреймами
3. Логический вывод от общего к частному называется
 - А) Индуктивным
 - В) Дедуктивным
 - С) Абдуктивным
 - Д) Формальным
4. Какая из моделей представления знаний отражает систему знаний в виде образа сети?
 - А) Семантическая модель
 - В) Фреймовая модель
 - С) Логическая модель
 - Д) Продукционная модель
 - Е) воображением
5. Автором фреймовой модели представления знаний является
 - А) Хопфильд
 - В) Хемминг
 - С) Мак-Каллоу
 - Д) Минский
6. Логическая модель представления знаний основана на
 - А) на правилах продукции
 - В) на системе исчисления предикатов
 - С) на тождественных преобразованиях
7. Пример платформы с открытой реализацией технологией MapReduce
 - А) Hadoop
 - В) CMS

C) Python

8. Сверточные нейронные сети чаще всего используются для задач..

- A) диагностики
- B) оптимизации
- C) распознавания образов
- D) Нет верного ответа

9. Примером алгоритма обучения персептрона является

- A) алгоритм Хебба
- B) алгоритм Канторовича
- C) алгоритм Монжа

10. Нейросетевые технологии применимы к задачам биометрической идентификации

- A) верно
- B) неверно

Приложение В

Тест для проверки знаний по самостоятельному изучению теоретического материала

1. Автором теста по определению наличия у компьютерных систем признаков искусственного интеллекта является
 - A) Питтс
 - B) Тьюринг
 - C) фон Нейман
 - D) Мак-Каллоу
2. Экспертные системы, решающие задачи в условиях неопределенных исходных данных - это
 - A) Доопределяющие системы
 - B) Трансформирующие системы
 - C) Классифицирующие системы
 - D) Нет верного ответа
3. На каком этапе проектирования экспертных систем происходит переход от тестовых примеров к решению реальных задач?
 - A) Формализация базы знаний
 - B) Тестирование
 - C) Опытная эксплуатация
 - D) Нет верных ответов
4. Под термином "очистка данных" в технологии KDD понимают:
 - A) транслитерация исходного набора данных
 - B) устранение пропусков, аномальных значений (выбросов)
 - C) преобразование данных к симметричному виду
 - D) выявление скрытых закономерностей и зависимостей
5. Агломеративные методы кластерного анализа характеризуются...
 - A) последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров
 - B) принадлежностью в начале работы алгоритма всех объектов одному кластеру, который на последующих шагах делится на меньшие кластеры
6. Согласно теории искусственных нейронных сетей, в строении биологического нейрона в качестве выходных каналов для нервных импульсов к другим нейронам используются
 - A) Дендриты
 - B) Аксоны
 - C) Сомы
 - D) Синапсы

7. Одной из первых искусственных сетей, способных к обучению при распознавании образов, является PERCEPTON Розенблатта, распознающий:
- A) Буквы алфавита
 - B) Ноты
 - C) Фотографии
 - D) Нет верного ответа
8. В нейронных сетях с прямыми связями:
- A) Нейроны могут соединяться только с нейронами нижележащих слоев
 - B) Допускаются связи внутри одного слоя
 - C) Используются связи с нейронами вышележащих слоев
9. Кто впервые предложил генетический алгоритм как метод решения задач оптимизации?
- A) Дж. Холланд
 - B) Ф. Розенблатт
 - C) Д. Нейман
10. Три закона робототехники были сформулированы
- A) К. Чапеком
 - B) А. Азимовым
 - C) Д. Нейманом
 - D) Л. Канторовичем